



An Improved Pathological Brain Detection System Based on Two-Dimensional PCA and Evolutionary Extreme Learning Machine

Deepak Ranjan Nayak¹ · Ratnakar Dash¹ · Banshidhar Majhi¹

Received: 11 August 2017 / Accepted: 15 November 2017 / Published online: 7 December 2017
© Springer Science+Business Media, LLC, part of Springer Nature 2017

Abstract

Pathological brain detection has made notable stride in the past years, as a consequence many pathological brain detection systems (PBDSs) have been proposed. But, the accuracy of these systems still needs significant improvement in order to meet the necessity of real world diagnostic situations. In this paper, an efficient PBDS based on MR images is proposed that markedly improves the recent results. The proposed system makes use of contrast limited adaptive histogram equalization (CLAHE) to enhance the quality of the input MR images. Thereafter, two-dimensional PCA (2DPCA) strategy is employed to extract the features and subsequently, a PCA+LDA approach is used to generate a compact and discriminative feature set. Finally, a new learning algorithm called MDE-ELM is suggested that combines modified differential evolution (MDE) and extreme learning machine (ELM) for segregation of MR images as pathological or healthy. The MDE is utilized to optimize the input weights and hidden biases of single-hidden-layer feed-forward neural networks (SLFN), whereas an analytical method is used for determining the output weights. The proposed algorithm performs optimization based on both the root mean squared error (RMSE) and norm of the output weights of SLFNs. The suggested scheme is benchmarked on three standard datasets and the results are compared against other competent schemes. The experimental outcomes show that the proposed scheme offers superior results compared to its counterparts. Further, it has been noticed that the proposed MDE-ELM classifier obtains better accuracy with compact network architecture than conventional algorithms.

Keywords Magnetic resonance imaging (MRI) · Two-dimensional PCA (2DPCA) · Extreme learning machine (ELM) · Pathological brain detection (PBD) · Modified differential evolution (MDE)

Introduction

Over the years due to brain diseases, the mortality rate increases vastly among individuals with different age groups across the globe. Pathological brain detection (PBD) has played significant role for early identification of various diseases such as Alzheimer's disease [36], mild cognitive

impairment, autism spectrum disorder [6], multiple sclerosis [33], hearing loss [34], and microbleeding [43]. The major objective of PBD is to assist radiologists to arrive at correct and quick clinical decisions. In PBD, a non-invasive imaging modality called magnetic resonance imaging (MRI) is often used since it supplies better resolution of brain tissues [32]. However, manual interpretation of MR images is a costly, troublesome and time-consuming task [2, 13, 15]. Hence, current trend is to develop automated PBD systems (PBDSs) with the help of image processing and machine learning algorithms which can detect brain diseases in less time. Further, it has been shown that PBDSs are effective and have practical applications.

Many attempts have been made toward the development of various PBDSs in the past decade [4]. However, the accuracy of these systems still requires notable improvement in order to meet the necessity of real world diagnostic situations.

This article is part of the Topical Collection on *Advanced Computational Intelligence and Soft Computing in Medical Imaging*

✉ Deepak Ranjan Nayak
depakranjannayak@gmail.com

¹ Pattern Recognition Lab, Department of Computer Science and Engineering, National Institute of Technology, Rourkela, 769 008, India

Hence, PBDS remains an open challenging issue in front of researchers. The goal of this study is to improve the performance of the system for pathological brain detection.

It has been observed that discrete wavelet transform (DWT) is the most used feature extractor in PBDSs as it analyzes images at several scales and handles one-dimensional (1D) singularities effectively. However, it has limited capability of representing two-dimensional 2D singularities (edges of an image). That is, DWT is not able to capture curve like features effectively from the images. Therefore, to handle such issue, application of advanced transforms are in great demand. Further, classifiers like support vector machine (SVM) and feed-forward neural network (FNN) are often used in earlier PBDSs. To train FNN, traditional gradient-based learning algorithms such as Levenberg-Marquardt (LM) and back-propagation (BP) are used which have many limitations such as trapping at local minima, slower learning speed, and learning epochs. Furthermore, traditional SVM classifier encounters higher computational complexity and it performs poorly on large datasets.

To overcome the aforementioned problems, we propose a novel PBDS in this paper. The main contributions of this study are summarized as follows:

- (a) Two dimensional PCA (2DPCA) is explored to extract the features from MR images.
- (b) To combat the issues of conventional learning algorithms, a simple and effective learning technique known as extreme learning machine (ELM) is employed.
- (c) To further enhance the performance of standard ELM, a new learning algorithm known as MDE-ELM based on modified differential evolution (MDE) and ELM is proposed.
- (d) To test the effectiveness of the suggested scheme, extensive experiments are conducted on three well-known datasets. In this context, the suggested scheme is compared against its counterparts with respect to classification accuracy and number of features required.

The remaining part of the article is structured as follows. Section “[Related work](#)” summarizes the related works. Section “[Datasets used](#)” offers the description of the datasets used in this study. Section “[Proposed work](#)” discusses the proposed methodology. In “[Experimental results and analysis](#)”, the experimental details and comparisons are presented. Finally, the concluding remarks are drawn in “[Conclusions and future work](#)”.

Related work

A significant number of PBDSs have been proposed in the past decade [4, 16]. Chaplot et al. [1] have suggested to use 2D discrete wavelet transform (2D DWT) and

support vector machine (SVM) for feature extraction and classification. El-Dahshan et al. [5] have employed 2D DWT and two classifiers such as k -nearest neighbor (KNN) and feed forward back-propagation artificial neural network (FP-ANN). To reduce the feature dimensionality, they have applied principal component analysis (PCA). The authors in [32, 38, 40] have used scaled conjugate gradient (SCG), particle swarm optimization (PSO), adaptive chaotic PSO (ACPSO), and scaled chaotic artificial bee colony (SCABC) to train the feed forward neural network (FNN) classifier. Zhang et al. [39] have combined DWT, PCA and kernel SVM (KSVM). In [2], a PBDS based on Ripplet transform (RT), PCA and least squares SVM (LS-SVM) is suggested. In [18], the authors harnessed wavelet entropy (WE) to extract features and probabilistic neural network (PNN) is used for classification. Later, in [4], the authors have combined feedback pulse coupled neural network (FPCNN), DWT, PCA and FNN to detect pathological brain. Zhang et al. [41] have used weighted-type fractional Fourier transform (WFRFT) and two individual classifiers such as generalized eigenvalue proximal SVM (GEPSSVM) and twin SVM (TSVM). Later, Yang et al. [26] have used wavelet energy values of as features. They have applied biogeography-based optimization (BBO) to train SVM classifier. Dong et al. [31] have utilized wavelet packet Shannon entropy (WPSE) and wavelet packet Tsallis entropy (WPTE) separately as features. In this, GEPSSVM is employed as classifier. Nayak et al. [13] have utilized 2D DWT, probabilistic PCA (PPCA) and AdaBoost with random forests (ADBRF) for identifying pathological brains. In [30], the authors have offered a PBDS which combines stationary wavelet transform (SWT), PCA, and GEPSSVM. In [12], a PCA+LDA technique is applied on the 2D DWT features. In [45], Naive Bayes classifier (NBC) based PBDS is proposed which uses WE features. While, in [29], wavelet energy and SVM is used. Sun et al. [37] have utilized GEPSSVM+RBF classifier on WE and Hu moment invariants (HMI) features. Wang et al. [23] have proposed a novel feature called fractional Fourier entropy (FRFE) and performed Welch’s t -test (WTT) to select the relevant features. Twin SVM (TSVM) classifier is employed for classification. Later, in [35], a PBDS based on FRFE features and multilayer perceptron (MLP) is proposed. They have employed an adaptive real coded BBO (ARCBBO) approach for training the MLP. In this case, the number of hidden neurons of MLP is found using three separate pruning methods, namely, Bayesian detection boundaries (BDB), dynamic pruning (DP) and Kappa coefficient (KC). Chen et al. [42] have utilized Minkowski-Bouligand dimension (MBD) features and proposed an improved PSO (IPSO) to train the single-hidden layer feedforward neural network. Dash et al. [14] have introduced a PBDS harnessing fast discrete curvelet transform and LS-SVM.

Later on, Wang et al. [22] have combined the variance and entropy (VE) values of dual-tree complex wavelet transform (DTCWT) and TSVM to detect pathological brain. Li et al. [21] have employed wavelet packet Tsallis entropy (WPTE) and FNN with real-coded biogeography-based optimization (RCBBO) for pathological brain detection.

The literature study shows that most PBDSs used different forms of wavelet like DWPT, SWT, DTCWT, etc., as feature extractor. Despite the merits of these approaches, it has been observed that none of the approaches are able to achieve perfect classification accuracy in all cases. Therefore, application of proper feature extraction algorithms needs to be explored. Further, classifiers like SVM and FNN are frequently used in the existing PBDSs in spite of many loopholes. Moreover, it has been found that few PBDSs need a large number of features and hence, there exists a scope to limit the feature requirement without compromising the accuracy. It is noted that Yang et al. [28] have proposed an efficient and novel image feature extraction technique called two dimensional PCA (2DPCA) which has gained tremendous attention from researchers in last decade. 2DPCA was initially applied to face recognition task and thereafter, it has been leveraged in many applications.

In order to combat the above issues, we have proposed an efficient PBDS to classify the MR image as healthy or pathological. The proposed PBDS utilizes 2DPCA for feature extraction. Subsequently, PCA+LDA approach is employed in order to decide the most significant feature set. Lastly, an improved learning algorithm called MDE-ELM is proposed which offers several advantages such as local minima avoidance, better generalization capability, faster learning rate, and well-conditioned over other classifiers like FNN, SVM, LS-SVM, ELM, etc.

Datasets used

The proposed PBDS has been evaluated on three benchmark datasets, namely, DS-I, DS-II, and DS-III which carries 66, 160 and 255 brain MR images respectively. The datasets accommodate T2-weighted brain MR images of size 256×256 in axial view plane which were downloaded from Medical School of Harvard University website [10]. Both DS-I and DS-II hold samples of seven categories of diseases such as sarcoma, glioma, meningioma, AD plus visual agnosia (VA), Pick's disease (PD), AD and Huntington's disease (HD) plus healthy brain samples. However, DS-III includes four more diseases such as cerebral toxoplasmosis (CTP), multiple sclerosis (MS), herpes encephalitis (HE), and chronic subdural hematoma (CSH). The proposed work deals with solving a binary class classification problem (healthy or pathological), where the pathological class

contains images from all kinds of diseases. Samples of all kinds of MR images are shown in Fig. 1.

Proposed work

The proposed system involves four stages such as contrast limited adaptive histogram equalization (CLAHE) based preprocessing, 2DPCA based feature extraction, PCA+LDA based feature reduction, and MDE-ELM based classification. The input of the system is an MR image and the output is the class label (healthy or pathological). The overview of the proposed PBDS is depicted in Fig. 2. A detail description of each stage is presented below.

Preprocessing using CLAHE

It is observed that most of the images in the datasets considered in this study are of low-contrast. Therefore, for contrast enhancement of the images, a standard technique named CLAHE is employed. CLAHE initially evaluates a histogram of gray values at a contextual region surrounded by every pixel and thereafter, allocates a value to each pixel intensity within the display range [17]. Additionally, it uses a fixed value dubbed clip limit which helps in clipping the histogram prior to the computation of cumulative distribution function (CDF). However, CLAHE redistributes those parts of the histogram equally among all histogram bins that surpass the clip limit.

Feature extraction using 2DPCA

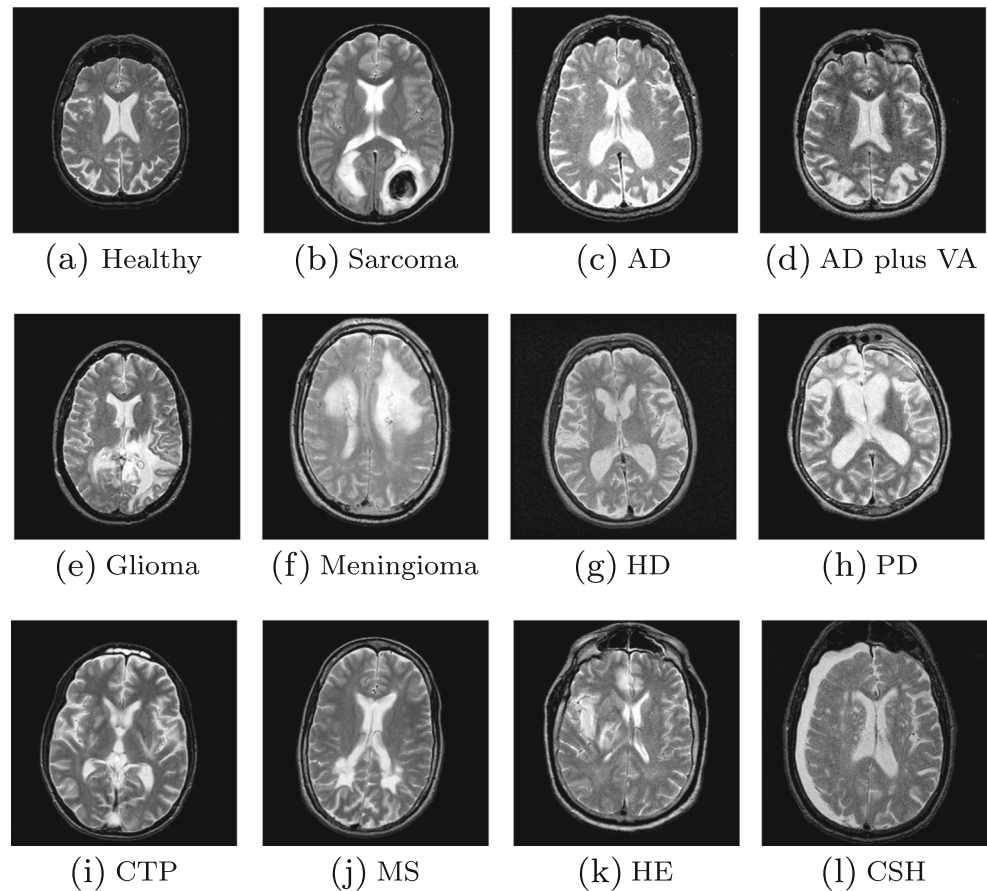
Two-dimensional PCA (2DPCA) has been shown to be promising in the domain of feature extraction and feature reduction over the last decade due to its salient properties like less memory storage and lower computational overhead [25]. In addition, 2DPCA enjoys decorrelation property and the feature vectors extracted from images are uncorrelated. It was originally applied to face recognition tasks and afterward, it has been successfully applied in several applications. This motivates us to employ 2DPCA for extracting features from brain MR images. Mathematically, it is described as follows.

For a given P training MR images ($I_j, j = 1, 2, \dots, P$) with size $m \times n$, the image covariance matrix in 2DPCA takes the form [28]

$$Cov = \frac{1}{P} \sum_{j=1}^P (I_j - \bar{I})^T (I_j - \bar{I}) \quad (1)$$

Here, Cov denotes a non-negative definite matrix of size $n \times n$ and \bar{I} is the mean of all the training images.

Fig. 1 T2-weighted brain MR samples [2]



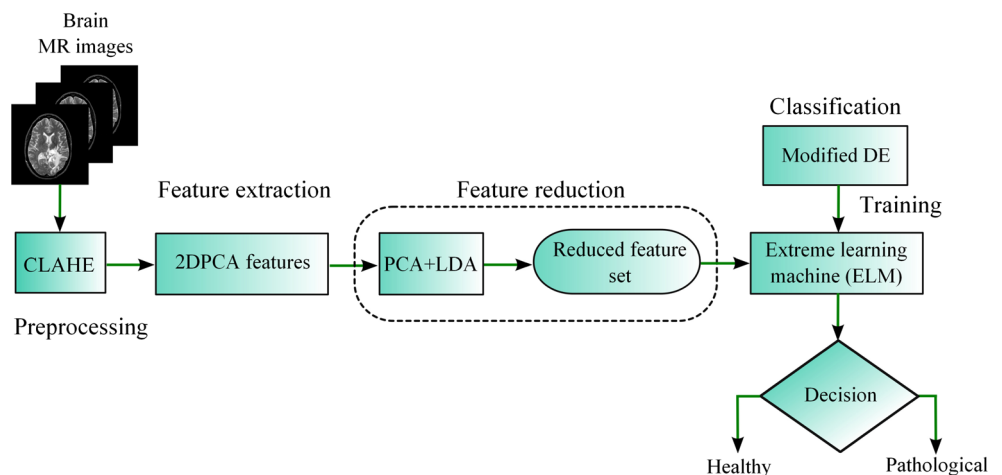
Next, we evaluate the eigenvalues and eigenvectors of matrix Cov . Then, α eigenvectors $V_1, V_2, \dots, V_\alpha$ (also called projection vector of 2DPCA) corresponding to α largest eigenvalues are selected as the transforming axes and these vectors are used for feature extraction. 2DPCA projects an image onto the transforming axes and serves the resultant α projections (projected vectors) as features which is stated as

$$R_i = IV_i, i = 1, 2, \dots, \alpha \tag{2}$$

It is worth mentioning here that α value is selected using a measure called normalized cumulative sum of variances (NCSV). The NCSV value for a^{th} eigenvector is calculated as

$$NCSV(a) = \frac{\sum_{u=1}^a \lambda(u)}{\sum_{u=1}^n \lambda(u)} ; 1 \leq a \leq n \tag{3}$$

Fig. 2 Overview of the proposed framework



where, $\lambda(u)$ represents the eigenvalue of the u^{th} eigenvector and n denotes the total number of the eigenvectors sorted in descending order of eigenvalues. Here, we choose a threshold value manually and the number of eigenvectors (for instance α) for which the NCSV value surpasses the threshold are selected. As mentioned earlier that these α eigenvectors are retained for extraction of features from the MR images.

For each input MR image, we apply 2DPCA and obtain the features. The implementation procedure of feature extraction is outlined in Algorithm 1.

Algorithm 1 Feature extraction using 2DPCA

Input: N input images: $I[x, y]$ of size $m \times n$
Output: Feature matrix: F_M of size $N \times D$

- 1: Calculate the mean of all the image \bar{I}
- 2: Evaluate image scatter matrix Cov using Eq. 1
- 3: Compute the eigenvalues and eigenvectors from matrix Cov and sort the eigenvectors in descending order according to eigenvalues.
- 4: Select α eigenvectors based on NCSV measure for a given threshold value.
- 5: **for** each image $I[x, y] \in N$ **do**
- 6: Find the projected vectors using Eq. 2
- 7: Arrange the projected vectors in a new vector of size $1 \times D$ and store in F_M .
- 8: **end for**
- 9: Obtain a feature matrix F_M containing features of images

Feature reduction using PCA+LDA

It has been observed that the features extracted using 2DPCA are of high dimension and the high dimensional feature vector prompts to high computational overhead and high storage space. Hence, application of dimensionality reduction techniques is of great importance. PCA has been found to be effective in reducing feature dimension which transforms high dimensional input data to a lower dimensional space while keeping maximum variations of the data. In contrast, linear discriminant analysis (LDA) attempts to find a feature subspace that best discriminates between the classes. But, conventional LDA performs poorly while dealing with high dimensional and small sample size problem as in this case the within-scatter matrix (S_w) is always singular [27]. Further, to make sure that S_w does not become singular, we need at least $D + C$ (where D =dimension of the feature vector and C =number of classes) number of samples which in general is practically not possible [11]. To address this issue, an approach called PCA+LDA is harnessed in the proposed system, where

a D -dimensional data is first reduced using PCA to an M -dimensional data and then to a L -dimensional data using LDA, $L \ll M < D$. It may be noted that the optimal number of features (L) required in our system is selected using the NCSV measure. The overall steps involved in the feature reduction stage is listed in Algorithm 2.

Algorithm 2 Feature reduction using PCA+LDA

Require: Feature matrix: F_M of size $N \times D$
Ensure: Reduced feature matrix: F_r of size $N \times L$

- ▷ Function $pca()$ and $lda()$ reduce the dimension using PCA and LDA respectively
- 1: Choose a dimension M ▷ Reduced dimension using PCA
- 2: $F'(N \times M) \leftarrow pca(F_M, M)$
- 3: Select a dimension L using NCSV measure ▷ Reduced dimension using LDA
- 4: $F_r(N \times L) \leftarrow lda(F', L)$
- 5: Output the reduced matrix F_r

Classification based on MDE-ELM

Extreme learning machine (ELM)

Extreme learning machine (ELM) is the most simple and efficient learning algorithm for training the single-hidden layer feed-forward neural networks (SLFNs) which avoids the limitations of gradient based learning schemes [8]. It has achieved dramatic successes in solving problems like multi-label classification problems and regression tasks. In contrast to conventional learning schemes such as BP, SVM and LS-SVM, ELM learns faster with better generalization performance [7]. In ELM, the hidden node parameters (the input weights and hidden biases) are randomly assigned, while the output weights of SLFNs are mathematically calculated by a simple inverse operation of the hidden layer output matrix.

Given N distinct training samples (x_j, t_j) , where $x_j = [x_{j1}, x_{j2}, \dots, x_{jL}]^T \in R^L$ and $t_j = [t_{j1}, t_{j2}, \dots, t_{jC}]^T \in R^C$, the hidden node number n_h and an activation function $\phi(\cdot)$, the ELM algorithm can be expressed as follows.

1. Generate hidden node parameters randomly (w_i^h, b_i) , $i = 1, 2, \dots, n_h$.
2. Compute the hidden layer output matrix \mathbf{H} .
3. Compute the output weight matrix $w^o = \mathbf{H}^\dagger \mathbf{T}$

Here, $w_i^h = [w_{i1}^h, w_{i2}^h, \dots, w_{iL}^h]^T$ represents the weight vector that links between i^{th} hidden neuron and the input neurons, $w_i^o = [w_{i1}^o, w_{i2}^o, \dots, w_{iC}^o]^T$ indicates the weight

vector that connects the i^{th} hidden neuron and the output neurons, and b_i is the bias of the i^{th} hidden neuron. \mathbf{H}^\dagger indicates the Moore-Penrose (MP) generalized inverse of matrix \mathbf{H} . The size of \mathbf{H} , w^o and \mathbf{T} are $N \times n_h$, $n_h \times C$ and $N \times C$ respectively. The smallest norm LS solution is unique and has the minimum norm among all the LS solutions. As the solution of ELM is obtained using an analytical method without iteratively tuning parameters, it converges faster than other traditional learning algorithms.

Modified DE algorithm

Differential evolution (DE) is a simple and effective population based meta-heuristic approach for global searching of optimization problems [3, 19]. The performance of DE is strongly influenced by its mutation strategy, crossover operation and control parameters. As a consequence, a significant amount of works have been proposed to improve its search performance and it has been reported that DE outperforms GA and PSO on various benchmark functions [9]. However, the standard DE faces problems of premature convergence at local optima and stagnation. Therefore, the recent trend is to improve the search performance of DE by means of novel strategies for mutation and parameter controlling. In this study, a novel mutation and random scale factor strategy is proposed to improve the performance of DE and hence, it is referred as modified DE (MDE). The stepwise description of the proposed MDE algorithm is as follows.

DE Initialization Randomly initialize the L -dimensional parameter vectors in a population of size N_p as $\{S_{j,It} | j = 1, 2, \dots, N_p\}$ with $S_{j,It} = [S_{1,j,It}, S_{2,j,It}, \dots, S_{L,j,It}]$, where It denotes the generation number.

Mutation For each target vector $S_{j,It}$, generate the mutant vector using the proposed mutation strategy as

$$V_{j,It} = S_{j,It} + f_s(S_{best,It} - S_{j,It}) + f_s(S_{best,It} - S_{r_1^j,It}) \quad (4)$$

where, r_1^j is a random integer between 1 to N_p and different from index j . $S_{best,It}$ denotes the best parameter vector having best fitness at generation It and f_s is the scaling factor which helps in scaling the difference vectors. In basic DE, the difference vector is scaled by a constant f_s . In the proposed scheme, however, f_s is set to change randomly using the following equation

$$f_s = 0.75 + [rand(.) / 4] \quad (5)$$

where $rand(.)$ is a uniformly distributed random number within the range $[0, 1]$.

Crossover Form a trial vector $U_{j,It} = [U_{1,j,It}, U_{2,j,It}, \dots, U_{L,j,It}]$ for the j^{th} target vector $S_{j,It}$ using binomial crossover as

$$U_{d,j,It} = \begin{cases} V_{d,j,It} & \text{if } randb(d) \leq C_r \text{ or } d = d_{rand} \\ S_{d,j,It} & \text{else} \end{cases}, \quad d = 1, 2, \dots, L \quad (6)$$

where, $randb(d)$ is the d^{th} evaluation of a uniform random number generator with outcome in $[0, 1]$, $d_{rand} \in [1, 2, \dots, L]$ is a randomly chosen index and $C_r \in [0, 1]$ is the crossover constant.

Selection Evaluate the fitness of the target and the trial vector and check the following condition to find the solution for next generation (i.e., $It = It + 1$)

$$S_{j,It+1} = \begin{cases} U_{j,It} & \text{if } f(U_{j,It}) \leq f(S_{j,It}) \\ S_{j,It} & \text{if } f(U_{j,It}) > f(S_{j,It}) \end{cases} \quad (7)$$

Here, $f(.)$ is the objective function which is to be minimized. Repeat the above procedure until a termination criterion gets satisfied.

Proposed evolutionary extreme learning machine

Since ELM utilizes random input weights and hidden biases, it leads to two critical issues [24, 46]: (i) high requirement of hidden neurons for which ELM responds slowly to unknown testing data and (ii) causing an ill-conditioned hidden layer output matrix \mathbf{H} in presence of large hidden neurons which induces poor generalization performance.¹

To overcome such issues, few research efforts have been reported in past years where population-based optimization schemes such as genetic algorithms (GA) [20], differential evolution (DE) [46] and PSO [24] are used to optimize the hidden node parameters of ELM. However, in this study, a new approach MDE-ELM by combining the modified DE (MDE) algorithm with ELM is proposed to enhance the performance of the proposed scheme compared to existing schemes. In this, MDE is used to optimize the hidden node parameters, whereas, MP generalized inverse is utilized to

¹Note: Condition number is shown to be an effective qualitative measure to find the conditioning of a matrix [44]. It may be noted that an ill-conditioned system has large condition number, while a well-conditioned system has small condition number. The 2-norm condition number of the matrix \mathbf{H} can be calculated as,

$$\mathcal{K}_2(\mathbf{H}) = \sqrt{\frac{\lambda_{max}(\mathbf{H}^T \mathbf{H})}{\lambda_{min}(\mathbf{H}^T \mathbf{H})}} \quad (8)$$

where, $\lambda_{max}(\mathbf{H}^T \mathbf{H})$ and $\lambda_{min}(\mathbf{H}^T \mathbf{H})$ denotes the largest and smallest eigenvalues of matrix $\mathbf{H}^T \mathbf{H}$.

analytically find the solution. It is worth mentioning here that the MDE algorithm searches global optima by considering both root-mean squared error (RMSE) and norm of the output weights of SLFNs which ensures in improving the generalization performance and the conditioning of the SLFN. The proposed MDE-ELM is stepwise listed as follows.

- (a) Randomly initialize all the parameter vectors in the population between [-1,1] such that each vector comprises a set of input weights and hidden biases as

$$S_j = \left[w_{11}^h, w_{12}^h, \dots, w_{1L}^h, w_{21}^h, w_{22}^h, \dots, w_{2L}^h, w_{n_h1}^h, w_{n_h2}^h, \dots, w_{n_hL}^h, b_1, b_2, \dots, b_{n_h} \right] \tag{9}$$

- (b) For each vector, evaluate the output weights and fitness. Here, for fitness evaluation, we compute the RMSE over

$$S_{j,It+1} = \begin{cases} U_{j,It} & \text{if } f(S_{j,It}) - f(U_{j,It}) > \epsilon f(S_{j,It}) \\ & \text{or } (|f(S_{j,It}) - f(U_{j,It})| < \epsilon f(S_{j,It}) \text{ and } \|w_{U_j}^o\| < \|w_{S_j}^o\|) \\ S_{j,It} & \text{otherwise} \end{cases} \tag{11}$$

where, $f(S_{j,It})$ and $f(U_{j,It})$ denotes the fitness value of the target vector j and its corresponding trial vector at iteration It respectively. $w_{S_j}^o$ and $w_{U_j}^o$ represents the output weights of target vector j and its corresponding trial vector, respectively. $\epsilon > 0$ is a user-defined tolerance rate.

- (e) To bound the input weights and biases in the range of [-1, 1], we use the following equation in the proposed MDE-ELM.

$$S_{d,j,It+1} = \begin{cases} -1 & \text{if } S_{d,j,It+1} < -1 \\ 1 & \text{if } S_{d,j,It+1} > 1 \end{cases}, \quad 1 \leq j \leq N_p, 1 \leq d \leq L \tag{12}$$

- (f) Repeat (c)–(e) until the point that the most extreme number of iterations are finished and obtain the optimal input weights and hidden biases.

The proposed scheme uses Eq. 11 to find the optimal input weights and hidden biases and hence, it tends to provide a lower value of norm of output weights of SLFNs. On the other hand, the smaller norm of the output weights leads to a smaller condition value of the output hidden matrix. To sum up, the proposed MDE-ELM offers the following advantages: (i) it improves the conditioning, (ii) it produces better generalization performance with a much more compact network. Compared to other gradient based methods and classical ELM, MDE-ELM approach does not need activation function to be differentiable.

Since the proposed PBDS includes techniques such as 2DPCA, PCA+LDA, and MDE-ELM, hereafter, in this

the validation set rather than the whole training set to overcome the overfitting issue. Hence, we can define fitness as

$$f() = \sqrt{\frac{\sum_{j=1}^{N_v} \left\| \sum_{i=1}^{n_h} w_i^o \phi(w_i^h \cdot x_j + b_i) - t_j \right\|_2^2}{N_v}} \tag{10}$$

where, N_v indicates the number of validation samples.

- (c) Find S_{best} of all the solutions in the population and generate the mutant vector V_j and trial vector U_j using Eqs. 4 and 6 respectively.
- (d) Update the vectors using the fitness value and the norm of the output weights and generate new population as follows:

paper, the proposed scheme is referred to as 2DPCA + PCA+LDA + MDE-ELM.

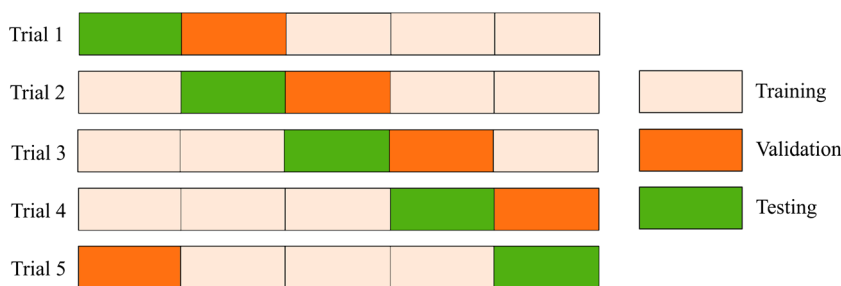
Experimental results and analysis

The parameters used and the statistical set up was kept similar to other competent schemes to derive relative comparisons.

Statistical set up

In order to validate the proposed scheme 2DPCA + PCA+LDA + MDE-ELM, simulation has been carried out on three different datasets, namely, DS-I, DS-II, and DS-III. For statistical analysis, cross-validation (CV) has been employed which avoids over-fitting problems. In this work, we have incorporated stratification into CV which splits the folds in such a way that each fold will have a similar class distribution. Figure 3 depicts the setting of a 5-fold CV for a single run. In each trial, one fold is used for testing, one for validation and the rests for training. The validation set is used to find the parameters of the MDE-ELM i.e., it helps us to know when to stop training. The test set is used to evaluate the performance in a run of five trials. Here, for DS-I, we employ 6-fold stratified cross validation (SCV) while for another two datasets, we select 5-fold SCV. The statistical setting for all the three datasets is given in Table 1. Here, the SCV procedure run for 10 times for three datasets.

Fig. 3 Illustration of 5-fold cross validation setting for a single run



Evaluation method

To decide whether the proposed scheme is effective or not, four different measures such as sensitivity (S_e), specificity (S_p), precision (P_r) and accuracy are computed. S_e is the fraction of pathological MR samples successfully predicted, while S_p is the fraction of healthy MR samples successfully predicted. However, accuracy (ACC) determines the fraction of the correctly predicted samples (both pathological and healthy) in the total number of testing samples. Moreover, to compare proposed MDE-ELM scheme against other schemes such as DE-ELM, PSO-ELM, basic ELM and BPNN, two parameters such as condition number and norm of output weights are used.

Experimental results

In the following, we discuss the results obtained at various stages of the proposed scheme.

Preprocessing and feature extraction results

In preprocessing stage, CLAHE is utilized which relies on the proper setting of its parameters. Here, the original MR image is divided into 64 contextual regions. The number of bins and the clip limit (β) are selected to be 256 and 0.01. The representative enhanced images corresponding to four original MR images are depicted in Fig. 4. From the figure, it is seen that the affected lesions are clear in the enhanced images than that of original images.

Next, 2DPCA algorithm is employed on the preprocessed images for feature extraction. In 2DPCA, the features are extracted using the projection vectors of the image scatter

matrix. If we use all the projection vectors for feature extraction of an image, then the total number of features will be too high. On the other hand, all projection vectors do not contain important information. Hence, a simple strategy based on NCSV measure is used in this study to select the optimal number of projection vectors (i.e., α). To test this strategy, we compute the NCSV values with varying number of projection vectors for all the three datasets as shown in Fig. 5. From the figure, it is seen that our algorithm needs more than 26 projection vectors for all the three datasets (in particular 23, 25 and 26 for DS-I, DS-II, and DS-III respectively) with a threshold of 0.8. Hence, we fix the α value as 26 in order to extract the salient features from the brain MRI of three datasets. As a consequence, the total number of features extracted from a single image is computed to be 6656 (i.e., 26×256). Here, the threshold value is determined experimentally.

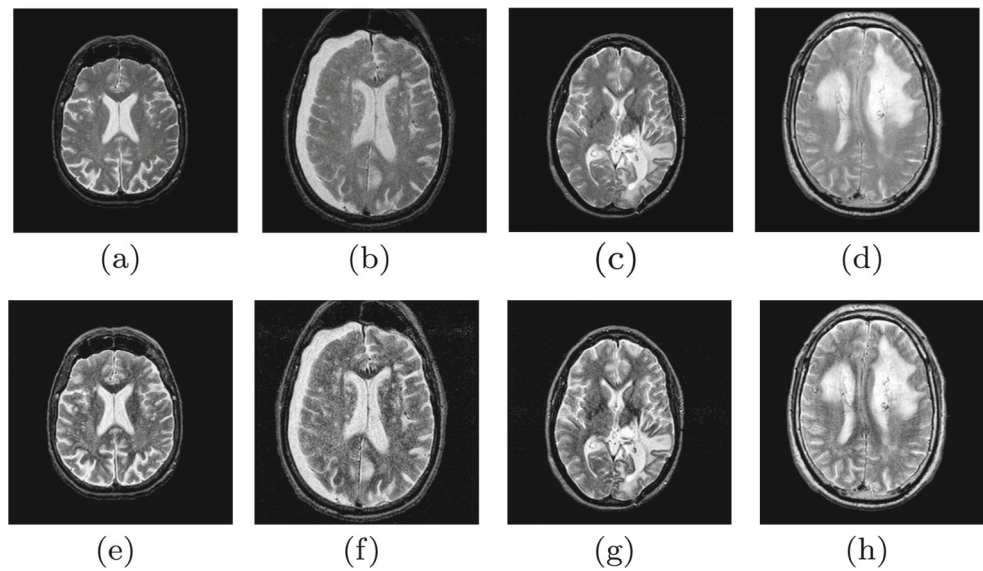
Feature reduction results

As the dimension of feature vector obtained by 2DPCA algorithm is much higher (i.e., 6656 features), we employ PCA+LDA to reduce the dimensionality. The number of significant features is obtained based on the NCSV values of different features. It has been observed that PCA preserves maximum information with more features compared to PCA+LDA. In this case, the threshold value for NCSV is set to 0.95. Moreover, the classification accuracy against the number of features for both PCA and PCA+LDA on three datasets is depicted in Fig. 6. From the figure, it is clear that PCA based scheme achieves higher accuracy with 14 features on all the three datasets, while PCA+LDA based scheme yields higher accuracy with only two features.

Table 1 Specification of three benchmark datasets [2, 13, 35]

Dataset	k -fold SCV	Total samples		Training		Validation		Testing	
		H	P	H	P	H	P	H	P
DS-I	6	18	48	12	32	3	8	3	8
DS-II	5	20	140	12	84	4	28	4	28
DS-III	5	35	220	21	132	7	44	7	44

Fig. 4 Preprocessing using CLAHE. Row 1 lists the original MR samples. Row 2 lists the corresponding preprocessed samples



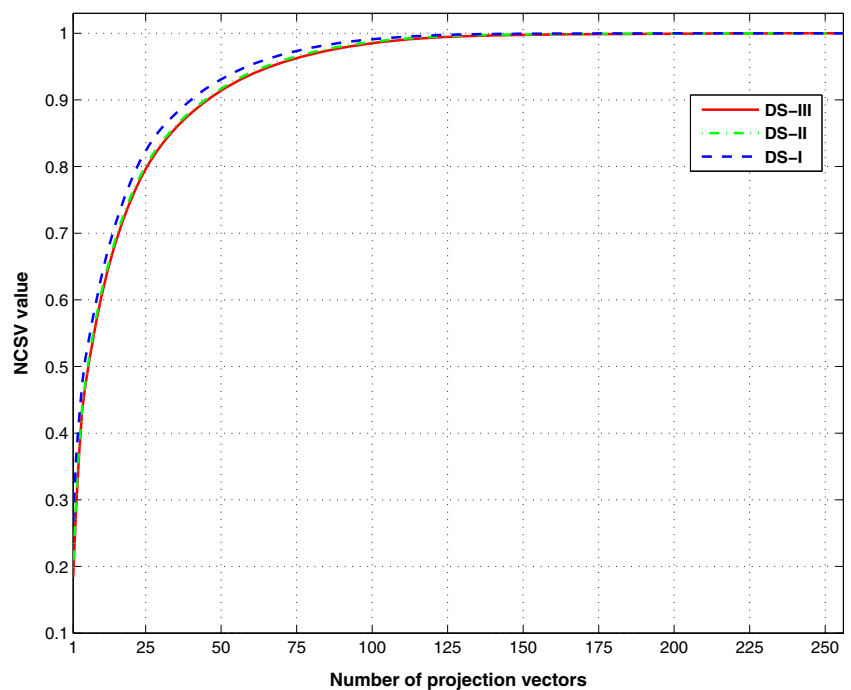
Classification results

The proposed system employs MDE-ELM for classification of MR images as healthy or pathological. Here, the performance of the proposed MDE-ELM is compared against other learning algorithms such as DE-ELM, PSO-ELM, ELM, and BPNN. The objective function is kept same for all the algorithms i.e., sigmoidal function and the inputs to the network are normalized into the range [-1,1]. It may be noted that we set 20 and 30 as the population size and the maximum number of iterations respectively for MDE-ELM, DE-ELM, and PSO-ELM algorithm. The ϵ value in the

proposed MDE-ELM is tested between a range [0.01, 0.2] at equally spaced intervals. However, it has been found that the proposed scheme achieves highest performance with ϵ value as 0.05. In case of PSO-ELM, the value of c_1 and c_2 are set as 2, while in DE-ELM, the crossover rate (C_r) and scaling factor (f_s) are set as 0.7 and 0.8 respectively.

Tables 2, 3 and 4 show the results obtained by MDE-ELM, DE-ELM, PSO-ELM, ELM and BPNN on three benchmark datasets. From the tables, it is clear that MDE-ELM outperforms others with less hidden neurons over all the datasets. It can also be noticed that basic DE-ELM earns perfect classification on DS-I and DS-II, however, it

Fig. 5 NCSV values with respect to different number of projection vectors for three datasets



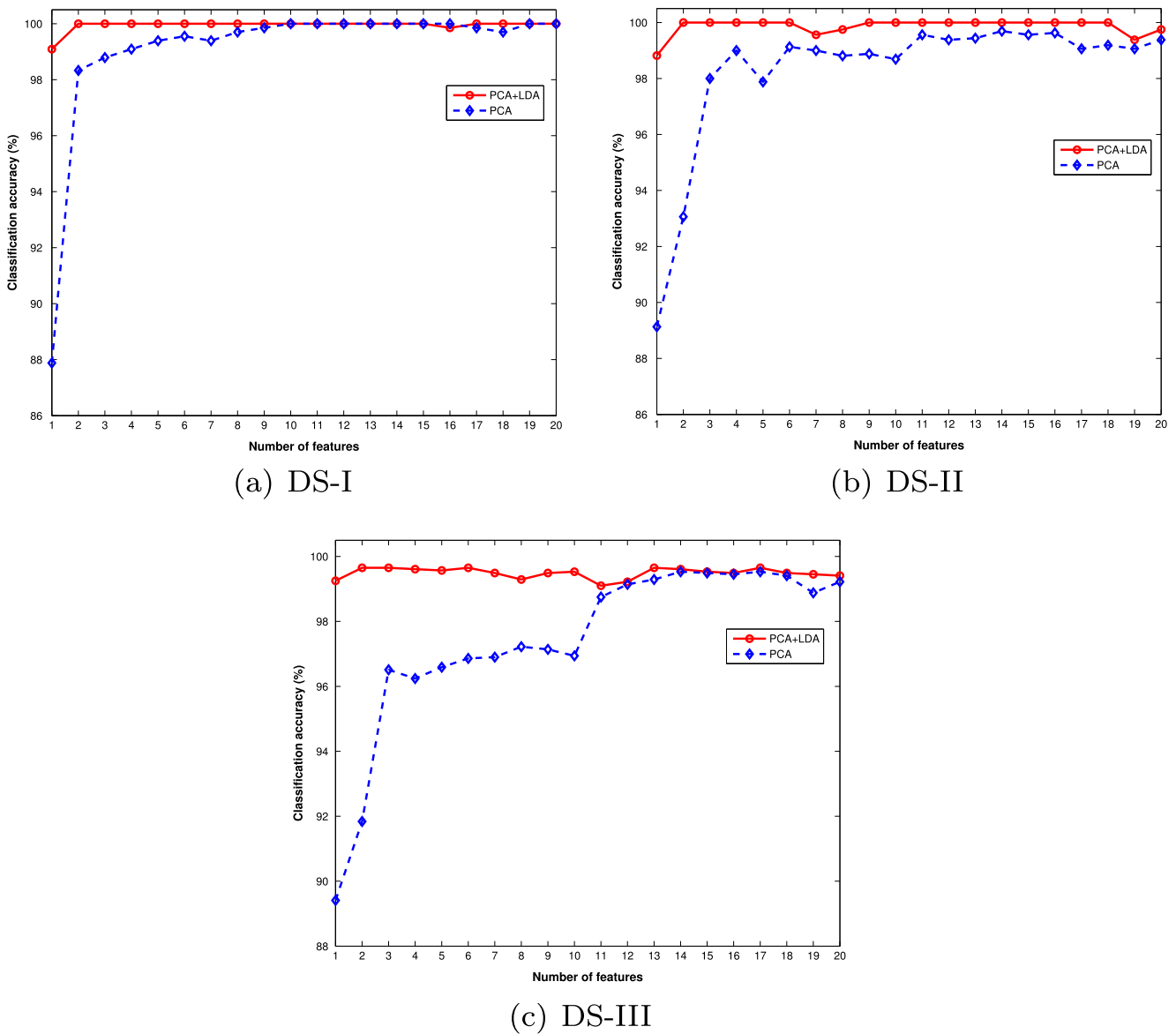


Fig. 6 Classification accuracy with respect to number of features for three datasets

earns comparable accuracy over DS-III. Compared to other algorithms, standard ELM demands more hidden neurons.

Further it is observed that the condition value of the matrix **H** obtained by MDE-ELM, DE-ELM and PSO-ELM

Table 2 Performance comparison of different algorithms on DS-I

Classifiers	ACC (%)	Hidden neurons (n_h)	Norm	Condition number (\mathcal{K}_2)
BPNN	99.85	4	–	–
ELM	100.00	5	36.1176	5.1029e+03
PSO-ELM	100.00	3	25.5475	101.5889
DE-ELM	100.00	3	19.0114	72.9316
MDE-ELM	100.00	3	13.3845	50.9073

Proposed scheme is highlighted in bold

Table 3 Performance comparison of different algorithms on DS-II

Classifiers	ACC (%)	Hidden neurons (n_h)	Norm	Condition number (\mathcal{K}_2)
BPNN	99.88	4	–	–
ELM	100.00	5	63.9829	2.1353e+03
PSO-ELM	99.94	3	22.1282	83.1536
DE-ELM	100.00	3	14.1658	55.5117
MDE-ELM	100.00	3	10.2688	46.5197

Proposed scheme is highlighted in bold

Table 4 Performance comparison of different algorithms on DS-III

Classifiers	ACC (%)	Hidden neurons (n_h)	Norm	Condition number (\mathcal{K}_2)
BPNN	99.37	4	-	-
ELM	99.49	5	112.5837	6.9975e+03
PSO-ELM	99.61	3	37.7973	173.3338
DE-ELM	99.53	3	22.9009	107.4050
MDE-ELM	99.65	3	14.6638	77.0868

Proposed scheme is highlighted in bold

algorithm is much smaller compared to the conventional ELM. Therefore, it is proved that the network trained by all these algorithms are highly well-conditioned compared to basic ELM. Further, their corresponding norm values are much smaller than basic ELM and hence, these algorithms tend to have better generalization performance compared to traditional ELM. It can be seen that the smaller norm value of w^o leads to a smaller condition value of matrix **H**. Compared with PSO-ELM and DE-ELM, the MDE-ELM obtains smaller condition and norm values. Therefore, it can be concluded that the proposed algorithm (MDE-ELM) can have better generalization performance with a compact network structure. It is worth mentioning here that the results reported in the tables are the average values of 50

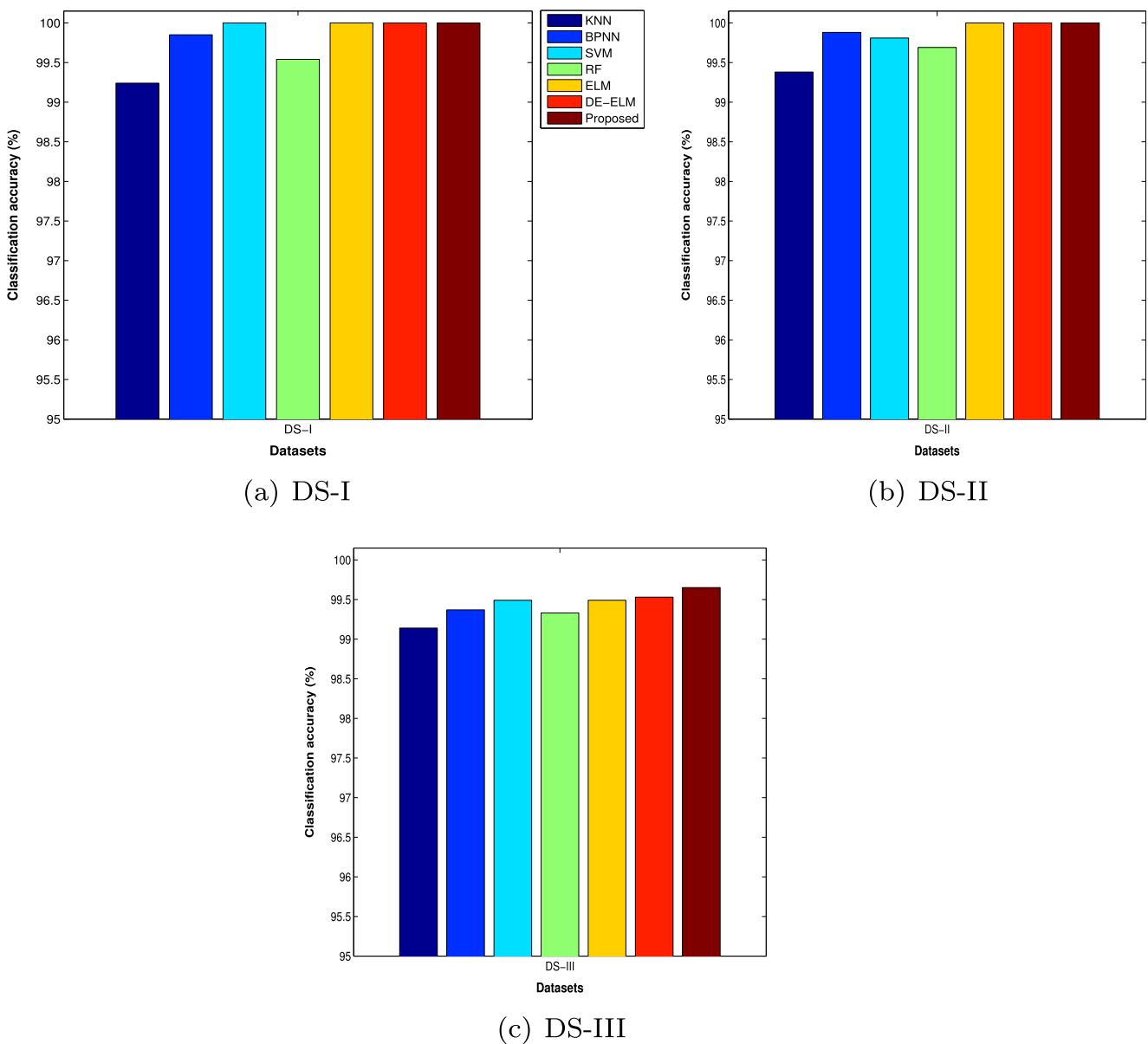


Fig. 7 Classification accuracy achieved by different classifiers for three datasets

Table 5 Correctly classified samples of the proposed scheme on DS-III

Run	F-1	F-2	F-3	F-4	F-5	Total	ACC(%)
1	51 (51)	50 (51)	51 (51)	51 (51)	51 (51)	254 (255)	99.61
2	51 (51)	51 (51)	51 (51)	51 (51)	50 (51)	254 (255)	99.61
3	50 (51)	51 (51)	51 (51)	51 (51)	51 (51)	254 (255)	99.61
4	51 (51)	51 (51)	51 (51)	51 (51)	51 (51)	255 (255)	100.00
5	51 (51)	50 (51)	51 (51)	51 (51)	51 (51)	254 (255)	99.61
6	51 (51)	51 (51)	51 (51)	51 (51)	51 (51)	255 (255)	100.00
7	51 (51)	51 (51)	50 (51)	51 (51)	51 (51)	254 (255)	99.61
8	51 (51)	51 (51)	51 (51)	49 (51)	51 (51)	253 (255)	99.22
9	51 (51)	51 (51)	51 (51)	51 (51)	50 (51)	254 (255)	99.61
10	50 (51)	51 (51)	51 (51)	51 (51)	51 (51)	254 (255)	99.61
Total						2541 (2550)	99.65

$x(y)$ indicating x brain images are correctly classified out of y brain images

trials and the parameters of all the schemes are determined through experimental evaluation.

Moreover, to prove the efficacy of the suggested MDE-ELM classifier, accuracy comparison is made against other classifiers like BPNN, KNN, random forest (RF), and SVM classifier on all the three datasets and the results are depicted in Fig. 7. For DS-I, KNN, BPNN, SVM, RF, ELM and DE-ELM yield an accuracy of 99.24%, 99.85%, 100.00%, 99.54%, 100.00% and 100.00% respectively; however, these classifiers obtain an accuracy of 99.38%, 99.88%, 99.81%, 99.69%, 100.00% and 100.00% respectively on DS-II. The accuracies yielded by KNN, BPNN, SVM, RF, ELM and DE-ELM are 99.14% 99.37%, 99.49%, 99.33%, 99.49%, and 99.53% respectively on DS-III. While MDE-ELM earns ideal classification on DS-I and DS-II datasets and an accuracy of 99.65% on DS-III dataset. This shows that the proposed algorithm outperforms all other classifiers

in DS-III and able to provide ideal results in other two datasets.

Table 5 indicates the number of correctly classified MR images obtained by the proposed scheme (2DPCA+PCA+LDA + MDE-ELM) over DS-III in each trial of a $10 \times k$ -fold SCV. It is found that the proposed scheme can successfully classify 2541 MR images out of 2550 samples (2200 pathological and 350 healthy MR images). In particular, 2195 pathological samples are successfully classified by our scheme and the rest five samples are misclassified to healthy class. However, the proposed system successfully predicts 347 healthy MR images and rest three samples are misclassified to pathological class. From these results, the sensitivity (S_e), specificity (S_p) and precision values (P_r) of the proposed scheme are computed as 99.82%, 98.57% and 99.77%, respectively which are shown in Table 6.

Table 6 Classification performance (%) of the proposed scheme based on PCA and PCA+LDA over three datasets

Dataset	Scheme No. of features	2DPCA+PCA+MDE-ELM 14	2DPCA+PCA+LDA+MDE-ELM 2
DS-I	S_e	100.00	100.00
	S_p	100.00	100.00
	P_r	100.00	100.00
	ACC	100.00	100.00
DS-II	S_e	99.79	100.00
	S_p	99.00	100.00
	P_r	99.86	100.00
	ACC	99.69	100.00
DS-III	S_e	99.64	99.82
	S_p	98.86	98.57
	P_r	99.73	99.77
	ACC	99.53	99.65

Table 7 Comparison against other competent PBDSs on three standard datasets

Existing PBDSs	Feature size	Run	ACC (%)		
			DS-I	DS-II	DS-III
DWT + SVM + POLY [1]	4761	5	98.00	97.15	96.37
DWT + PCA + BPNN + SCG [32]	19	5	100.00	98.29	97.14
DWT + PCA + FNN + SCABC [40]	19	5	100.00	98.93	97.81
DWT + PCA + FNN + ACP SO [38]	19	5	100.00	98.75	97.38
DWT + PCA + KSVM [39]	19	5	100.00	99.38	98.82
WPSE + GEPSVM [31]	16	10	99.85	99.62	98.78
WPTE + GEPSVM [31]	16	10	100.00	100.00	99.33
WPTE + FNN + RCBBO [21]	16	10	100.00	100.00	99.49
WE + HMI + GEPSVM [37]	14	10	100.00	99.56	98.63
DWT + PCA + ADBRF [13]	13	10	100.00	99.18	98.35
FRFE + WTT + SVM [23]	12	10	100.00	99.69	98.98
DTCWT + VE + GEPSVM [22]	12	10	100.00	99.75	99.25
FRFE + WTT + DP-MLP + ARCBBO [35]	12	10	100.00	99.19	98.24
RT + PCA + LS-SVM [2]	9	5	100.00	100.00	99.39
DWT + PCA + <i>k</i> -NN [5]	7	5	98.00	97.54	96.79
FPCNN + DWT + PCA + FNN [4]	7	10	100.00	98.88	98.43
SWT + PCA + GEPSVM [30]	7	10	100.00	99.62	99.02
WE + NBC [45]	7	10	92.58	91.87	90.51
DWT + PCA + LDA + RF [12]	7	10	100.00	99.75	99.14
MBD + SLFN + IPSO [42]	5	10	100.00	98.19	98.08
2DPCA + PCA + MDE-ELM	14	10	100.00	99.69	99.53
2DPCA + PCA+LDA + MDE-ELM (Proposed)	2	10	100.00	100.00	99.65

Proposed scheme is highlighted in bold

Comparison to PCA based PBDS

To test the effectiveness of PCA+LDA approach over PCA, another experiment is done over three datasets. The performances of both the schemes, namely, 2DPCA+ PCA + MDE-ELM and 2DPCA+ PCA+LDA + MDE-ELM are listed in Table 6. It may be noticed that the proposed 2DPCA+ PCA+LDA + MDE-ELM scheme achieves better sensitivity, precision and accuracy than 2DPCA+ PCA + MDE-ELM over all the datasets with a relatively less number of features. Moreover, 2DPCA + PCA+LDA MDE-ELM obtains slightly less specificity than 2DPCA+ PCA + MDE-ELM in DS-III. However, it is worth addressing here that the CAD system with higher sensitivity values leads to have better performance. Therefore, it can be concluded that the proposed 2DPCA+ PCA+LDA + MDE-ELM scheme holds greater potential in taking accurate clinical decisions.

Comparison to existing PBDSs

To benchmark the performance of the suggested scheme in context of the number of number of features required and

classification accuracy, extensive comparison with twenty existing schemes has been done over three datasets and is shown in Table 7. It is found that most of the earlier PBDSs yield ideal classification on DS-I; however, three PBDSs such as RT + PCA + LS-SVM [2], WPTE + FNN + RCBBO [21] and WPTE + GEPSVM [31] offer ideal classification on DS-II. Further, there is no PBDS available which can yield perfect classification over DS-III. However, our proposed PBDS obtains higher accuracy i.e., 99.65% compared to other PBDSs with a minimum number of features. Since MDE-ELM is used as classifier, the proposed system earns better generalization performance and responds faster to unknown testing data.

From the experiments, it has been observed that the proposed system has been tested on three openly accessible datasets accommodating images from patients during the late and middle stages of diseases, but a larger dataset with images from all stages of diseases can be validated to achieve better generalization performance. The present study deals with solving a two-class classification problem, however solving a multi-class brain disease classification problem is more challenging. Further, MDE demands more

parameter to tune, hence there exists a scope to investigate on an optimization scheme which may need less number of parameters.

Conclusions and future work

This paper proposed an improved pathological brain detection system based on 2DPCA and an evolutionary ELM. In the proposed PBDS, 2DPCA is used for feature extraction followed by a PCA+LDA approach for feature reduction. Thereafter, a novel learning algorithm called MDE-ELM is introduced to perform classification of MRI brain which offers several advantages over traditional classifiers. The goal of using MDE in MDE-ELM is to optimize the hidden node parameters of standard ELM. The performance of the proposed scheme is evaluated on three standard datasets and the experimental results confirm that the effectiveness of the proposed scheme in improving classification accuracy compared to the existing schemes. Further, the number of features required is shown to be much less than others.

The proposed MDE-ELM algorithm can be tested over real regression and classification problems. Despite the merits of the proposed PBDS, it has been observed that the PBDS is benchmarked on three accessible datasets which are smaller in size; hence, a larger dataset collected online will further prove its effectiveness. Further, the images in the chosen datasets are assembled from the last and the middle stage of the diseases, images collected during all the stages need to be validated. In future, it would be interesting to hybridize ELM with other metaheuristic algorithms like grey wolf optimizer (GWO), firefly algorithm (FA), gravitational search algorithm (GSA) etc. In addition, harnessing deep learning algorithms for analyzing 3D MR images is another possible future work.

Compliance with Ethical Standards

Conflict of interests We have no conflicts of interest.

Ethical approval This article does not contain any studies with human participants or animals performed by any of the authors.

References

1. Chaplot, S., Patnaik, L. M., and Jagannathan, N. R., Classification of magnetic resonance brain images using wavelets as input to support vector machine and neural network. *Biomedical Signal Processing and Control* 1(1):86–92, 2006.
2. Das, S., Chowdhury, M., and Kundu, K., Brain MR image classification using multiscale geometric analysis of ripplelet. *Prog. Electromagn. Res.* 137:1–17, 2013.
3. Das, S., and Suganthan, P. N., Differential evolution: a survey of the state-of-the-art. *IEEE Trans. Evol. Comput.* 15(1):4–31, 2011.
4. El-Dahshan, E. A., Mohsen, H. M., Revett, K., and Salem, A. B. M., Computer-aided diagnosis of human brain tumor through MRI: a survey and a new algorithm. *Expert Syst. Appl.* 41(11):5526–5545, 2014.
5. El-Dahshan, E. S. A., Honsy, T., and Salem, A. B. M., Hybrid intelligent techniques for MRI brain images classification. *Digital Signal Processing* 20(2):433–441, 2010.
6. Hazlett, H. C., Gu, H., Munsell, B. C., Kim, S. H., Styner, M., Wolff, J. J., Elison, J. T., Swanson, M. R., Zhu, H., Botteron, K. N., and et al., Early brain development in infants at high risk for autism spectrum disorder. *Nature* 542(7641):348–351, 2017.
7. Huang, G. B., Wang, D. H., and Lan, Y., Extreme learning machines: a survey. *International Journal of Machine Learning and Cybernetics* 2(2):107–122, 2011.
8. Huang, G. B., Zhu, Q. Y., and Siew, C. K., Extreme learning machine: theory and applications. *Neurocomputing* 70(1):489–501, 2006.
9. Islam, S. M., Das, S., Ghosh, S., Roy, S., and Suganthan, P. N., An adaptive differential evolution algorithm with novel mutation and crossover strategies for global numerical optimization. *IEEE Trans. Syst. Man Cybern. B (Cybernetics)* 42(2):482–500, 2012.
10. Johnson, K. A., and Becker, J. A., The Whole Brain Atlas. <http://www.med.harvard.edu/AANLIB/>.
11. Martínez, A. M., and Kak, A. C., PCA Versus LDA. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 23(2): 228–233, 2001.
12. Nayak, D. R., Dash, R., and Majhi, B., Classification of brain MR images using discrete wavelet transform and random forests. In: *5th National Conference on Computer Vision, Pattern Recognition, Image Processing and Graphics (NCVPRIPG)*, pp. 1–4. *IEEE*, 2015.
13. Nayak, D. R., Dash, R., and Majhi, B., Brain MR image classification using two-dimensional discrete wavelet transform and AdaBoost with random forests. *Neurocomputing* 177:188–197, 2016.
14. Nayak, D. R., Dash, R., and Majhi, B., Pathological brain detection using curvelet features and least squares SVM. *Multimedia Tools and Applications* 75:1–24, 2016.
15. Nayak, D. R., Dash, R., and Majhi, B., Stationary wavelet transform and adaboost with SVM based pathological brain detection in MRI scanning. *CNS & Neurological Disorders Drug Targets* 16:137–149, 2017.
16. Nayak, D. R., Dash, R., Majhi, B., and Prasad, V., Automated pathological brain detection system: a fast discrete curvelet transform and probabilistic neural network based approach. *Expert Syst. Appl.* 88:152–164, 2017.
17. Pizer, S. M., Johnston, R. E., Erickson, J. P., Yankaskas, B. C., and Muller, K. E., Contrast-limited adaptive histogram equalization: speed and effectiveness. In: *Proceedings of the 1st Conference on Visualization in Biomedical Computing*, pp. 337–345. *IEEE*, 1990.
18. Saritha, M., Joseph, K. P., and Mathew, A. T., Classification of MRI brain images using combined wavelet entropy based spider web plots and probabilistic neural network. *Pattern Recogn. Lett.* 34(16):2151–2156, 2013.
19. Storn, R., and Price, K., Differential evolution—a simple and efficient heuristic for global optimization over continuous spaces. *J. Glob. Optim.* 11(4):341–359, 1997.
20. Suresh, S., Babu, R. V., and Kim, H., No-reference image quality assessment using modified extreme learning machine classifier. *Appl. Soft Comput.* 9(2):541–552, 2009.
21. Wang, S., Li, P., Chen, P., Phillips, P., Liu, G., Du, S., and Zhang, Y., Pathological brain detection via wavelet packet tsallis entropy and real-coded biogeography-based optimization. *Fundamenta Informaticae* 151(1–4):275–291, 2017.

22. Wang, S., Lu, S., Dong, Z., Yang, J., Yang, M., and Zhang, Y., Dual-tree complex wavelet transform and twin support vector machine for pathological brain detection. *Appl. Sci.* 6(6):169, 2016.
23. Wang, S., Zhang, Y., Yang, X., Sun, P., Dong, Z., Liu, A., and Yuan, T. F., Pathological brain detection by a novel image feature—fractional Fourier entropy. *Entropy* 17(12):8278–8296, 2015.
24. Xu, Y., and Shu, Y., Evolutionary extreme learning machine based on particle swarm optimization. In: *International Symposium on Neural Networks*, pp. 644–652. Springer, 2006.
25. Xu, Y., Zhang, D., Yang, J., and Yang, J. Y., An approach for directly extracting features from matrix data and its application in face recognition. *Neurocomputing* 71(10):1857–1865, 2008.
26. Yang, G., Zhang, Y., Yang, J., Ji, G., Dong, Z., Wang, S., Feng, C., and Wang, Q., Automated classification of brain images using wavelet-energy and biogeography-based optimization. *Multimedia Tools and Applications* 75:1–17, 2015.
27. Yang, J., and Yang, J. Y., Why can LDA be performed in PCA transformed space? *Pattern Recognit.* 36(2):563–566, 2003.
28. Yang, J., Zhang, D., Frangi, A. F., and Yang, J. Y., Two-dimensional PCA: a new approach to appearance-based face representation and recognition. *IEEE transactions on pattern analysis and machine intelligence* 26(1):131–137, 2004.
29. Zhang, G., Wang, Q., Feng, C., Lee, E., Ji, G., Wang, S., Zhang, Y., and Yan, J., Automated classification of brain MR images using wavelet-energy and support vector machines. In: *2015 International Conference on Mechatronics, Electronic, Industrial and Control Engineering (MEIC-15)*, pp. 683–686, 2015.
30. Zhang, Y., Dong, Z., Liu, A., Wang, S., Ji, G., Zhang, Z., and Yang, J., Magnetic resonance brain image classification via stationary wavelet transform and generalized eigenvalue proximal support vector machine. *Journal of Medical Imaging and Health Informatics* 5(7):1395–1403, 2015.
31. Zhang, Y., Dong, Z., Wang, S., Ji, G., and Yang, J., Preclinical diagnosis of magnetic resonance (MR) brain images via discrete wavelet packet transform with Tsallis entropy and generalized eigenvalue proximal support vector machine (GEPSVM). *Entropy* 17(4):1795–1813, 2015.
32. Zhang, Y., Dong, Z., Wu, L., and Wang, S., A hybrid method for MRI brain image classification. *Expert Syst. Appl.* 38(8):10,049–10,053, 2011.
33. Zhang, Y., Lu, S., Zhou, X., Yang, M., Wu, L., Liu, B., Phillips, P., and Wang, S., 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, 2016.
34. Zhang, Y., Ranjan Nayak, D., Yang, M., Yuan, T. F., Liu, B., Lu, H., and Wang, S., Detection of unilateral hearing loss by stationary wavelet entropy. *CNS & Neurological Disorders-Drug Targets (Formerly Current Drug Targets-CNS & Neurological Disorders)* 16(2):122–128, 2017.
35. Zhang, Y., Sun, Y., Phillips, P., Liu, G., Zhou, X., and Wang, S., A multilayer perceptron based smart pathological brain detection system by fractional Fourier entropy. *J. Med. Syst.* 40(7):1–11, 2016.
36. Zhang, Y., Wang, S., Phillips, P., Dong, Z., Ji, G., and Yang, J., Detection of alzheimer's disease and mild cognitive impairment based on structural volumetric MR images using 3d-DWT and WTA-KSVM trained by PSOTVAC. *Biomedical Signal Processing and Control* 21:58–73, 2015.
37. Zhang, Y., Wang, S., Sun, P., and Phillips, P., Pathological brain detection based on wavelet entropy and Hu moment invariants. *Bio-medical Materials and Engineering* 26(s1):S1283–S1290, 2015.
38. Zhang, Y., Wang, S., and Wu, L., A novel method for magnetic resonance brain image classification based on adaptive chaotic PSO. *Prog. Electromagn. Res.* 109:325–343, 2010.
39. Zhang, Y., and Wu, L., An MR brain images classifier via principal component analysis and kernel support vector machine. *Prog. Electromagn. Res.* 130:369–388, 2012.
40. Zhang, Y., Wu, L., and Wang, S., Magnetic resonance brain image classification by an improved artificial bee colony algorithm. *Prog. Electromagn. Res.* 116:65–79, 2011.
41. Zhang, Y. D., Chen, S., Wang, S. H., Yang, J. F., and Phillips, P., Magnetic resonance brain image classification based on weighted-type fractional Fourier transform and nonparallel support vector machine. *Int. J. Imaging Syst. Technol.* 25(4):317–327, 2015.
42. Zhang, Y. D., Chen, X. Q., Zhan, T. M., Jiao, Z. Q., Sun, Y., Chen, Z. M., Yao, Y., Fang, L. T., Lv, Y. D., and Wang, S. H., Fractal dimension estimation for developing pathological brain detection system based on Minkowski-Bouligand method. *IEEE Access* 4:5937–5947, 2016.
43. Zhang, Y. D., Zhang, Y., Hou, X. X., Chen, H., and Wang, S. H., Seven-layer deep neural network based on sparse autoencoder for voxelwise detection of cerebral microbleed. *Multimedia Tools and Applications* 76:1–18, 2017.
44. Zhao, G., Shen, Z., Miao, C., and Man, Z., On improving the conditioning of extreme learning machine: a linear case. In: *7Th International Conference on Information, Communications and Signal Processing, ICICS*, pp. 1–5. IEEE, 2009.
45. Zhou, X., Wang, S., Xu, W., Ji, G., Phillips, P., Sun, P., and Zhang, Y., Detection of pathological brain in MRI scanning based on wavelet-entropy and naive Bayes classifier. In: *Bioinformatics and Biomedical Engineering*, pp. 201–209, 2015.
46. Zhu, Q. Y., Qin, A. K., Suganthan, P. N., and Huang, G. B., Evolutionary extreme learning machine. *Pattern Recognit.* 38(10):1759–1763, 2005.