Research Article

Supervised meta-heuristic extreme learning machine for multiple sclerosis detection based on multiple feature descriptors in MR images

Adele Rezaee[1] · Khosro Rezaee[2] · Javad Haddadnia[3] · Hamed Taheri Gorji[4]

Abstract

In this study, we propose a hybrid approach involving feature extraction, feature selection, and optimized learning for the diagnosis of multiple sclerosis (MS), which can detect the lesion caused by MS plaques in the brain using magnetic resonance imaging analysis. A major challenge associated with lesion diagnosis by neurologists is that it is a time-consuming process and demands high expertise; therefore, researchers have been stimulated to find an auto-diagnose method of the disease. Given the high resemblance of MS plaque-induced lesions and other lesions such as Alzheimer's or dementia, scant research has explored the diagnosis of MS-induced lesions, most of which suffering from the lack of an efficient and accurate method. Informed by the need for a precise hybrid model for the classification of MS plaques and other comparable lesions, a solution is proposed that utilizes an efficient model. In this method, after image pre-processing, the feature vector is created by applying fractal and Pseudo-Zernike Moments descriptors. Feature selection using the Difference Evolution) algorithm to select the minimum subset of features will reduce the number of Extreme Learning Machine (ELM) inputs for classification. To improve the classification effect, the ELM wavelet kernel parameters are also regulated by the Shuffled Frog-Leaping Algorithm. By applying the proposed model to a set of brain MR images obtained from healthy subjects and MS patients during different experimental iterations, an average accuracy of 97% was obtained. The results of the method were estimated under specific conditions, and finally the proposed model yielded desirable outputs compared to similar methods.

Keywords Multiple sclerosis · Extreme learning machine · Shuffled frog-leaping algorithm · Differential evolution · Fractal descriptor

1 Introduction

Multiple Sclerosis is an inflammatory disease in which myelin sheaths of the nerve cells in the brain and spinal cord are damaged [1–6]. This damage can disrupt the ability of parts of the nervous system which are responsible for communication and causes many signs and symptoms such as physical problems [7]. Regional estimates suggest that this disease has a moderate prevalence in countries and is in prevalence range of the European and Far East countries [8]. Symptoms of MS disease appear in several forms and its new symptoms occur either as step recurrence -multiple disease reversal or alternately over time [9–11]. Initially, recovery from attacks is almost complete, but slowly, neuropsychiatric disabilities with different degrees will remain from each attack [12]. MRI is the most practical method for detecting the masses left in the brain, which can greatly help the specialist. Sometimes there are different interpretations of MR brain images and so it is necessary to have an accurate statistical analysis of

✉ Khosro Rezaee, Kh.rezaee@meybod.ac.ir | [1]Department of Biomedical Engineering, Islamic Azad University, Tabriz Branch, Tabriz, Iran. [2]Department of Biomedical Engineering, Meybod University, Meybod, Yazd, Iran. [3]Department of Biomedical Engineering, Hakim Sabzevari University, Sabzevar, Iran. [4]Department of Electrical Engineering, University of North Dakota, Grand Forks, ND 58202-7165, USA.

alternating signals. Sometimes due to the inherent nature of MRI, it is not possible to distinguish lesions caused by diseases such as Alzheimer's, MS, and other common diseases in the brain. If it is possible to detect lesions from this disease at early stages, then it will be possible to treat the disease.

The discovery of lesions by the neurologist requires expedition, time and high accuracy and typically, with increasing number of MR images of the brain, the efficiency required for the diagnosis will be significantly reduced. The inadequacies of contrast and image clarity, as well as the similarity of lesions caused by the disease to other brain tissues, lead to different interpretations of brain MRI images; therefore, a precise analysis of the resulting images is required. Thus, employing a robust automated approach can provide satisfactory outcomes, whereas additionally dominant the time needed for diagnosis. The major contributions of MR processing work lie in applying the feature extraction and classification to detect the lesions and non-lesions as shown in Fig. 1.

Several strategies have been proposed to date in order to automate the intelligent brain disease diagnosis, especially MS. Among the studies in the field of diagnosis, separation and classification of lesions caused by MS disease, Ballin et al. [13] mapped a three-dimensional lesion of the MS onto the MR image in addition to conduct classification by combination methods. Zhang et al. [14] also used tissue analysis in order to diagnose MS by extracting the features and then selecting the best ones in brain MR images. The best feature analysis had been used by linear methods and statistical analysis and the diagnostic accuracy of this disease in MR images can reach up to 88% by choosing the intelligent tissue analysis method.

Roy et al. [15] provided an automated method using tissue features and support vector machines, in which the brain lesion separation process was performed and then the classification occurred. The strong point of their work was to use the cumulative distribution function and also normalization of the area of interest to improve the image quality. The texture features, local brightness, and initial spatial information were used as the main features and in fact, the classification process occurred for image pixels, not for the image itself.

Elliott et al. [16] proposed the consistent probabilistic detection for new MS lesions in MR images. They used a series of sequential MR scans. Their work was a two-step classification; first, they used a classifier to find out the possibility of a voxel for being lesion. Cabezas et al. [17] suggested conducting the classification mechanism through the boosting classification technique in order to classify the classification for a set of features. They prepared MR images from 45 subjects during the three recording periods. This set of techniques has led to a typical, precise separation and as a result, parts of the MR image with the potential of disease are revealed.

Sweeney et al. [18] compared different methods of machine learning to build the best feature vector and they did this using a multi-quality structure in MR images. This resulted in a fairly precise separation of the images and the separation of lesions caused by MS disease. Ardakani et al. [19] investigated MS diagnosis through tissue analysis. They extracted features from the images for 50 patients through the analysis of MR images, initially by analyzing the main components. Then, they perform classification by relying on the linear separation method and obtained the sensitivity of 100% and the surface below the ROC curve had the value of 1 in their work. Separating a favorite area and classifying the three classes was one of the innovations of their work.

Liu et al. [20] diagnosed MS by removing confounding factors that did not include information on the lesion area and they did this using limited clustering. Weygandt et al. [21] relied on the pursuit of biomarkers based on segmentations such as thresholding and combining atlas images and brain MR to find the location of brain lesions based on which, the early stages of the formation of these lesions were possible from the images. Karimaghaloo et al. [22] proposed a system based on randomized conditional stratification and classification to separate the areas suspected to MS-related lesions, which distinguished small pieces which were considered as lesion.

Brosch [23] used three-dimensional encoded deep convolutional neural networks to detect lesions caused by MS disease. Zhang et al. [24] relied on methods such as static wavelet and entropy, and statistical characteristics. They suggested using the decision tree, k-NN classifier, and
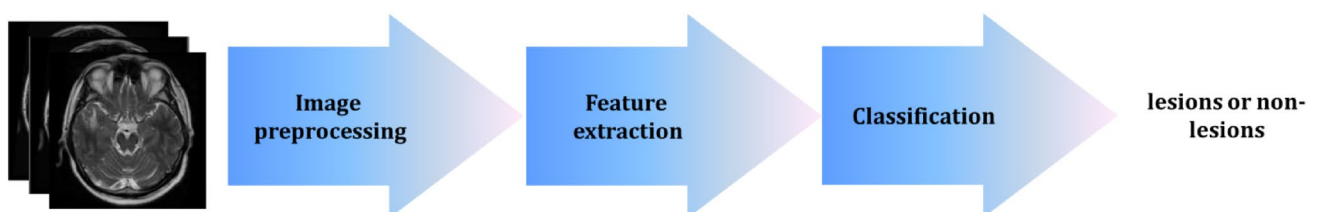

Fig. 1 Processing steps utilized in the lesions and non-lesions categorization

supporting vector machines for the classification process. They then achieved the accuracy of 97.9% for their data sets, and sensitivity and specificity values were higher than 95%. The use of two-dimensional separation and 3D mapping and final detection has been proposed through deep convolutional neural networks [25].

Recently, Gheshlaghi et al. [26] proposed a super-pixel segmentation based technique for multiple sclerosis lesion detection. Their study includes SVM as an effective classifier with polynomial kernels that is applied to have better performance in distinguishing specified decision classes. Also, they use discrete wavelet decomposition (DWT) which extracts local features from analyzing MR images.

Improving automated multiple sclerosis lesion segmentation with a cascaded 3D convolutional neural network approach have been conducted by Valverde et al. [27]. They propose an automated White Matter (WM) lesion segmentation method for MS patient images and their approach relies on a cascade of two 7-layer convolutional neural networks. The comparison of the previous approaches over the last few years has been shown in Table 1.

In this study, different lesions of the MR images are segmented by using the clustering method, and then, features of segmented areas will be extracted. There are some objectives in the research, including the following: (1) Detection of MS lesions from magnetic resonance images in two classes of no disease and disease incidence; (2) Promoting the feature vector structure by aggregating results from descriptors such as Fractal and PZM. Additionally, we focus to demonstrate the following hypotheses: (1) Compatible extracted features based on efficient descriptors decrease the time and space complexity for building the prediction model. Also, feature subset selection by an evolutionary-differential algorithm improves the prediction accuracy of the ELM classifier and reduces the false prediction ratio. (2) Appropriate ELM configuration parameters, such as the number of neurons in hidden layer and the training error optimized by the memetic meta-heuristic SFLA algorithm. The performance of proposed methods is analyzed with the focus on dimensionality reduction of extracted features and improvement of classification accuracy using MR images.

The rest of this paper is organized as follows: Sect. 2 describes the proposed method with the different sections of the algorithm. Section 3, considers the implementation of the algorithm and the experimental results. Finally, the conclusion of the paper will be presented in Sect. 4.

2 Materials and methods

Separation in image processing is an approach which is conceptually used in the image segmentation process. We first described the MRI data used for the evaluation of the proposed segmentation and classification approach. Then, the different steps of the proposed classification process were detailed. The trend of the proposed stages could be summarized according to the following Fig. 2.

2.1 Pre-processing

In the preprocessing step, two steps are implemented on the basis of slices received from a person, which are Histogram equalization and segmentation by csFCM.

2.1.1 Histogram equalization

Retinex multi-scale transformation method was used for histogram equalization, which is one of the optimization methods, which also has acceptable results in the preprocessing. In the desired transform, first, the image brightness pattern is calculated based on the weighted neighboring mean in the specified radius for each pixel. The Retinex multi-scale transformation is expressed according to Eq. (1) [37]:

$$L_{(i_{cent}, j_{cent})} = \frac{\sum i, j_{cneighboring\ of\ cent}\ I(i,j).W(i,j)}{\sum i, j_{cneighboring\ of\ cent}\ W(i,j)} \tag{1}$$

where I is the input image received from MRI devices and W is the weight of the matrix. Neighborhood weights (NW) are calculated for each pixel (i_{cent}, j_{cent}) based on the neighborhood of pixels and in accordance with Eq. (2) [37]:

$$NW_{i,j} = e^{-(Dist_{i,j}/\ Radius)^2}$$
$$st. \quad Dist_{i,j} = \left[(i_{cent} - i)^2 + (j_{cent} - j)^2\right]^{0.5} \tag{2}$$

The cent refers to the pixel coordinates where we want to calculate their brightness and the Radius term is the default radius for the neighborhood, where different values can be adopted. Finally, the final image is calculated by difference of the brightness pattern from the original image in the logarithmic range. In other words, R and L

Table 1 The comparison of the previous approaches

Authors	Hybrid algorithm	Dataset	Performance	Advantage	Disadvantage
Souplet et al. [28]	Image enhancement and segmentation lesions	Laboratory database at the University of Cyprus	Accuracy ≥ 80%	Is useful in MRI lesions segmentation	The method is sensitive to noise and lesion type-low accuracy
Schmidt et al. [29]	Segmentation based on lesions	Omages collected from the medical clinic	Accuracy ≥ 82%	A tool for detection of FLAIR-hyperintense white-matter lesions	The method is sensitive to noise and lesion type-low accuracy
Kwok et al. [30]	A model based fusion of features and SVM	MedPix database and their collected MR images	Accuracy ≥ 61%	Suitable in detection	Low accuracy and variation of their algorithm is high-quality of image must be high
Cerasaet al. [31]	A model based cellular neural network to segment lesions	Collected MR images	Accuracy ≥ 83%	Suitable in detection	The method is sensitive to noise and lesion type-low accuracy
Cabezas et al. [32]	Lesion segmentation based on cascaded 3D CNNs	Laboratory database at the University of Cyprus	Accuracy >= 87%	They presented a deep 3D CNNs to segment- non-sensitive to image with different quality	Uncertainty has not been analyzed -low accuracy
Brosch et al. [23]	Deep 3D convolutional encoder networks	Collected MR images	Accuracy ≥ 87%	They presented a deep 3D CNNs-non-sensitive to image with different quality	Uncertainty has not been analyzed -low accuracy
Kawahara et al. [33]	Morphological and appearance features	Database of clinical Avaris image	Accuracy ≥ 88%	The model is able to predict the physical disability from MR images	Uncertainty has not been analyzed -low accuracy- Missing some image information
Siddiqui et al. [34]	The used discrete wavelet transform (DWT), principal component analysis (PCA), and least squares support vector machine (LS-SVM)	MedPix database and Collected MR images	Accuracy ≥ 95%	Enhancing MR images and combining hybrid model-good accuracy	Uncertainty has not been analyzed—LSSVM isn't optimum model
Deshpande et al. [35]	Dictionary learning and PCA method	Collected MR images	Accuracy ≥ 95%	They proposed a novel supervised approach using sparse representations and dictionary learning	Uncertainty has not been analyzed—limitation of PCA to handle only linear data
Yoo et al. [25]	Unsupervised deep learning and random forest applied to myelin and T1w images	Collected MR images	Accuracy ≥ 88%	Deep learning of myelin maps has not been previously attempted	Uncertainty has not been analyzed—low accuracy
Alshayeji et al. [36]	Morphological operations, and a multilayer feed-forward neural network	Collected MR images	Accuracy ≥ 98%	Their method provided a significant recognition rate of up to 98.5% and high dice coefficient (DC) value (0.71 ± 0.18)	Uncertainty has not been analyzed—the method is sensitive to noise and lesion type

Fig. 2 Proposed model steps for brain lesion and non-lesion classification
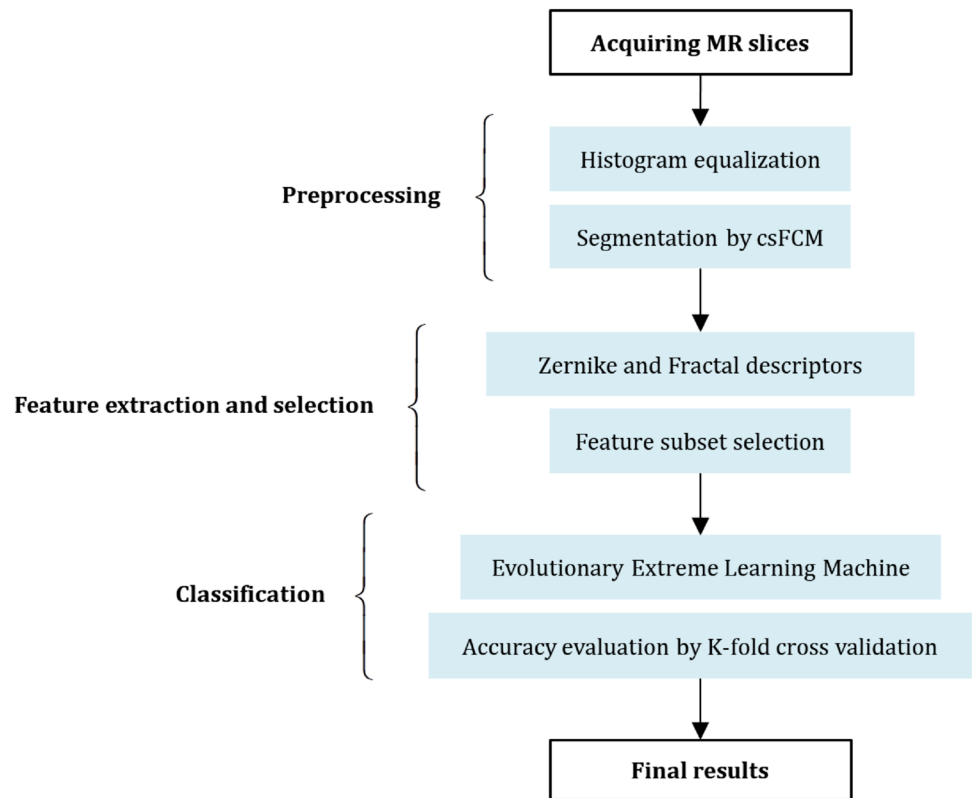
are respectively the components of the reflection and the component of the image brightness:

$$Log(R) = Log(I) - Log(L) \qquad (3)$$

2.1.2 csFCM

The conditional spatial fuzzy C-means (csFCM) method is derived from the clustering technique followed by the segmentation of magnetic resonance images [38] which can be used for noise images and provides good results. The process of the csFCM algorithm has been displayed in the Region of Interest (ROI) segmentation in Algorithm 1.

In the set of equations of Algorithm 1, x_k is the equivalent of the pixel k of the magnetic resonance image. The conditional spatial FCM algorithm with parameter p and q is denoted as $csFCM_{p,q}$. Again, it may be noted that $csFCM_{1,0}$ is identical to the conventional FCM algorithm.

2.2 Feature extraction and selection

In the process of extracting and selecting features from the pre-processing images, the feature vector is constructed from the combination of features extracted by the Pseudo-Zernike moments and fractal algorithms and, then the modified DE algorithm reduces the extracted vector dimension.

2.2.1 Pseudo-Zernike moments (PZm)

Pseudo-Zernike moments (PZMs) are a set of complex polynomials which form an orthogonal set within a unit circle ($x^2 + y^2 = 1$) which is represented by $V_{nm}(x, y)$ and its structure is defined as PZMs Eq. (4) [39]:

$$V_{nm}(x, y) = V(\rho, \theta) = R_{nm}(\rho)e^{im\theta} \qquad (4)$$

Algorithm 1: The conditional spatial fuzzy C-means (csFCM) in ROI segmentation of slices

Input: set values for the number of FCM clusters C: the degree of fuzziness m, p, and $q=2$, and the error ε.

Output: an image with clustered area in which the ROI is segmented based on selected clusters.

Begin

1.   Initialize randomly the centers of clusters $v_i^{(0)}$,
2.   Initialize randomly the centers of joint clusters $w_i^{(0)}$
3.   Set j=1
4.   **while** stopping criterion not met **do**
5.   j = j+1

        Calculate the membership value $\mathbf{U^{(j)}}$ using the cluster centers $\mathbf{v_i^{(j-1)}}$ as follows:

6.
$$\mu_{ik} = ||x_k - v_i||^{2/1-m} \sum_{c=1}^{C} ||x_k - v_c||^{2/m-1}$$

        Calculate the conditional spatial membership value $\mathbf{u_{ik}^{(j)}}$ using the centers $\mathbf{v_i^{(j-1)}}$ as follows:

7.
$$\mu_{ik} = f_{ik} \times ||x_k - v_i||^{2/1-m} \sum_{c=1}^{C} ||x_k - v_c||^{2/m-1}$$

        $Where\ f_{ik}\ is\ f_{ik} = M^{-1} \times \sum_{j \in N(x_k)}^{1} \mu_{ik}$

        Calculate weighted membership value $\mathbf{z_{ik}^{(j)}}$ as follows:

8.
$$z_{ik} = \mu_{ik}^p \times u_{ik}^q \times (\sum_{c=1}^{C} \mu_{ck}^p \times u_{ck}^q)^{-1}$$

        Calculate weighted membership value $\mathbf{z_{ik}^{(j)}}$ as follows:

9.
$$z_{ik} = \mu_{ik}^p \times u_{ik}^q \times (\sum_{c=1}^{C} \mu_{ck}^p \times u_{ck}^q)^{-1}$$

        Update joint cluster value $w_i^{(j)}$ as follows:

10.
$$w_i = (\sum_{k=1}^{N} z_{ik}^m x_k) \times (\sum_{k=1}^{N} z_{ik}^m)^{-1}$$

        Update centers $\mathbf{v_i^{(j)}}$ as follows:

11.
$$v_i = (\sum_{k=1}^{N} \mu_{ik}^m x_k) \times (\sum_{k=1}^{N} \mu_{ik}^m)^{-1}$$

12.   **end while**: until $||w_{ik}^p - w_{ik}^p||^1 < \varepsilon$
13.   Return the cluster center $\mathbf{w_i}$ and the membership value $\mathbf{z_{ik}}$; i = 1,2,...,C; k = 1,2,..., N.

end

Where $j = (-1)^{0.5}$, $\theta = \tan^{-1}(yx^{-1})$, $|\rho| \leq 1$, $n \geq 0$, $m \leq n$, $n - |m| = even$, and R is a radial polynomial and it should also be noted

that ρ is the source vector length to the point (x, y) and θ is the angle between the vector ρ and the axis x in the direction of counterclockwise. In the above relation n is a non-negative integer that shows the polynomial order. M is a negative or positive integer that shows the order of the horizontal arc repetition and its absolute value is always smaller or equal to n, and the difference m of n is always an even value. In the above relation, R_{nm} are radial polynomials calculated in accordance with relationships (5) and (6) [39]:

$$R_{nm}(x,y) = \left[\sum_{s=0}^{0.5 \times (n-|m|)} S_{n,|m|,s}(x^2 + y^2) \right]^{n-2s} \quad (5)$$

$$S_{n,|m|,s} = (-1)^s \frac{(n-s)!}{s! \left(\frac{n+|m|}{2} - s \right)! \left(\frac{n-|m|}{2} - s \right)!} \quad (6)$$

Zernike moments (ZMs) are a mapping of the image into a set of Zernike complex polynomials. One of the important features of Zernike moments is their orthogonally; for this reason, image features can be displayed without any redundancy in information or overlap between moments. If the rotation is made in the images, it will not have any effect on its Zernike moments. The complex Zernike moments of the order n and with the repetition of m are calculated using Eq. (7) [39]:

$$ZM_{nm} = \frac{n+1}{\pi} \sum_x \sum_y f(x,y) V^*_{nm}(x,y) \tag{7}$$

where, $f(x,y)$ is the brightness intensity function of a digital image at the location of x and y and the $*$ sign also denotes the complex conjugate.

2.2.2 Fractal

A fractal descriptor, known as one of the tissue descriptors of the image, is implemented in two stages based on segmentation [40–42]; first, the gray level image is divided into a group of images with binary values, and this process is performed by the bi-threshold binary decomposition algorithm. The bi-threshold binary decomposition algorithm considers the image $I(x,y)$ as an input and retrieves binary images. By bi-threshold binary decomposition, a pair of thresholds is selected

from T as a result; the segmentation is carried out in accordance Eq. (8) [40]:

$$I_b(x,y) = \begin{cases} 1 & \text{if } t_l < I(x,y) \le t_u \\ o & \text{Otherwise} \end{cases} \tag{8}$$

where t_l and t_u respectively represent the minimum and maximum threshold values. By applying Eq. (8), the binary images are decomposed and then proceeded from $T \cup \{n_l\}$. On the one hand, all threshold pairs of $\{t, n_l\}$, $t \in T$ can be achieved in which, n_l represent the maximum possible gray level in $I(x,y)$. So the number of binary images is $2n_t$ and the value of n_t is set to 8 value. After applying the bi-threshold binary decomposition, the fractal tissue analysis feature vector is constructed according to the segmentation based on the size of the binary images, the mean gray level and the fractal dimensions of the boundaries. The boundaries of the image regions $I_b(x,y)$ are selected as marginal segments, which are represented by $\Delta(x,y)$ and calculated as Eq. (9) [40]:

Fig. 3 The implementation stages of an algorithm for extracting features from MRI slices
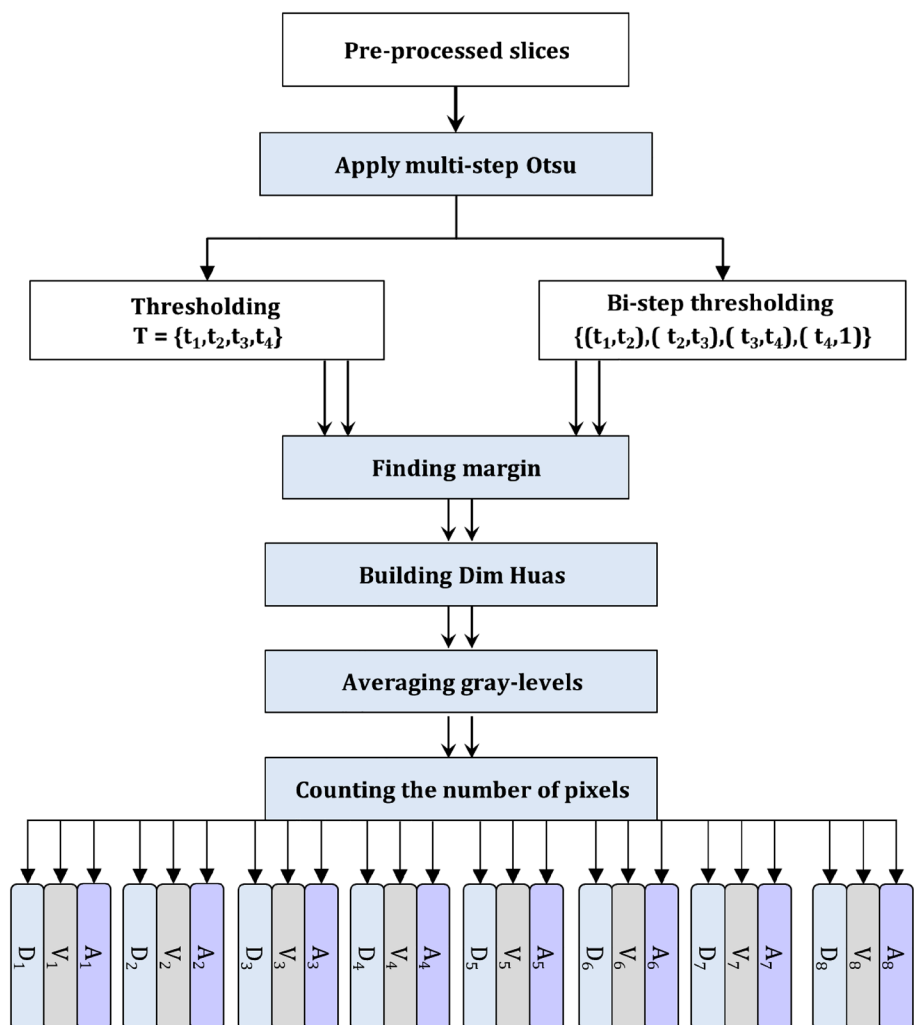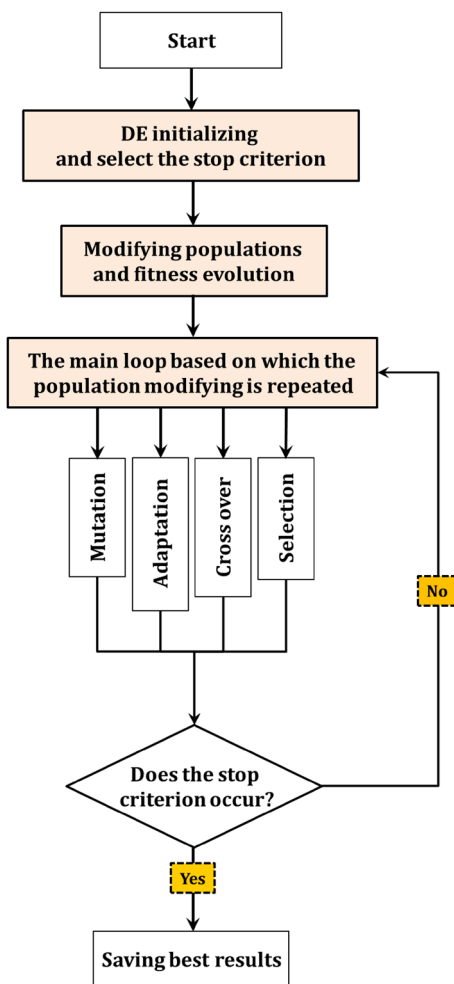
Fig. 4 Steps of the DE algorithm

$$\Delta(x,y) = \begin{cases} & if\ \exists\ (x,y) \in N_B[(x,y)] \\ 1 & I_b(x,y) = 0\ \Lambda \\ & I_b(x,y) = 0 \\ o & Otherwise \end{cases} \qquad (9)$$

where $N_B[(x,y)]$] represents a group of pixels with 8 connections to (x,y). Also, $\Delta(x,y)$ will have a value of 1 if the pixel is at the position (x,y), which refers to the binary image I_b and, if there is at least one neighborhood with a pixel with zero has the sum of 1. Figure 3 indicates the implementation stages of fractal algorithm. In this figure, area $(A_1, A_2, \ldots, A_n)$ and mean gray level $(V_1, V_2, \ldots, V_n)$ features are computed directly from the binary image. Fractal dimension $(D_1, D_2, \ldots, D_n)$ features is computed from the border or margin image. In these dimension, we select n equal 8 because this would result in a lower error rate.

2.2.3 Differential Evolution (DE) algorithm

We conduct the feature selection with a modified Differential Evolution (DE) algorithm. In the feature selection problem, the most effective subset features is chosen which has a better solution. The general evolutionary-differential algorithm has been shown in Fig. 4.

In order to increase the speed and at the same time maintain the heuristic nature of the algorithm and prevent early convergence, in both Cross Over and Mutation vectors, we have established a balance between local and global search, where random and uniform numerical definitions in the range of zero to one can be used. Additionally, the use of additional parameters in the control of the algorithm is eliminated and therefore, adjusting the DE algorithm in the feature selection is greatly optimized and finally we do not eliminate the automatic and random mode of the algorithm. In the proposed DE algorithm, the mutation vector controls the mutation effect in adaptive form using the feedback and thus, a kind of self-adjustment will increase the accuracy and speed of the algorithm. In other words, when the average fitness approaches the best fitness, it indicates that the algorithm reaches the best final value. At this time, the effect of the mutation should be reduced so that the algorithm does not go away from the final solution. Conversely, when the difference between the average fitness and the best fitness is high, it is necessary to map the problem space algorithm to increase the effect of the mutation.

An ELM based on wavelet kernel is used to calculate the feature selection error and in the fitness function, the initial train data is divided into new train data and validation data with the K-fold cross validation method, with K equal to 5 and, the optimal subset is selected after a limited number of repetitions. Therefore, the error resulting from the classification of the validation data along with the minimum number of features will be the output of the cost function.

2.3 Evolutionary extreme learning machine

The Extreme Learning Machine (ELM) is a configuration with a hidden layer where the weights between the first layer and the random hidden layer are updated and, the weights of its second layer are weighted normally similar to the performance of the nervous system. ELM outputs with m neurons and the activation function f can be displayed as follows:

$$o_j = \sum_{i=1}^{m} \beta_i f(l_i x_r + b_i) \qquad (10)$$

The algorithm is fast but may offer a better overall performance [43–45]. Learning error and output weights should

be minimized simultaneously to reduce learning errors in the ELM algorithm. Therefore, the overall performance of the neural networks will increase:

$$Min \ \ \|AS - C\|,$$
$$\|S\| \tag{11}$$

where, it can be written that:

$$S = A^T(E^{-1} + AA^T)^{-1}C \tag{12}$$

where, E is the adjustment coefficient, A is the output matrix of the hidden layer, and C is the expected output matrix of the samples. Therefore, the output function of the ELM algorithm can be written as Eq. (13):

$$u(r) = v(r)A^T\left(\frac{1}{E} + AA^T\right)^{-1}C \tag{13}$$

If the feature vector $v(r)$ is unspecified, the ELM kernel matrix can be rewritten based on the Mercer conditions:

$$D = AA^T \ : \ k_{jz} = v(r_j)v(r_z) = x(r_j, r_z) \tag{14}$$

where, $u(r)$ is the output function of the wavelet kernel for ELM, which can be represented as Eq. (15):

$$u_r = [b(r, r_1), \dots, b(r, r_M)]\left(\frac{1}{E} + D\right)^{-1}C \tag{15}$$

where, $D = AA^T$ and $b(r, g)$ is the kernel function ELM. In this case, there are some kernel functions including linear kernel, Polynomial kernel, Gaussian kernel, and Exponential kernel. But the Wavelet kernel function is beneficial for simulations and performance [14]:

$$b(r, g) = \cos\left(w \times \|r - g\| \times x^{-1}\right) \ \exp(-\|r - g\|^2 \times y^{-1}) \tag{16}$$

Determining these parameters will greatly affect the desired accuracy. To find the best structure of ELM, we suggest that the Shuffled Frog-Leaping Algorithm (SFLA) algorithm perform the optimization. The steps of SFLA to find best parameters of wavelet kernel in ELM classifier are as follows:

A.   Global search

In SFLA, the combination increases the quality of the memetics affected by different subgroups. Global and Local searches continue to meet the convergence condition. The balance between global transfer and local search allows the algorithm to leap easily from the local minimum and extend until reaching optimization. One of the features of the SFLA algorithm is fast convergence. Global search to find best parameters is conducted based on Algorithm 2.

Algorithm 2. Global search by SFLA

Step 1)	Initialization: select m and n, m is the number of memeplexes, and n is the number of frogs per memeplex, thus the total population in the pool is obtained through the relation $F = m.n$.
Step 2)	Virtual population generation: F virtual frogs ($U(1)$), $U(2),\dots, U(F)$) are sampled from the possible space. The value of fitness $f(i)$ of each frog $U(i)$ is calculated for each $U(i) = (U1i, U2i,\dots, Udi)$ in which d is the number of decision variables.
Step 3)	Frog Ranking: Frogs are sorted according to their fitness in descending order and stored in the array $X = \{U(i), f(i), i = 1,\dots, F\}$. The best frog position of P_X is selected in the entire population ($Px = U(1)$)
Step 4)	Partitioning frogs into mempelexes: The array X is divided into m memeplex ($Y1, Y2,\dots, Ym$), each containing n frogs.
Step 5)	Memetic evolution in each memeplex: each memeplex ($Y_k, k = 1,\dots, m$) evolves by local search.
Step 6)	Shuffling memeplexes: After a certain number of memetic evolutions are completed in each memeplex, memeplexes ($Y_1,\dots, Y_m$) are placed in X so that the relation $X = (Y_k, k = 1,\dots, m)$ is established. The best frog position in the P_X population is updated. If convergence conditions are met, it stops. Otherwise it goes to the fourth stage of global search.

B.   Local search

In the fifth step of global search, the evolution of each memeplex is performed for n times independently. After the memeplex evolved, the algorithm returns to the global search for the shuffling. The details of local search are described in each memeplex. Weights are assigned with triangular probability distribution according to Eq. (17).

$$p_j = \frac{2(n + 1 - j)}{n(n + 1)} \ \ j = 1, \dots, n \tag{17}$$

where, j is the index of j-th member and n is the number of elements. To construct submemeplex array, q frog is randomly selected from each n frogs in each memeplex. The frogs in the submemeplex are arranged in descending order in terms of their fitness. The positions of the best frog and the worst frog in the submemeplex are identified by P_B and P_w, respectively. The new position of the worst frog in the submemeplex that is the frog with the worst performance and then is calculated through Eq. (18). Also, S is the step size or mutation rate of the frog and is calculated as Eq. (19):

$$U(q) = P_w + S \tag{18}$$

$$S = max\left[round\left(rand\left(P_B - P_w\right)\right), Smax\right] \quad (19)$$

Also, we can conduct the local search based on Algorithm 3.

Algorithm 3. Local search by SFLA

Step 1)	Initialization: Parameters i_m and i_n get the initial value of zero. In this concept, i_m counts the number of memeplexs and in counts the number of evolution stages.
Step 2)	Generating submemeplexes: The goal of the frogs is to move optimal positions by improving their memeplexes. The submemeplex selection method is to allocate higher weights to frogs with higher performance and lower weights to the frogs with lower performance. Weights are assigned with triangular probability distribution according to Eq. (17).
Step 3)	Correcting the position of the worst frog based on Eq (18). If the new position is better than the previous position, then replace the new $U(q)$ with the previous $U(q)$ and goes to *step 6* of the local search. Otherwise, go to *step 4* of the local search.
Step 4)	S or step size of the frog and is calculated by Eq. (19). If U(q) is within the possible space, the new performance value $f(q)$ is calculated. If the new $f(q)$ is better than the previous one, then $U(q)$ replaces the previous $U(q)$ and goes to *step 6* of the local search. Otherwise, it goes to *step 5* of the local search.
Step 5)	Calculating step size by P_x: If better result is not obtained in the *step 3*, then after calculating the step size of the frog, the new position $U(q)$ is calculated by (19).
Step 6)	Censorship: If the new position is not in the possible area or not better than the previous position, a new *frog* (r) will randomly be generated in a possible position and replaces by a frog whose new position is not suitable for progress. $f(r)$ is calculated and $U(q)$ is set to r and $f(q)$ equal to $f(r)$).

Step 7)	Upgrading memeplex: After changing the memetic of the worst frog in the submemeplex, the frogs in Z are placed in their original position on Y_{Im}. Y_{Im} is sorted by descending order in terms of performance. If $I_n < n$, it goes to step 3 of the local search. If $I_m < m$, go to *step 1* of the local search. Otherwise, it will return to global search for shuffling the memeplexes.

3 Experimental results

3.1 Data

The main data, including slices selected by a specialist – 10 to 15 important slices have been separated from each person from a variety of people with different levels of MS disease which was collected during the course of a year and a half from the Vasei Hospital in Sabzevar, Iran from 2016 to 2018. The sample of MRI selected from MS patients has been shown in Fig. 5. Healthy individuals are also included in this research to be considered as control or study variables. Patients are divided into two groups with and without the history of pain inflammation in the head region and various magnetic resonance imaging centers provided reports on imaging. People with MS disease with different stages of the disease have long been monitored by a doctor and the definitive symptoms of the disease have been listed. Sampling was done randomly according to the opinion of the specialist. The personal characteristics of the patients after imaging were recorded in the questionnaires or along with two DICOMDIR and Radiant setup software and complete neurological and visual examination has been performed on patients. Data gathering research was a two-group cross-sectional analytical study.

The sample of the study including 125 subjects (at least 10 slices according to physician diagnosis) in patient
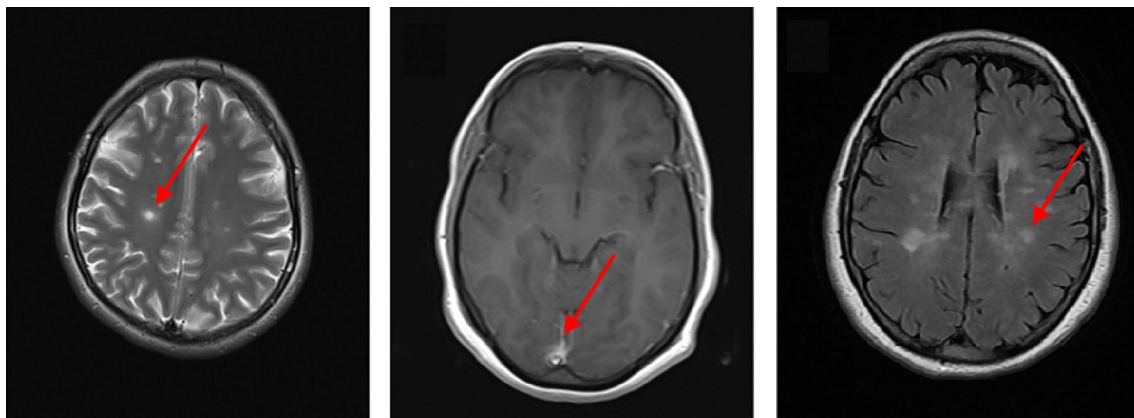

Fig. 5 The two physicians have definitely commented on the MS lesion and even estimated its possible location in our MR images

subjects, definitive diagnosis of disease and age below 50 years old referring to the MS clinic, neurology ward and, patients referred to the magnetic resonance imaging center of Vasei Hospital in Sabzevar, in two groups 64 people and 61 people were examined. 64 patients were diagnosed with MS and 61 subjects were healthy without any complications. The two experts have definitely commented on the disease and even estimated its possible location. *FLAIR* images and often T_1 and T_2 modes have been studied.

3.1.1 Evaluations

Three factors of accuracy, sensitivity and specificity, which have been introduced to measure the diagnosis accuracy of the proposed system, are calculated in accordance with Eqs. (21) to (23) and used to evaluate the system:

$$Accuracy = \left(\frac{N_{TP} + N_{TN}}{N_{TP} + N_{FN} + N_{TN} + N_{FP}} \right) \qquad (20)$$

$$Sensitivity = \left(\frac{N_{TP}}{N_{TP} + N_{FN}} \right) \qquad (21)$$

$$Specificity = \left(\frac{N_{TN}}{N_{TN} + N_{FP}} \right) \qquad (22)$$

where N_{TP} is the number of brain MRI images containing MS-induced lesion, and the proposed algorithm has correctly diagnosed the disease. In addition, N_{TN} indicates the number of brain MRI images not induced by MS lesions and the proposed algorithm has correctly detected the absence of disease. N_{FP} is the number of brain MRI images that did not contain any MS-induced lesions, but the proposed algorithm had erroneously diagnosed the disease. Finally, N_{FN} indicates the number of brain MRI images that contained MS-caused lesion, but the proposed algorithm had erroneously ruled out the disease.

3.2 Model initializing

Because the number of areas suspected to MS lesion may vary in the preprocessing stage that is, cases that resemble lesions resulting from Alzheimer's disease, tumor or pathologic damage, or masses other than the target class, the experiment was repeated for 10 times. Thus, all images with an initial dimension after the pre-processing are imported to the optimized decision maker network as input.

The parameters of the DE algorithm in feature selection are based on the cost function due to the classification using the ELM (wavelet kernel) neural network. Table 2 represents the parameters of the DE algorithm in order to select the best features.

3.3 Results

The proposed algorithm was implemented in Matlab and the required system for data processing included Due CPU 2.6 GHz with a RAM of 4 GB and performance of the algorithm was between 10 and 20 s in short term. The change in accuracy is due to the resizing of the image that, in the current research it has been tried to consider three different models of image resizing. As mentioned, the first-stage benchmarks have been represented by the K-fold validation method in Tables 3, 4 and 5 for dimensional change. Each table consists of train and test stage, and accuracy varies between 90% and 98% due to the changes in dimensions. Also Figs. 6, 7, 8 and 9 indicate an error reduction and convergence due to the feature selection and kernel parameters regularization from all features as one-third of the total features of the model.

4 Discussion

Experiments were carried out in the feature selection phase, in which the appropriate subset selection was made based on which selection of subsets including 10%,

Table 2 Initialization for an adaptive DE algorithm in feature selection

Parameter	Parameter description	Initialization
Inp_{Num}	Total number of decision variables	Equivalent to the total number of features (48) formed of features extracted from fractal descriptors and PZm
$Selected_{Feat}$	Number of selected features	Usually, within a quarter to two-thirds of the total features, the variable is discrete (12–32)
Max_{Iter}	Maximum repetition for search	10, 20 and 30
W_{Start}	Initial internal weight	0.95
W_{End}	Final internal weight	0.35
Err	Error due to classification in cost function	A continuous number in the range of zero to one

Table 3 The assessment of benchmarks of MS recognition model without feature selection and kernel parameters optimization for different image dimensions. The bold values are best accuracies

	K	32×32 dimension			64×64 dimension			128×128 dimension			256×256 dimension		
		Accuracy	Specificity	Sensitivity	Accuracy	Specificity	Sensitivity	Accuracy	Specificity	Sensitivity	Accuracy	Specificity	Sensitivity
Testing level	1	70.32	78.53	66.91	73.37	81.95	69.03	76.92	83.33	71.42	77.11	85.46	70.21
	2	87.56	78.06	89.45	90.22	81.28	93.10	92.85	83.33	94.54	94.23	85.09	97.93
	3	82.60	74.47	90.17	86.65	79.63	92.00	90.66	80.00	96.54	89.20	81.39	94.48
	4	75.86	59.72	83.63	77.09	62.59	85.81	78.57	66.66	87.50	80.01	67.61	89.32
	5	78.69	80.68	81.24	80.38	81.04	84.17	85.71	83.33	87.50	87.29	81.13	90.19
	6	80.16	80.83	82.26	84.21	82.80	85.26	87.13	83.23	88.17	87.29	85.74	91.53
	7	72.27	79.36	70.41	74.40	81.45	72.43	78.57	83.33	75.00	74.89	82.90	73.07
	8	69.65	65.42	80.68	74.86	69.26	83.56	76.92	66.66	85.71	77.11	69.11	87.23
	9	83.18	79.31	91.57	88.32	81.37	93.14	92.85	83.33	95.65	96.32	87.57	98.44
	10	84.31	88.39	79.08	90.03	91.83	84.39	92.85	98.17	87.50	94.03	97.50	85.03
Average	–	78.46	76.47	81.54	81.95	79.32	84.28	84.93	78.20	86.95	**85.74**	**82.35**	**87.73**

Table 4 The assessment of benchmarks of MS recognition model with feature selection for different image dimensions. The bold values are best accuracies

	K	32×32 dimension			64×64 dimension			128×128 dimension			256×256 dimension		
		Accuracy	Specificity	Sensitivity	Accuracy	Specificity	Sensitivity	Accuracy	Specificity	Sensitivity	Accuracy	Specificity	Sensitivity
Testing level	1	93.30	100	87.50	100	100	100	92.31	83.33	100	90.15	80.81	98.80
	2	93.30	100	87.50	85.71	34.62	85.71	100	100	100	98.11	99.37	95.25
	3	85.71	100	75.00	92.86	100	85.71	85.71	100	75.00	83.65	97.26	77.37
	4	84.61	100	66.67	92.86	100	85.71	87.61	72.48	100	85.21	70.73	97.15
	5	75.00	80.00	71.42	91.67	27.07	100	92.58	100	87.50	91.35	97.56	85.46
	6	92.30	100	85.71	92.86	29.02	100	87.13	83.23	88.17	88.43	81.91	84.77
	7	87.13	83.23	88.17	92.23	28.58	100	100	100	100	96.78	97.30	94.08
	8	92.85	100	87.50	85.71	24.28	85.71	87.61	72.48	100	84.09	70.51	97.91
	9	84.61	100	75.00	84.62	27.17	83.33	100	100	100	97.84	96.19	100
	10	91.66	80.00	100	84.62	26.11	83.33	92.31	83.33	100	90.55	81.29	98.44
Average	–	88.04	94.32	82.44	90.32	89.94	91.37	**91.70**	**95.00**	**90.37**	90.61	87.28	92.92

Table 5 The assessment of benchmarks of MS recognition model with feature selection and kernel parameters optimization for different image dimensions. The bold values are best accuracies

Testing level	K	32×32 dimension			64×64 dimension			128×128 dimension			256×256 dimension		
		Accuracy	Specificity	Sensitivity	Accuracy	Specificity	Sensitivity	Accuracy	Specificity	Sensitivity	Accuracy	Specificity	Sensitivity
	1	92.62	96.22	89.85	94.26	94.83	93.75	92.62	90.56	94.21	95.86	94.33	97.05
	2	94.21	96.22	92.64	95.80	98.11	93.94	100	100	100	98.34	98.11	98.52
	3	85.95	90.56	82.35	87.80	89.47	86.36	95.86	94.33	97.05	96.74	96.29	97.10
	4	97.54	98.11	97.19	95.90	94.55	97.01	100	100	100	97.52	98.11	97.05
	5	95.93	100	92.73	95.93	96.55	95.38	95.86	94.33	97.05	94.21	94.33	94.11
	6	86.79	86.79	79.71	86.89	89.29	84.85	100	100	100	90.08	92.45	88.23
	7	90.74	90.74	82.35	89.34	92.86	86.36	100	100	100	100	100	100
	8	96.22	96.22	92.73	94.36	94.55	94.03	95.86	94.33	97.05	98.31	98.11	98.55
	9	94.33	94.33	91.30	95.08	94.64	95.45	97.52	98.11	97.05	96.69	96.22	97.05
	10	96.22	96.22	89.70	95.00	94.74	95.27	97.52	98.11	97.05	89.25	88.67	89.97
Average	–	91.45	94.54	89.06	93.03	93.96	92.24	**97.56**	**97.94**	**96.98**	95.70	95.66	95.73

15%, 20%, 30%, 40%, 50%, 60%, 65%, 70%, 80%, 90%, and 100% of the total features, and finally all the features are included in the calculation and in Table 6, the accuracy level changes in 10 repetitions (the 5-fold final output) are observed. In the feature selection step, two 5-fold cross validation stages were used in the cost function for generating three train, validation, and test data.

Figure 10 shows a comparison between features selection approaches that output based on the lowest error rate with 5-fold cross validation and it performed with other methods such as Genetic algorithm (GA), Particle Swarm Optimization (PSO) and Ant Colony Optimization (ACO) and the most optimal features have been computed in comparison with the DE method of this study in four repetitions. DE performance is expressed based on *Kappa* coefficient by Eq. (23) in which was more appropriate than other solutions in feature selection level.

$$Kappa = \frac{N \sum_{i=1}^{r} x_{ii} - \sum_{i=1}^{r} (x_{i+} \times x_{+i})}{N^2 - \sum_{i=1}^{r} (x_{i+} \times x_{+i})} \tag{23}$$

where, N is the total number of samples, r is the number of classes, x_{ii} is the main diagonal of adjacency matrix of error, x_{i+} is the marginal sum of the rows and x_{+i} is the marginal sum of the columns.

Using other evolutionary algorithms such as PSO, ACO, GA, instead of the SFLA algorithm in optimizing the wavelet kernel of the ELM and its training did not result in less errors. The main reason for this is the ability of the SFLA to find the global and local optimum and persistence to achieve the optimal solution. Additionally, the processing time of the SFLA algorithm is more appropriate than other solutions and the advantage of not having much parameters and simplicity in implementation can be considered as the main reason for choosing an algorithm to improve ELM. Hence, in Fig. 10, SFLA optimizer performance has also been compared to improve ELM performance with other evolutionary optimizers such as PSO, ACO and GA and, the error in any type of data due to lower optimality has been reported.

The overall performance of the ELM algorithm has been shown after the improvement by the SFLA in the set of Fig. 11, in which the error converges to the minimum values. Figure 12 also compared ELM classifier with linear, RBF, polynomial, and wavelet kernels as decision makers for each class of sample slices, with and without MS and, the final result indicates the proper performance of this wavelet kernel in distinguishing healthy individuals and MS patients. The ELM classifier with different kernels is optimized by SFLA method.

The use of fractal and PZMs descriptors to extract feature yielded optimal responses. The primary reason for employing the proposed methods in feature extraction is
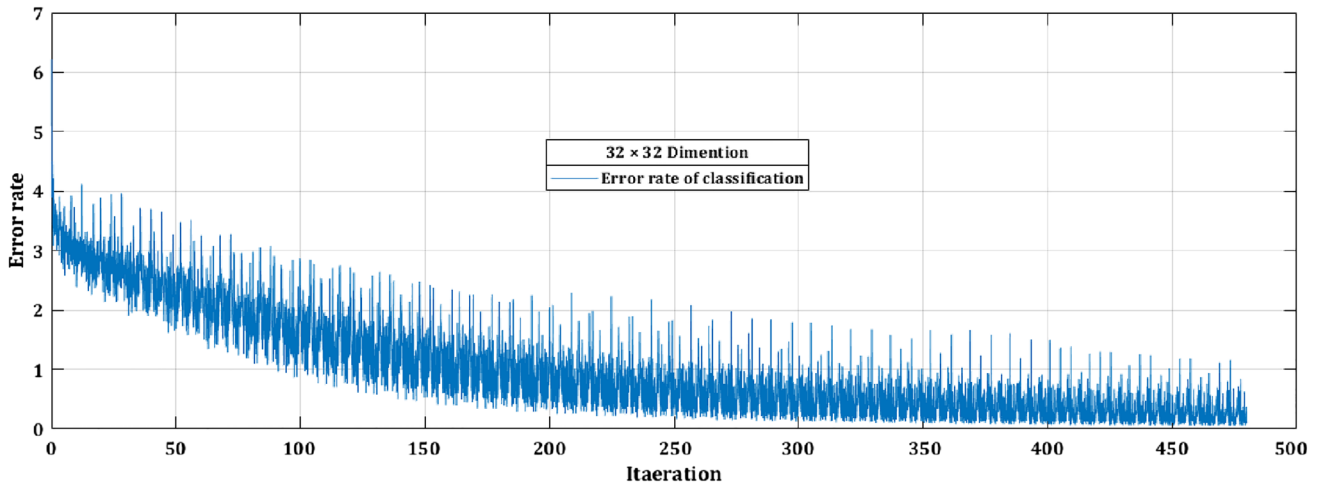
Fig. 6 The error rate and convergence of the proposed model after 500 iterations for images with 32×32 dimension
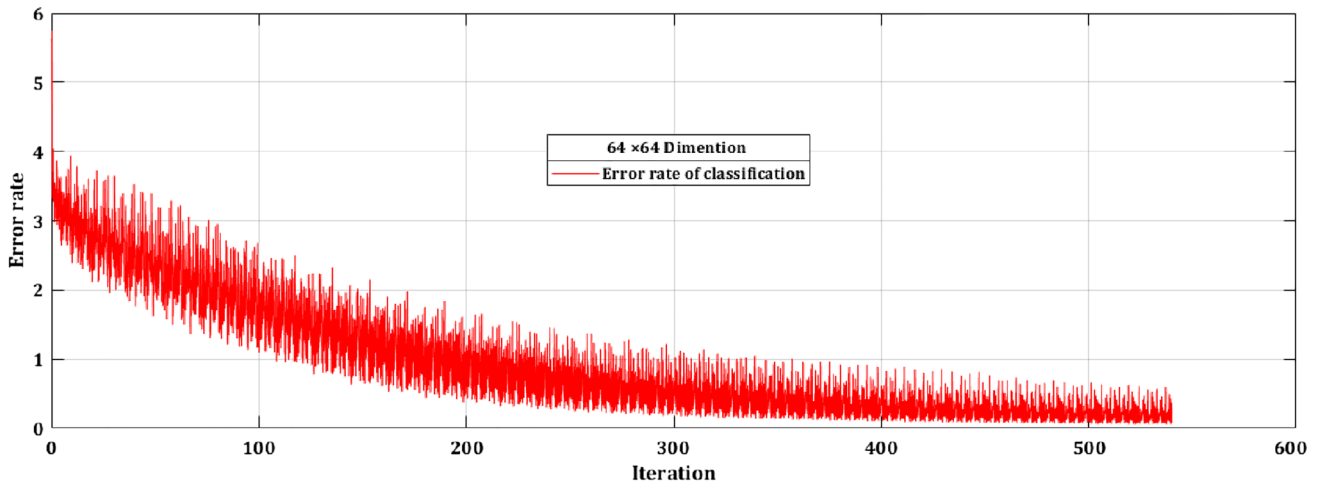

Fig. 7 The error rate and convergence of the proposed model after 500 iterations for images with 64×64 dimension
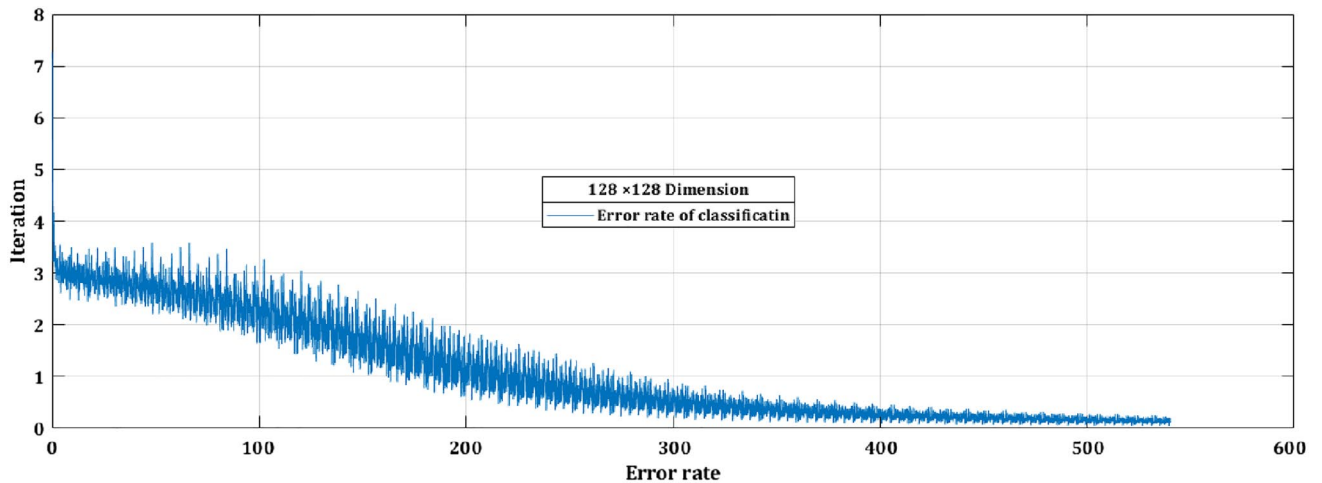

Fig. 8 The error rate and convergence of the proposed model after 500 iterations for images with 128×128 dimension
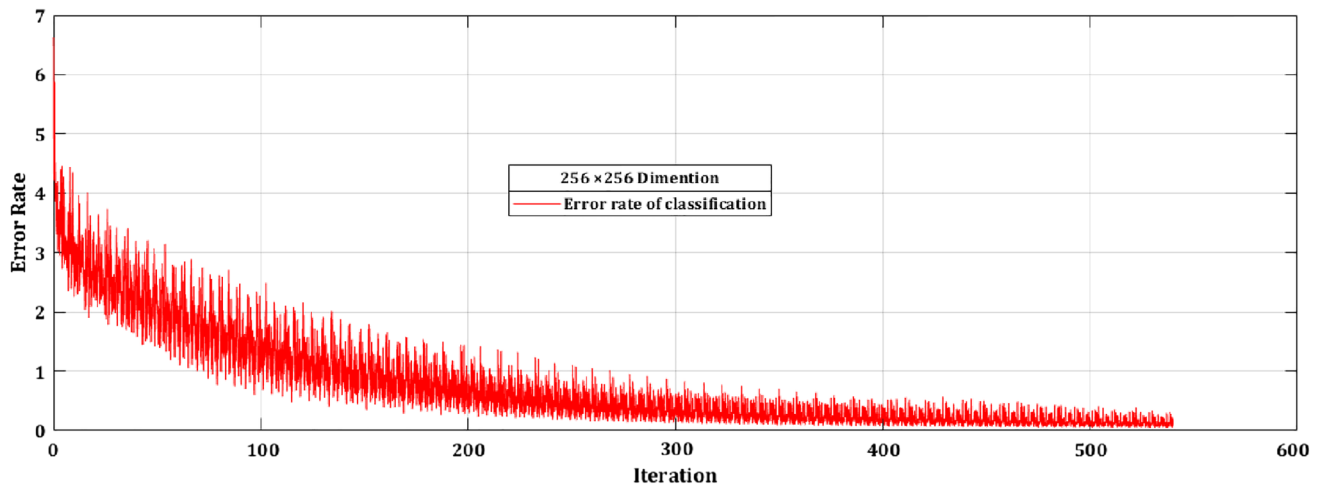
Fig. 9 The error rate and convergence of the proposed model after 500 iterations for images with 256×256 dimension

Table 6 The effect of dimension reduction in the calculated accuracy for the train and test steps. The bold values are best accuracies

|	Mean	Minimum number			Medium–low number			Medium–high number			Maximum number		
		10%	15%	20%	30%	40%	50%	60%	65%	70%	80%	90%	100%
All data	5-fold (1)	88.81	92.54	96.72	**98.22**	95.17	95.46	94.85	93.11	92.75	92.01	91.19	91.12
	5-fold (2)	90.47	93.69	95.54	**98.65**	96.12	95.62	94.16	94.90	93.13	92.14	91.66	90.81
	5-fold (3)	88.12	93.95	96.03	96.37	**97.18**	93.48	93.76	92.49	92.25	93.93	92.35	89.46
	5-fold (4)	87.70	92.75	94.14	96.03	**96.84**	94.73	93.43	93.84	93.51	93.25	92.56	91.23
	5-fold (5)	90.91	92.96	95.19	**98.61**	97.60	96.61	95.51	94.71	94.50	93.75	92.54	92.28
	5-fold (6)	87.34	91.49	96.42	**97.83**	96.29	95.44	94.54	95.34	93.62	92.84	91.93	90.17
	5-fold (7)	90.09	91.15	95.18	97.39	**97.71**	94.53	93.32	94.58	94.69	93.07	91.72	90.31
	5-fold (8)	91.75	93.75	96.91	**96.22**	95.82	94.18	93.07	93.52	94.88	93.81	92.37	91.49
	5-fold (9)	89.27	92.47	95.57	**99.65**	96.35	94.79	94.67	94.22	93.91	94.24	92.21	91.50
	5-fold (10)	88.85	93.06	95.73	**97.55**	96.09	95.48	94.19	93.38	92.54	91.35	91.77	91.13
Aveeage	–	**92.61**			**96.40**			**93.82**			**91.74**		
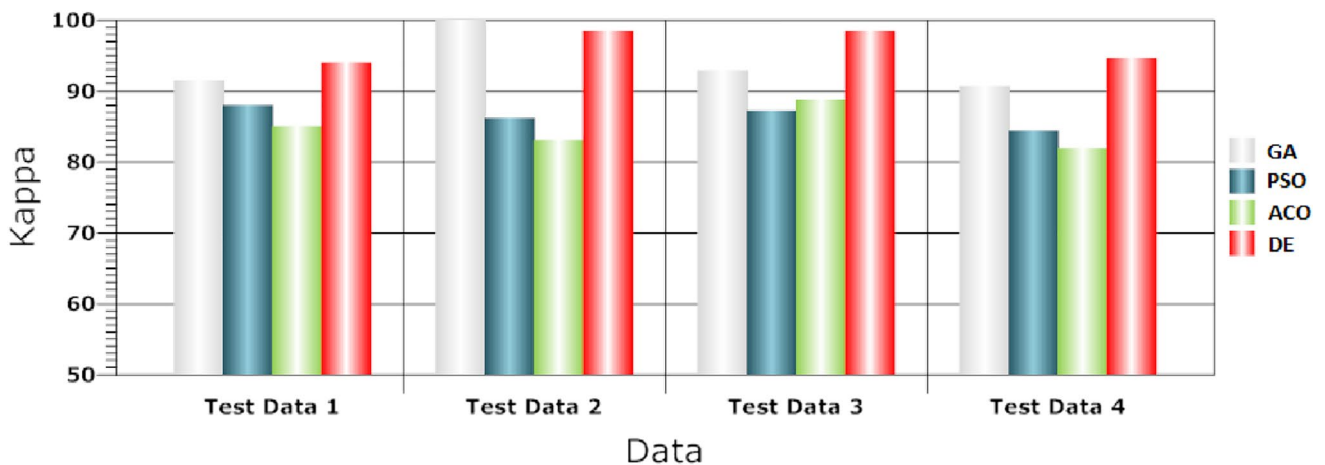
Fig. 10 The performance and comparison of the DE algorithm versus GA, PSO, and ACO algorithms in the feature selection step for four random sections of test data

its high accuracy in the analysis of MR images, in addition to its novelty and limited application in MR image analysis. Second, the set of images was analyzed by identical methods including GLCM, LBP, and HOG, which were compared with the proposed method in terms of feature extraction. Figure 13 shows that the calculation of performance accuracy at different test data by fractal and PZMs aggregation can produce responses that are more desirable than that of conventional methods such as GLCM, LBP and HOG. The third reason is that due to the nature of these descriptors, as explained in [39, 46–48], they are capable of extracting features even in small-scaled or rotated image or where the mass has been altered [47]. The fractal descriptor yielded high accuracy in Lahmiri's research for the detection of Alzheimer's lesions. It has been shown that fractal is effectively correlated with human perception of surface roughness.

In addition to the comparison of descriptors, the repeatability of the algorithm was tested in four random datasets by calculating Accuracy, Sensitivity, and Specificity factors.

Figure 14 shows the algorithm's repeatability with the minimum dispersion in the outcomes.

The algorithm performance was evaluated by calculating the evaluation factors and calculating their mean values, in such a way that:

(A) Processing model: Information processing was performed on the basis of the comparison between the solutions of the magnetic resonance image analysis with the proper accuracy.
(B) Repetition: The variance in responses is low and unmodeled uncertainty is solvable.
(C) Processing time: The time reduction rate was lower during the design phase and evaluation. However, the data processing time is half the overall time, which can be important in saving time.

Consisting of descriptions of features based on fractal and PZM methods, feature selection with DE algorithm and wavelet kernel optimization by SFLA, the
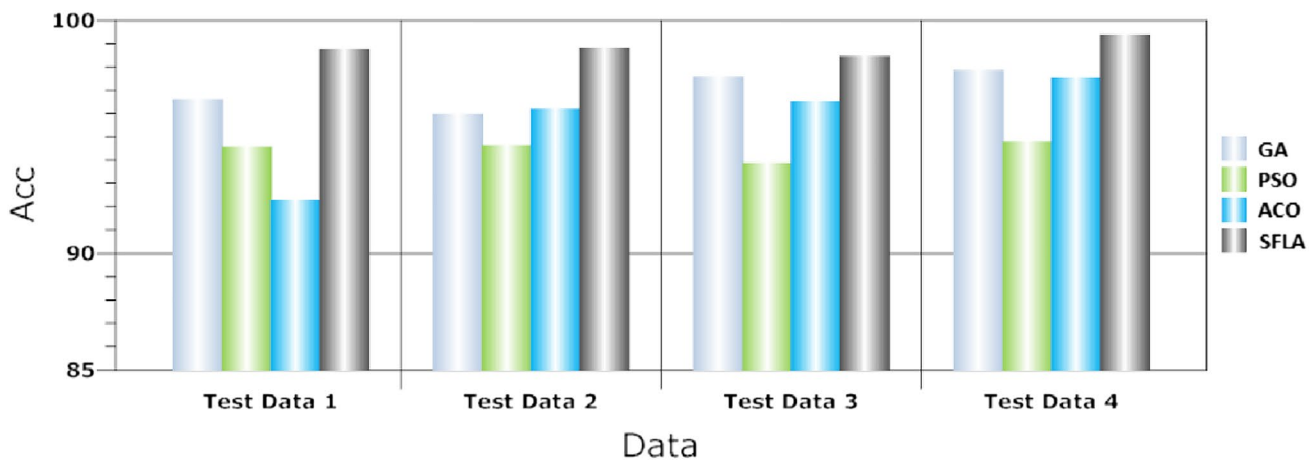

Fig. 11 The parameters regularization of wavelet kernel by SFLA versus GA, PSO, and ACO algorithms for four random sections of test data
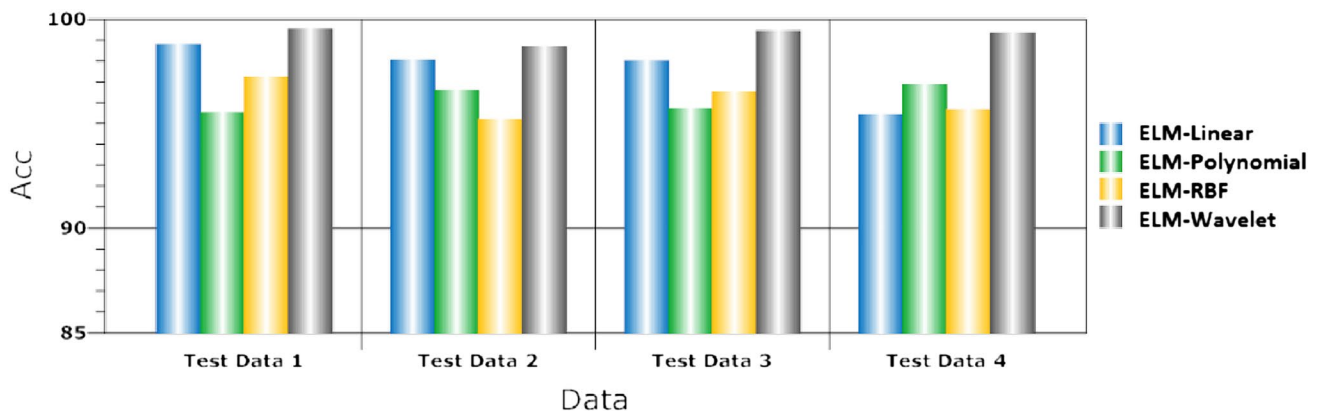

Fig. 12 The performance of ELM classifier with linear, RBF, polynomial and wavelet kernels for four random sections of test data

Fig. 13 Comparison between conventional feature extraction methods in MR images to detect the MS lesions for four random sections of test data
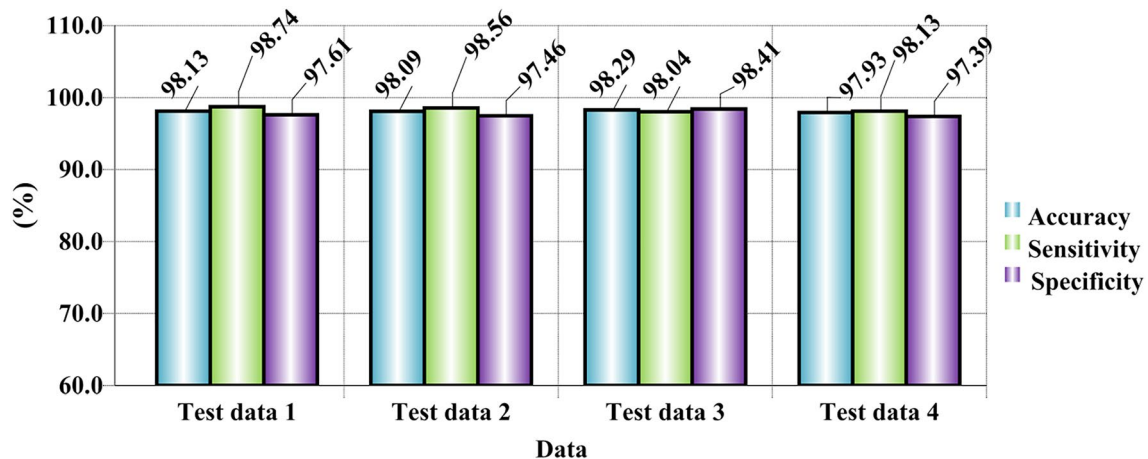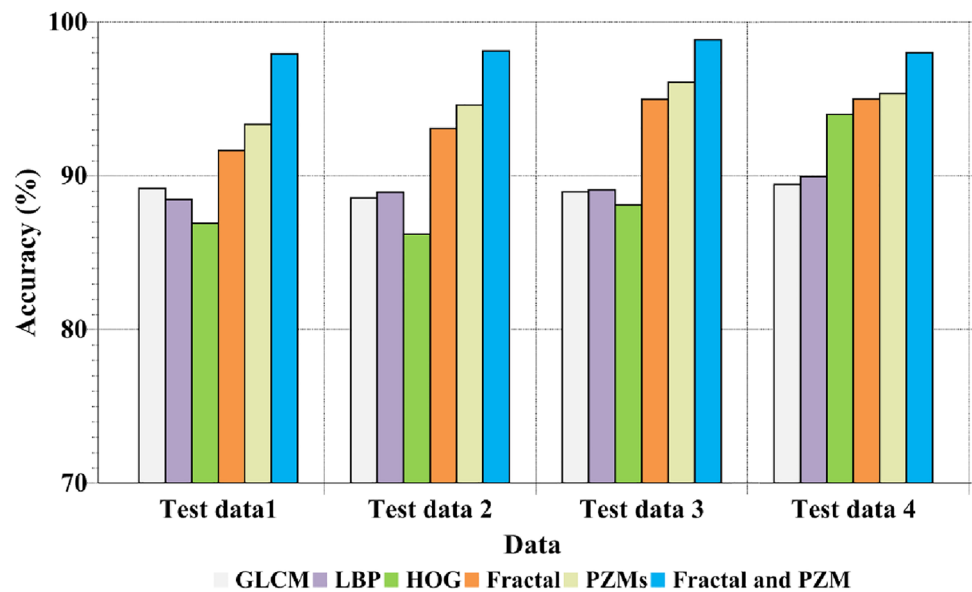

Fig. 14 The accuracy, sensitivity, and specificity factors of proposed model for four random sections of test data

ELM performance is better than other similar classifiers. In overall, we achieve accuracy higher than 97% at train and test stage via dividing the data by K-Fold cross-validation.

5 Conclusion

Considering the nature of magnetic resonance image of brain, cancerous masses, brain lesions and their areas, in this paper, a method has been proposed on the basis of MS diagnosis based on the use of automatic processing technique. MS-related lesions have often abnormal

position in MRI and because of compression in their slice, they exhibit a different brightness. Using csFCM, pixels containing probable MS lesions were separated from others. Feature extraction from the same separated pixels, the type and shape of the lesion by the two descriptors, makes it possible to create the proper feature vector of slices. The DE algorithm removed some of the data from the feature vector that, the DE algorithm function cost function was achieved by data segmentation through 5-fold cross validation on the basis of the calculation of ELM class error. The optimal ELM classifier by the SFLA has better capability for classification and the existing claims were proven. By repeating the test,

there was a slight difference in the responses obtained which indicated un-modeled uncertainty. Undoubtedly, more improvements can be made in terms of learning and processing time and the error can be minimized. The suggestion is that a larger volume of slices of each subject will be introduced into the classifier by modeling methods such as deep learning.

Compliance with ethical standards

Conflict of interest We have no conflict of interest to declare.

Ethical approval We have all permissions for working on the real medial data from the ethical committee of Vasei hospital in Sabzevar, Iran, regarding the healthy and patient individuals. Also, the results of our research can be reported based on the same permission received.

References

1. Filippi M, Brück W, Chard D, Fazekas F, Geurts JJ, Enzinger C, Hametner S, Kuhlmann T, Preziosa P, Rovira À, Schmierer K (2019) Association between pathological and MRI findings in multiple sclerosis. Lancet Neurol 18(2):198–210

2. Eichinger P, Schön S, Pongratz V, Wiestler H, Zhang H, Bussas M, Hoshi MM, Kirschke J, Berthele A, Zimmer C, Hemmer B (2019) Accuracy of unenhanced MRI in the detection of new brain lesions in multiple sclerosis. Radiology 291(2):429–435

3. Schmierer K, McDowell A, Petrova N, Carassiti D, Thomas DL, Miquel ME (2018) Quantifying multiple sclerosis pathology in post mortem spinal cord using MRI. Neuroimage 15(182):251–258

4. Fechner A, Savatovsky J, El Methni J, Sadik JC, Gout O, Deschamps R, Gueguen A, Lecler A (2019) A 3T phase-sensitive inversion recovery MRI sequence improves detection of cervical spinal cord lesions and shows active lesions in patients with multiple sclerosis. Am J Neuroradiol 40(2):370–375

5. Silveira F, Sánchez F, Miguez J, Contartese L, Gómez A, Patrucco L, Cristiano E, Rojas JI (2020) New MRI lesions and topography at 6 months of treatment initiation and disease activity during follow up in relapsing remitting multiple sclerosis patients. Neurol Res 42(2):148–152

6. Rovira À, Wattjes MP, Tintoré M, Tur C, Yousry TA, Sormani MP, De Stefano N, Filippi M, Auger C, Rocca MA, Barkhof F (2015) Evidence-based guidelines: MAGNIMS consensus guidelines on the use of MRI in multiple sclerosis—clinical implementation in the diagnostic process. Nat Rev Neurol 11(8):471

7. Ysraelit MC, Fiol MP, Gaitán MI, Correale J (2018) Quality of life assessment in multiple sclerosis: different perception between patients and neurologists. Front Neurol 11(8):729

8. Brundin L, Kobelt G, Berg J, Capsa D, Eriksson J (2017) European multiple sclerosis platform: new insights into the burden and costs of multiple sclerosis in Europe: Results for Sweden. Mult Scleros J 23(2):179–191

9. Khosravi M, Newberg A, Jahangiri P, Raynor W, Al-zaghal A, Werner T, Alavi A (2018) Innovative applications of combined PET/MRI modality in diagnosis and follow-up of Multiple Sclerosis. J Nucl Med 59(supplement 1):1220

10. Novakova L, Axelsson M, Khademi M, Zetterberg H, Blennow K, Malmeström C, Piehl F, Olsson T, Lycke J (2017) Cerebrospinal fluid biomarkers as a measure of disease activity and treatment efficacy in relapsing-remitting multiple sclerosis. J Neurochem 141(2):296–304

11. Thompson AJ, Banwell BL, Barkhof F, Carroll WM, Coetzee T, Comi G, Correale J, Fazekas F, Filippi M, Freedman MS, Fujihara K (2017) Diagnosis of multiple sclerosis: 2017 revisions of the McDonald criteria. Lancet Neurol 1:1

12. Rezaee K, Azizi E, Ghezelbash MR, Madanian M, Haddania J (2012) A novel intelligent system to accurately segmentation of brain tumors in MR images by using image processing and discrete wavelet transform. Majlesi J Multimed Process 1(4):1

13. Akselrod-Ballin A, Galun M, Basri R, Brandt A, Gomori MJ, Filippi M, Valsasina P (2006) An integrated segmentation and classification approach applied to multiple sclerosis analysis. In: IEEE computer society conference on computer vision and pattern recognition, vol 1. IEEE, pp 1122–1129

14. Zhang J, Wang L, Tong L (2007) Feature reduction and texture classification in MRI-Texture analysis of multiple sclerosis. In: IEEE/ICME international conference on complex medical engineering, 2007, CME 2007. IEEE, pp 752–757

15. Roy PK, Bhuiyan A, Ramamohanarao K (2013) Automated segmentation of multiple sclerosis lesion in intensity enhanced flair MRI using texture features and support vector machine. In: IEEE international conference on image processing 2013. IEEE, pp 4277–4281

16. Elliott C, Arnold DL, Collins DL, Arbel T (2013) Temporally consistent probabilistic detection of new multiple sclerosis lesions in brain MRI. IEEE Trans Med Imaging 32(8):1490–1503

17. Cabezas M, Oliver A, Valverde S, Beltran B, Freixenet J, Vilanova JC, Ramió-Torrentà L, Rovira À, Lladó X (2014) BOOST: a supervised approach for multiple sclerosis lesion segmentation. J Neurosci Methods 30(237):108–117

18. Sweeney EM, Vogelstein JT, Cuzzocreo JL, Calabresi PA, Reich DS, Crainiceanu CM, Shinohara RT (2014) A comparison of supervised machine learning algorithms and feature vectors for MS lesion segmentation using multimodal structural MRI. PLoS One 9(4):e95753

19. Ardakani AA, Gharbali A, Saniei Y, Mosarrezaii A, Nazarbaghi S (2015) Application of texture analysis in diagnosis of multiple sclerosis by magnetic resonance imaging. Glob J Health Sci 7(6):68

20. Liu J, Brodley CE, Healy BC, Chitnis T (2015) Removing confounding factors via constraint-based clustering: an application to finding homogeneous groups of multiple sclerosis patients. Artif Intell Med 65(2):79–88

21. Weygandt M, Hummel HM, Schregel K, Ritter K, Allefeld C, Dommes E, Huppke P, Haynes J, Wuerfel J, Gärtner J (2015) MRI-based diagnostic biomarkers for early onset pediatric multiple sclerosis. NeuroImage Clin 7:400–408

22. Karimaghaloo Z, Rivaz H, Arnold DL, Collins DL, Arbel T (2015) Temporal hierarchical adaptive texture CRF for automatic detection of gadolinium-enhancing multiple sclerosis lesions in brain MRI. IEEE Trans Med Imaging 34(6):1227–1241

23. Brosch T, Tang LY, Yoo Y, Li DK, Traboulsee A, Tam R (2016) Deep 3D convolutional encoder networks with shortcuts for multiscale feature integration applied to multiple sclerosis lesion segmentation. IEEE Trans Med Imaging 35(5):1229–1239

24. Zhang Y, Lu S, Zhou X, Yang M, Wu L, Liu B, Phillips P, Wang S (2016) Comparison of machine learning methods for stationary wavelet entropy-based multiple sclerosis detection: decision

tree, k-nearest neighbors, and support vector machine. Simulation 92(9):861–871

25. Yoo Y, Tang LY, Brosch T, Li DK, Kolind S, Vavasour I, Rauscher A, MacKay AL, Traboulsee A, Tam RC (2018) Deep learning of joint myelin and T1w MRI features in normal-appearing brain tissue to distinguish between multiple sclerosis patients and healthy controls. NeuroImage Clin 17:169–178

26. Gheshlaghi SH, Ranjbar A, Suratgar AA, Menhaj MB, Faraji F (2019) A superpixel segmentation based technique for multiple sclerosis lesion detection. arXiv preprint arXiv:1907.03109

27. Valverde S, Cabezas M, Roura E, González-Villà S, Pareto D, Vilanova JC, Ramió-Torrentà L, Rovira À, Oliver A, Lladó X (2017) Improving automated multiple sclerosis lesion segmentation with a cascaded 3D convolutional neural network approach. NeuroImage 15(155):159–168

28. Souplet, JC, Lebrun C, Ayache N, Malandain G (2008) An automatic segmentation of T2-FLAIR multiple sclerosis lesions. In: Grand challenge work: multiple sclerosis lesion segmentation. Challenge. pp 1–11

29. Schmidt P, Gaser C, Arsic M, Buck D, Förschler A, Berthele A, Hoshi M, Ilg R, Schmid VJ, Zimmer C, Hemmer B (2012) An automated tool for detection of FLAIR-hyperintense white-matter lesions in multiple sclerosis. Neuroimage 59(4):3774–3783

30. Kwok PP, Ciccarelli O, Chard DT, Miller DH, Alexander DC (2012) Predicting clinically definite multiple sclerosis from onset using SVM. In: Machine learning and interpretation in neuroimaging 2012. Springer, Berlin, pp 116–123

31. Cerasa A, Bilotta E, Augimeri A, Cherubini A, Pantano P, Zito G, Lanza P, Valentino P, Gioia MC, Quattrone A (2012) A cellular neural network methodology for the automated segmentation of multiple sclerosis lesions. J Neurosci Methods 203(1):193–199

32. Cabezas M, Oliver A, Roura E, Freixenet J, Vilanova JC, Ramió-Torrentà L, Rovira À, Lladó X (2014) Automatic multiple sclerosis lesion detection in brain MRI by FLAIR thresholding. Comput Methods Programs Biomed 115(3):147–161

33. Kawahara J, McIntosh C, Tam R, Hamarneh G (2014) Novel morphological and appearance features for predicting physical disability from MR images in multiple sclerosis patients. In: Computational methods and clinical applications for spine imaging. Springer, Cham, pp 61–73

34. Siddiqui MF, Reza AW, Kanesan J (2015) An automated and intelligent medical decision support system for brain MRI scans classification. PLoS One 10(8):e0135875

35. Deshpande H, Maurel P, Barillot C (2015) Classification of multiple sclerosis lesions using adaptive dictionary learning. Comput Med Imaging Graph 1(46):2–10

36. Alshayeji MH, Al-Rousan MA, Ellethy H (2018) An efficient multiple sclerosis segmentation and detection system using neural networks. Comput Electr Eng 1(71):191–205

37. Jobson DJ, Rahman ZU, Woodell GA (1997) Properties and performance of a center/surround retinex. IEEE Trans Image Process 6(3):451–462

38. Adhikari SK, Sing JK, Basu DK, Nasipuri M (2015) Conditional spatial fuzzy C-means clustering algorithm for segmentation of MRI images. Appl Soft Comput 1(34):758–769

39. Gorji HT, Haddadnia J (2015) A novel method for early diagnosis of Alzheimer's disease based on pseudo Zernike moment from structural MRI. Neuroscience 1(305):361–371

40. Costa AF, Humpire-Mamani G, Traina AJ (2012) An efficient algorithm for fractal analysis of textures. In: 25th SIBGRAPI conference on graphics, patterns and images. IEEE, pp 39–46

41. Khawaled S, Zibulevsky M, Zeevi YY (2019) Texture and Structure Two-view Classification of Images. arXiv preprint arXiv:1908.09264

42. Krohn S, Froeling M, Leemans A, Ostwald D, Villoslada P, Finke C, Esteban FJ (2019) Evaluation of the 3D fractal dimension as a marker of structural brain complexity in multiple-acquisition MRI. Human brain mapping

43. Ahmadi A, Davoudi S, Daliri MR (2019) Computer aided diagnosis system for multiple sclerosis disease based on phase to amplitude coupling in covert visual attention. Comput Methods Programs Biomed 1(169):9–18

44. Zhang YD, Zhao G, Sun J, Wu X, Wang ZH, Liu HM, Govindaraj VV, Zhan T, Li J (2018) Smart pathological brain detection by synthetic minority oversampling technique, extreme learning machine, and Jaya algorithm. Multimed Tools Appl 77(17):22629–22648

45. Nayak DR, Dash R, Majhi B (2017) An improved extreme learning machine for pathological brain detection. In: TENCON 2017—2017 IEEE region 10 conference. IEEE, pp 13–18

46. Zhang YD, Jiang Y, Zhu W, Lu S, Zhao G (2018) Exploring a smart pathological brain detection method on pseudo Zernike moment. Multim Tools Appl 77(17):22589–22604

47. Lahmiri S (2016) Image characterization by fractal descriptors in variational mode decomposition domain: application to brain magnetic resonance. Physica A 15(456):235–243

48. Backes AR, Casanova D, Bruno OM (2012) Color texture analysis based on fractal descriptors. Pattern Recogn 45(5):1984–1992