Multimed Tools Appl (2018) 77:22705–22733
https://doi.org/10.1007/s11042-017-5281-x

Development of pathological brain detection system using Jaya optimized improved extreme learning machine and orthogonal ripplet-II transform

Deepak Ranjan Nayak[1] · **Ratnakar Dash**[1] · **Banshidhar Majhi**[1]

Received: 13 July 2017 / Revised: 11 September 2017 / Accepted: 4 October 2017 /
Published online: 27 November 2017

Abstract Pathological brain detection systems (PBDSs) have drawn much attention from researchers over the past two decades because of their significance in taking correct clinical decisions. In this paper, an efficient PBDS based on MR images is introduced that markedly improves the recent results. The proposed system makes use of contrast limited adaptive histogram equalization (CLAHE) and orthogonal discrete ripplet-II transform (O-DR2T) with degree 2 to enhance the quality of the input MR images and extract the features respectively. Subsequently, relevant features are obtained using PCA+LDA approach. Finally, a novel learning algorithm called IJaya-ELM is proposed that combines improved Jaya algorithm (IJaya) and extreme learning machine (ELM) for segregation of MR images as pathological or healthy. The improved Jaya algorithm is utilized to optimize the input weights and hidden biases of single-hidden-layer feedforward neural networks (SLFN), whereas one analytical method is used for determining the output weights. The proposed algorithm performs optimization according to both the root mean squared error (RMSE) and the norm of the output weights of SLFNs. Extensive experiments are carried out using three benchmark datasets and the results are compared against other competent schemes. The experimental results demonstrate that the proposed scheme brings potential improvements in terms of classification accuracy and number of features. Moreover, the proposed IJaya-ELM classifier achieves higher accuracy and obtains compact network architecture compared to conventional ELM and BPNN classifier.

Keywords Pathological Brain Detection System (PBDS) · Magnetic Resonance Imaging (MRI) · Orthogonal Discrete Ripplet-II Transform (O-DR2T) · Extreme Learning Machine (ELM) · Jaya Algorithm

✉ Deepak Ranjan Nayak
depakranjannayak@gmail.com

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

1 Introduction

Brain disease is one of the prime factors for causing death in people with different age groups. Different types of brain diseases exist such as neoplastic diseases (brain tumor), cerebrovascular diseases (stroke), degenerative diseases, and infectious diseases; some of these diseases may cause severe problems in the human brain and may prompt to death. This necessitates the development of automated computer-based decision systems which help the physicians to take correct and fast clinical decisions at early stages. Magnetic resonance imaging (MRI) is a common medical imaging modality used in PBDS because of its advantage of providing huge information about the soft tissues [5, 44, 53]. In addition, MRI is a non-invasive imaging modality compared other modalities including X-ray and CT scan. However, manual interpretation is difficult due to large data storage in MRI. Further, manual interpretation is costly, tedious and time-consuming process [8, 27, 31]. To overcome such issues, automated pathological brain detection systems (PBDSs) need to be developed to assist radiologists in taking accurate and quick decisions. PBDS utilizes various image processing and machine algorithms at different stages.

A significant amount of work has been done in developing various PBDSs in the past decades [10, 54, 65]. However, the development of an ideal PBDS is still challenging because of the difficulty in selecting proper algorithms for feature extraction, feature reduction, and classification. Further, these three phases should combinedly work in all cases regardless of the type of image modalities and the dataset size. Hence, PBDS remains an open problem for researchers. Our objective here is to enhance the performance of the PBDS with respect to existing systems for abnormality detection in the human brain.

It has been observed that discrete wavelet transform (DWT) is the mostly used feature extractor in PBDS since it analyzes images at several scales and handles one-dimensional (1D) singularities effectively. However, DWT can not handle two-dimensional (2D) singularities (edges of an image). That is, DWT is not able to capture curve like features effectively from the images. Therefore, finding a transform in order to capture 2D singularities is highly in demand. Further, classifiers like feed forward neural network (FNN) and support vector machine (SVM) are often used in earlier PBDSs because of their capability in separating nonlinear input patterns and predicting continuous functions. But, conventional gradient-based learning algorithms such as back-propagation (BP) and Levenberg-Marquardt (LM) used for the training of FNN causes many problems such as local minima, slower learning speed, and learning epochs. Few hybrid models have been designed with the help of population-based optimization strategies to overcome the limitations of traditional learning algorithms. Further, the traditional SVM classifier encounters higher computational complexity and performs poorly on large datasets [26]. 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.

Considering these concerns, we propose a novel PBDS with the following characteristics.

(a) Orthogonal discrete ripplet-II transform (O-DR2T) is used for feature extraction to capture 2D singularities along with a group of curves from MR images.
(b) To address the problems of traditional learning algorithms, a recently proposed learning algorithm known as extreme learning machine (ELM) is employed which provides faster learning speed and better generalization performance compared to conventional learning algorithms. However, ELM suffers from various limitations such as slower response speed on testing data, high requirement of hidden neurons, and ill-conditioned problem.

(c) To further enhance the performance of standard ELM, a hybrid learning algorithm based on improve Jaya optimization algorithm and ELM (IJaya-ELM) is proposed.
(d) To validate the proposed scheme, extensive experiments are carried out on three well-known datasets. In this context, the proposed PBDS is compared against other competent methods with respect to classification accuracy and number of features.

The remaining part of the article is organized as follows. Section 2 presents a review of the current PBDSs. Section 3 provides the materials used in the experiments. The proposed methodology adopted in the article is discussed in detail in Section 4. The statistical setting and pseudocode of the proposed scheme are presented in Section 5. In Section 6, a detail experimental evaluation and comparative analysis have been presented. Finally, the concluding remarks are presented in Section 7.

2 Related work

In the past years, a number of PBDSs have been reported in the literature for detection of brain diseases. MRI has been used as the imaging modality in almost all PBDSs. PBDSs can be broadly divided into two classes: direct-feature-based PBDS and indirect-feature-based PBDS depending on the type of features used. The former class uses coefficients of image transform as the key features. Indirect-feature-based PBDS, however, extracts features using statistical descriptors such as energy, entropy, mean and standard deviation from the coefficients. The PBDSs of the first category require feature transformation or selection techniques to get relevant feature sets. However, it is optional in case of the second category.

Chaplot et al. [5] were the forebears who proposed a PBDS with the help of 2D DWT features and two separate classifiers, namely, self-organizing map (SOM) and SVM. The authors in [23] have proposed a PBDS where Slantlet transform (ST) is employed for feature extraction and back-propagation neural network (BPNN) is used for classification. Later, El-Dahshan et al. [11] have suggested a hybrid approach with the assistance of 2D DWT and two separate classifiers such as k-nearest neighbor (k-NN) and feed forward back-propagation artificial neural network (FP-ANN). In order to reduce the feature dimensionality, they have applied principal component analysis (PCA). Further, with same features the authors in [53, 57, 59, 61], have proposed a variety of PBDSs. In these works, gradient-based and population-based optimization algorithms such as scaled conjugate gradient (SCG), particle swarm optimization (PSO), adaptive chaotic PSO (ACPSO), and scaled chaotic artificial bee colony (SCABC) are used to optimize the parameters of FNN, BPNN and kernel SVM (KSVM) classifier. Zhang et al. [60] have suggested a PBDS where DWT plus PCA based features are given to a KSVM classifier. The authors in [8] have derived features from Ripplet transform (RT) and reduced the feature dimensionality using PCA. Subsequently, they have applied least squares SVM (LS-SVM) for classification. In [10], the authors have utilized DWT and PCA for feature extraction and reduction prior to the employment of feedback pulse coupled neural network (FPCNN). Finally, they have applied FP-ANN for classification. Afterward, Wang et al. [42] offered a PBDS based on stationary wavelet transform (SWT), PCA and FNN. In this, the parameters of FNN classifier are optimized using artificial bee colony (ABC) and PSO and hence the schemes are coined as IABAP-FNN, ABC-SPSO-FNN, and HPA-FNN. In another work, Zhang et al. [62] have deployed weighted-type fractional Fourier transform (WFRFT) and PCA for feature extraction and reduction, respectively. For classification, they have applied generalized eigenvalue proximal SVM (GEPSVM) and twin SVM (TSVM). Later, Nayak et

al. [27] have proposed a PBDS with the support of 2D DWT and probabilistic PCA (PPCA). In this, AdaBoost with random forests (ADBRF) method is employed for classification. While in [51], authors have used SWT, PCA, and GEPSVM for feature extraction, reduction, and classification respectively. The authors in [25] have utilized HL_3 coefficients of 2D DWT as features and harnessed PCA+LDA strategy for dimensionality reduction. Later, Dash et al. [28] have utilized curvelet transform for feature extraction, however, PCA and LS-SVM is employed for feature reduction and classification. Chen et al. [63] have used Minkowski-Bouligand dimension (MBD) features for MR image classification. Edge detection is performed using Canny edge detector prior to feature extraction. Subsequently, an improved PSO based on three-segment particle representation, time-varying acceleration coefficient, and chaos theory (PSO-TTC) is proposed to train the single-hidden layer feedforward neural network. In [31], authors have used fast discrete curvelet transform for feature extraction after segmentation using simple pulse coupled neural network (SPCNN). Eventually, PNN is applied for classification.

Recent articles on PBDS have used feature descriptors like energy, entropy [13], mean and standard deviation etc., in the feature extraction stage. For example, in [36], the entropy values of the wavelet coefficients are used as features. A spider web plot and t-test strategy is used to select the significant features. Subsequently, probabilistic neural network (PNN) is employed for classification. Later, Yang et al. [48] have computed energy values from a level-3 DWT coefficients to serve as features. For classification, biogeography-based optimization (BBO) technique is integrated into SVM. In [52], a discrete wavelet packet transform (DWPT) based PBDS is proposed. Two different types of entropies namely, Shannon entropy (SE) [30] and Tsallis entropy (TE) are evaluated from the sub-bands and finally, GEPSVM is utilized to classify MR images as healthy or pathological. Furthermore, in [56], a hybrid BBO and PSO based method known as HBP for training of FNN is suggested. In this, wavelet entropy values are used as features. In [67], wavelet entropy (WE) and a Naive Bayes classifier (NBC) based PBDS is proposed. While in [50], wavelet energy and SVM are used. Thereafter, Zhang et al. [64] have used Tsallis entropy of DWPT for feature extraction and fuzzy support vector machine (FSVM) for classification. On the other hand, in [58], WE and Hu moment invariants (HMI) features are used followed by a GEPSVM+RBF classifier. Wang et al. [43] have proposed a PBDS based on a novel feature called fractional Fourier entropy (FRFE) which is the combination of FRFT and Shannon entropy. Two separate test such as Welch's t-test (WTT) and Mahalanobis distance (MD) is performed to select the relevant features. Subsequently, TSVM is employed for classification. Later, in [55], a PBDS based on FRFE features and multilayer perceptron (MLP) is proposed. In this, three pruning methods, namely, Bayesian detection boundaries (BDB), dynamic pruning (DP), and Kappa coefficient (KC) are utilized to get the optimal hidden neurons in MLP. Subsequently, an adaptive real coded BBO (ARCBBO) approach has been employed to update the weights of MLP. In [40], the authors have employed three varieties of binary PSO (BPSO) to select significant features from the entropy values of an 8-level DWT. PNN is deployed for classification. While in [39], the variance and entropy (VE) values of a dual-tree complex wavelet transform (DTCWT) are used as features. Both GEPSVM and TSVM are used as the classifier. Later, Nayak et al. [29] have used energy and entropy values of 2D-SWT as features. They have employed symmetric uncertainty ranking (SUR) filter for feature selection and AdaBoost with support vector machine (ADBSVM) for classification. The authors in [38] have employed wavelet packet Tsallis entropy (WPTE) for feature extraction and FNN with real-coded biogeography-based optimization (RCBBO) for classification.

The literature study reveals that in most PBDSs wavelet and its variants (like SWT, DWPT, DTCWT, etc.) have been frequently used for feature extraction. However, traditional

DWT suffers from many drawbacks such as limited directional selectivity and translation variance. SWT can resolve the translation variance issue; however, it leads to redundancy and is not able to capture higher dimensional singularities. DTCWT is efficient and less redundant, which offers more directional selectivities (i.e., six) compared to SWT and DWT. It can be concluded here that all these transforms are less capable of handling 2D singularities. Hence, further improvements in directional selectivity need to be done. Additionally, it has been observed that FNN and SVM are commonly used in many PBDSs which require more parameters to tune and are time-consuming. Further, most of the schemes have been validated on small datasets and shown higher accuracies; however, they perform poorly when evaluated on large datasets. Thus, there exists a scope to eradicate the shortcomings of the existing schemes.

To combat these issues, we have proposed an efficient PBDS to classify the MR images as healthy or pathological. The proposed PBDS uses O-DR2T for feature extraction due to its ability in capturing directional features (edges and curves). Subsequently, a PCA+LDA based approach is employed in order to determine the most significant feature set. Eventually, a hybrid learning algorithm IJaya-ELM for SLFN is introduced which offers many advantages such as avoiding local minima issue, better generalization capability, faster learning rate, and well-conditioned in contrast to classifiers like FNN, SVM, LS-SVM, ELM, etc. These improvements lead the proposed PBDS to a more robust and accurate system over other current existing schemes.

3 Datasets used

The performance of the proposed PBDS are tested on three benchmark datasets, namely, DS-66, DS-160, and DS-255 accommodating 66, 160 and 255 images, respectively. These datasets hold T2-weighted brain MR images of size 256×256 in axial view plane which is available in Medical School of Harvard University website [22]. Along with the healthy brain samples, the datasets DS-66 and DS-160 have samples from seven classes of diseases, namely, sarcoma, AD (Alzheimer's disease), AD plus visual agnosia (VA), glioma, meningioma, Huntington's disease (HD), and Pick's disease (PD). DS-255 contains four more diseases, viz., cerebral toxoplasmosis (CTP), multiple sclerosis (MS), herpes encephalitis (HE), and chronic subdural hematoma (CSH). Samples of all kind of MR images are shown in Fig. 1.

Out of 11 types of diseases, glioma, meningioma, and sarcoma are of brain tumor type; while CTP, MS, and HE are of infectious type. The diseases such as AD, AD plus VA, PD, and HD are called the degenerative diseases; whereas CSH is a cerebrovascular disease. The proposed work is a two-class classification problem (healthy or pathological) in which the pathological class contains images from all kinds of diseases.

4 Proposed methodology

This section describes the methods involved in the proposed PBDS. The proposed PBDS consists of four steps, namely, preprocessing, feature extraction, feature reduction, and classification. The input of the system is an MR image and the output is the class label (healthy or pathological). In the preprocessing step, contrast limited adaptive histogram equalization (CLAHE) is employed. In feature extraction step, we use orthogonal discrete ripplet-II transform to extract features and in feature dimension reduction step, PCA+LDA approach

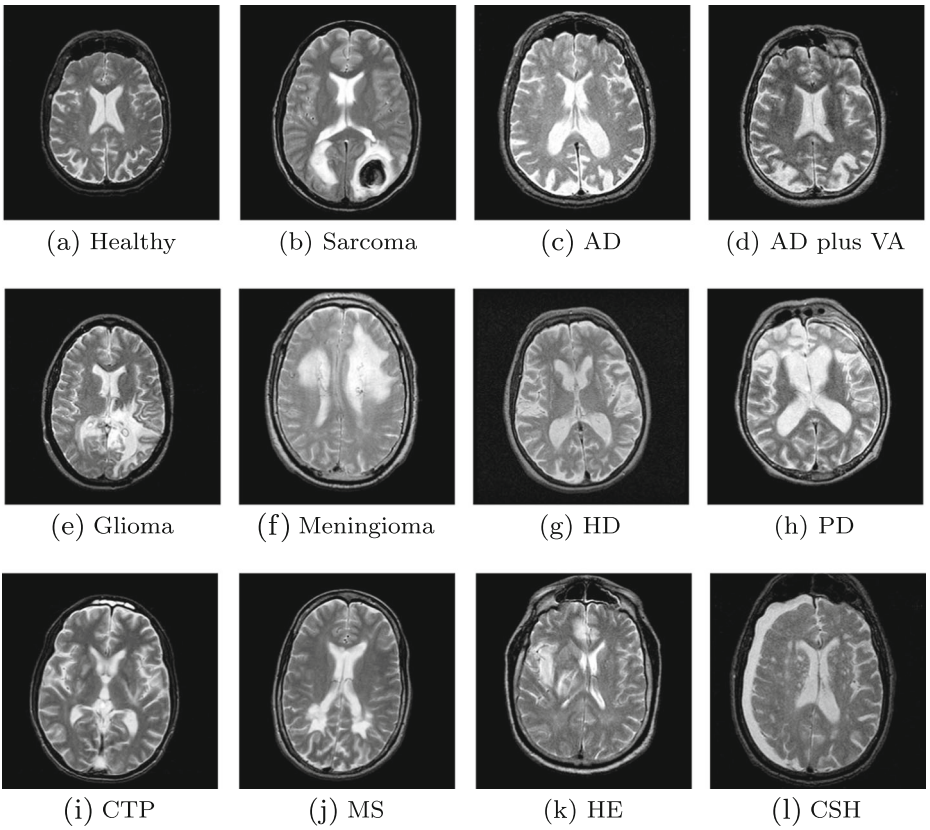

Fig. 1 Samples of T2-weighted brain MR images [8]

is harnessed. Thereafter, for classification, a hybrid learning algorithm IJaya-ELM is utilized, where improved Jaya (IJaya) algorithm is used to optimize the initial weights and biases of the SLFN. The proposed PBDS works in two parts, namely, offline learning and online prediction. The former part includes the training and evaluation process of the system; whereas, the latter part predicts a class label for the query MR image. The detailed block diagram of the proposed PBDS is depicted in Fig. 2. All the steps are delineated below.

4.1 Preprocessing based on CLAHE

It is observed that most of the images in the datasets considered in this work are of low-contrast. Therefore, for contrast enhancement of the images, a standard technique named contrast limited adaptive histogram equalization (CLAHE) is employed. CLAHE initially evaluates a histogram of gray values in a contextual region centered around each pixel and then, it allocates a value to each pixel intensity within the display range [32]. 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.

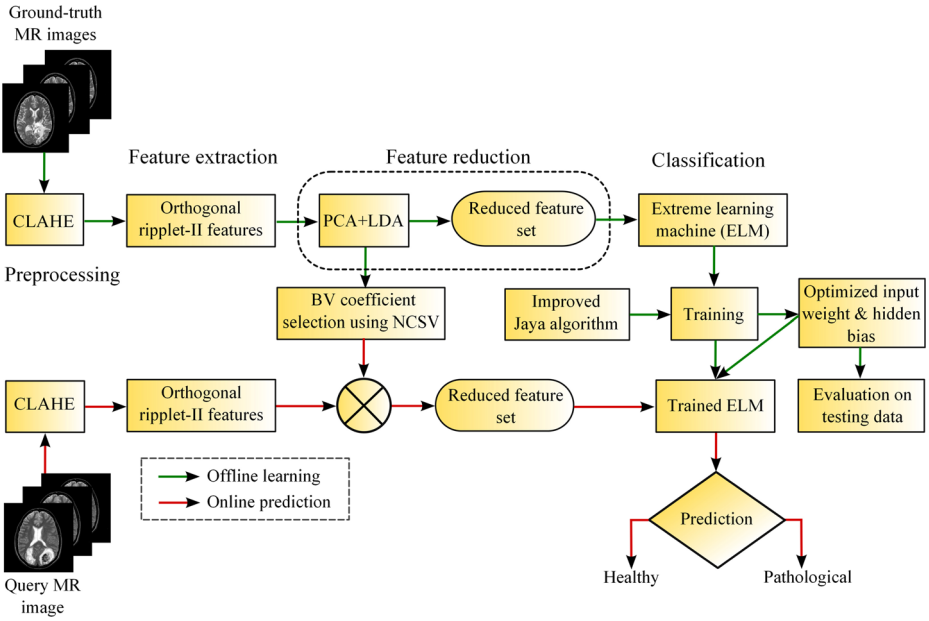

Fig. 2 Detailed block diagram of the proposed PBD system

4.2 Feature extraction based on O-DR2T

Fourier transform has been found to be less suitable for feature extraction in images as it loses the time information and can not handle 1D singularities. Hence, this transform fails to provide efficient representation of images that contains edges, however, it works well for only smooth images. In contrast, wavelet transform performs better in representing 1D singularities (i.e., point singularities). But conventional wavelet transform is not capable of representing 2D singularities along arbitrarily shaped curves. In order to resolve the problem that conventional wavelet suffers from, another transform called ridgelet transform was introduced which is based on Radon transform [2, 9]. Ridgelet holds great potential in representing line singularities (i.e., it is capable of extracting lines of arbitrary orientation), but it is not able to handle 2D singularities. Thereafter, first generation curvelet transform based on multiscale ridgelet was proposed by Candes and Dohono in [3] to resolve the 2D singularities along smooth curves. Later on, they proposed the second generation curvelet transform [1] which is simple, fast, and less redundant than the former one. Because of the capabilities like multiresolution, more directional selectivity, anisotropy, and localization, it has drawn attentions over last decades. The anisotropic property guarantees solving 2D singularities along C^2 curves and to accomplish this, curvelet utilizes a parabolic scaling law [4]. However, the reason behind the selection of parabolic scaling is not clear. In order to resolve this issue, a new transform called as ripplet-I transform is proposed which generalizes the scaling law [12, 46]. In general, ripplet-I transform generalizes the curvelet transform by adding two parameters such as support c and degree d. When $c = 1$ and $d = 2$, ripplet-I transform becomes curvelet transform. These two parameters provide ripplet-I transform with anisotropy capability of representing 2D singularities along arbitrarily shaped curves. Then, they proposed ripplet-II transform [45] based on generalized Radon

transform (GRT) [6, 7] to further improve the capability of representing 2D singularities. It satisfies the properties like multiresolution, localization, good directionality, and flexibility. Moreover, compared to wavelet and ridgelet transform, ripplet-II has the fastest decay in coefficients and for which the sparser representation of images having edges is possible. Another variant of ripplet-II transform known as orthogonal ripplet-II transform generates even more sparse feature vectors than ripplet-II transform which is crucial for classification task. Therefore, it has been leveraged in applications like texture classification and image retrieval [45]. As the orthogonal ripplet-II transform is efficient in representing edges and textures than other conventional transforms and the affected regions in MR images contain edges and textures of arbitrary shapes, it is used as a feature extraction tool in this work.

4.2.1 Ripplet-II transform

Given a 2D function $g(x, y)$, the continuous ripplet-II transform in polar coordinates (ρ, α) is defined as

$$RT2_g(s, t, d, \theta) = \int\int \bar{\psi}_{s,t,d,\theta}(\rho, \alpha) g(\rho, \alpha) \rho \, d\rho \, d\alpha \tag{1}$$

where, $g(\rho, \alpha)$ is the polar coordinate conversion of $g(x, y)$, $\psi_{s,t,d,\theta} : \mathbb{R}^2 \to \mathbb{R}^2$ is known as ripplet-II function and $\bar{\psi}$ is the complex conjugate of ψ. The ripplet-II function is stated as

$$\psi_{s,t,d,\theta}(\rho, \alpha) = s^{-1/2} \varphi((\rho \cos^d((\theta - \alpha)/d) - t)/s) \tag{2}$$

where $\varphi : \mathbb{R} \to \mathbb{R}$ is a smooth univariate wavelet function, and $s > 0$, $t \in \mathbb{R}$, $d \in \mathbb{N}$ and $\theta \in [0, 2\pi)$ indicates scale, translation, degree and orientation parameters, respectively. By tuning these parameters, ripllet-II transform can capture structural information along arbitrary curves. Using (1) and (2), we have

$$RT2_g(s, t, d, \theta) = \langle \varphi_{s,t}(r), GR_d[g] \rangle \tag{3}$$

where $GR_d[g]$ is the GRT of function g and is defined as

$$GR_d(r, \theta) = \int\int g(\rho, \alpha) \delta(r - \rho \cos^d((\alpha - \theta)/d)) \rho \, d\rho \, d\alpha \tag{4}$$

The GRT can also be evaluated using Fourier transform [45]. Equation (3) indicates that ripplet-II transform is the inner product between GRT and 1D wavelet. It can also be represented as

$$g(\rho, \alpha) \xrightarrow{GRT} GR_d[g](r, \theta) \xrightarrow{1D-WT} RT2_g(s, t, d, \theta) \tag{5}$$

which defines that ripplet-II transform in two steps: first compute GRT of g and then compute 1D WT of the GRT of g.

The discrete version of ripplet-II transform (DR2T) can be defined as

$$g(\rho, \alpha) \xrightarrow{DGRT} GR_d[g](r, \theta) \xrightarrow{1D-DWT} RT2_g(s, t, d, \theta) \tag{6}$$

in which the discrete GRT (DGRT) of g is first computed and subsequently, the 1D discrete WT (DWT) of the DGRT of g is computed. The computing procedure for discrete ripplet-II transform becomes more simpler when $d = 2$. In this case, the GRT is dubbed as 'parabolic Radon transform' and is defined as follows [45]

$$GR_2(r, \theta) = 2\sqrt{r} R[g(\rho'^2, 2\alpha')](\sqrt{r}, \theta/2) \tag{7}$$

where, $R[g(\rho,\alpha)](r,\theta)$ is the classical Radon transform (CRT) in polar coordinates. However, in general, the GRT of function g for $d > 0$ takes the form in Fourier domain as

$$GR_d^F(r,\theta) = 2\sum_{n=-\infty}^{+\infty}\left[\int_r^{\infty}\int g(\rho,\alpha)e^{-\text{in }\alpha}d\alpha \times (1-(r/\rho)^{2/d})^{-1/2} \times T_{nd}((r/\rho)^{1/d})d\rho\right]e^{\text{in }\theta} \tag{8}$$

where $T_n(.)$ denotes the Chebyshev polynomial of degree n.
In summary, the forward DR2T with $d = 2$ of an input image can be computed as follows:

(i) Convert the input function from Cartesian coordinates to polar coordinates i.e., $g(x,y)$ to $g(\rho,\alpha)$. Replace (ρ,α) by $(\rho'^2, 2\alpha')$ in $g(\rho,\alpha)$. Subsequently, generate a new image $g'(x,y)$ by interpolation after converting polar coordinates (ρ',α') to Cartesian coordinates (x,y). The variables x and y hold integer values.
(ii) Employ discrete CRT on $g'(x,y)$ that produces $R(r',\theta')$ and then substitute (r',θ') with $(\sqrt{r},\theta/2)$ in $R(r',\theta')$ as in (7). And obtain the DGRT coefficients $GR_2(r,\theta)$.
(iii) Apply 1D DWT to DGRT coefficients w.r.t. r and obtain the discrete ripplet-II coefficients.

The above substitution from (r',θ') to $(\sqrt{r},\theta/2)$ makes DR2T coefficients more sparser than others.

4.2.2 Orthogonal ripplet-II transform

Orthogonal ripplet-II transform is an extension of ripplet-II transform which is achieved by applying 2D WT to GRT coefficients in place of 1D WT along r and θ. The additional WT along angle θ helps in improving the sparsity of transform coefficients. The continuous orthogonal ripplet-II transform of the function g takes the form

$$RT2_g^{orth}(s,t_1,t_2,d) = 2\sum_{n=-\infty}^{+\infty}\int\int\frac{1}{s}\bar{\varphi}(\frac{r-t_1}{s})\bar{\varphi}(\frac{r-t_2}{s})\int_r^{\infty}\int g(\rho,\alpha)e^{-\text{in }\alpha}d\alpha \times(1-(r/\rho)^{2/d})^{-1/2} \times T_{nd}((r/\rho)^{1/d})\,d\rho e^{\text{in }\theta}dr\,d\theta \tag{9}$$

Now, the discrete orthogonal ripplet-II transform (O-DR2T) can be stated as

$$g(\rho,\alpha) \xrightarrow{DGRT} GR_d[g](r,\theta) \xrightarrow{2D-DWT} RT2_g^{orth}(s,t_1,t_2,d) \tag{10}$$

where the DGRT of g is first evaluated and thereafter 2D DWT of the DGRT of g is harnessed. It is worth mentioning that unlike DR2T, O-DR2T has no explicit direction parameter and for which it may lose the explicit directional information. However, in [45], it is shown that orthogonal ripplet-II transform supplies more sparser features of the images than wavelet, ridgelet and standard ripplet-II transform because of the replacement of 2D DWT with 1D DWT. Hence, in the proposed system, O-DR2T is used as feature extractor.

4.2.3 Feature generation

For each training input MR image, we apply O-DR2T and obtain the coefficients. Then, the transform coefficients are arranged in a feature vector of dimension D, where $D = m * n$, and m and n are the number of rows and columns of the image. This vector is calculated for

each training images and a feature matrix is formed finally. The implementation procedure of the feature generation is outlined in Algorithm 1.

Algorithm 1 Feature extraction using orthogonal discrete ripplet-II transform

Input: N input images: $g[x, y];\ 0 \leq x < m, 0 \leq y < n$
Output: Feature matrix: F_M of size $N \times D$

1: **for** each image $g[x, y] \in N$ **do**
2: Transform $g(x, y)$ into polar coordinates $g(\rho, \alpha)$ and substitute (ρ, α) with $(\rho^{'2}, 2\alpha^{'})$
3: Transform polar coordinates $(\rho^{'}, \alpha^{'})$ to Cartesian coordinates (x, y) and obtain another image $g'(x, y)$ by 2D bilinear interpolation
4: Compute 1D FFT of $g'(x, y)$, $G'(u, v)$ along θ (columns)
5: Compute $GR_d(r, \theta)$ in Fourier domain i.e., $GR_d^F(r, \theta)$ for $G'(u, v)$ and $d = 2$ using (8)
6: Compute inverse 1D FFT g_{inv} on $GR_d^F(r, \theta)$ along θ (columns)
7: Apply 2D DWT on g_{inv} along r and θ, and obtain the D-dimensional $(= m * n)$ coefficients
8: Arrange the coefficients in a vector of size $1 \times D$ and store in F_M
9: **end for**
10: Obtain a feature matrix F_M containing all vectors

4.3 Feature reduction based on PCA+LDA

It is noticed that the features generated by O-DR2T are of high dimension which leads high computational overhead and large storage space requirement. Therefore, application of dimensionality reduction techniques is of great importance. PCA is a frequently used feature dimension reduction technique which transforms high dimensional input data to a lower dimensional space while keeping maximum variations of the data [35]. 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 [49]. Further, to make sure that S_w does not become singular, at least $D + C$ (where, D=dimension of feature vector and C=number of classes) number of samples are required which in general is practically not possible [24]. To address this issue, a well-known method dubbed as PCA+LDA has been applied in this study, where a D-dimensional data is first reduced to an M-dimensional data using PCA and then reduced to a l-dimensional data using LDA, $l << M < D$.

In order to get a relevant feature set, we fist sort the eigenvalues of different features in decreasing order and then the normalized cumulative sum of variances (NCSV) corresponding to each feature is calculated. The NCSV value for j^{th} feature is defined as

$$NCSV(j) = \frac{\sum_{u=1}^{j} \alpha(u)}{\sum_{u=1}^{D} \alpha(u)} \quad ;\ 1 \leq j \leq D \tag{11}$$

where, $\alpha(u)$ represents the eigenvalue of the u^{th} feature and D denotes the dimensionality of the feature vector. Finally, a threshold value is set manually and the number of features for which the NCSV value surpasses the threshold are selected. Relevant features selected are determined experimentally to have a maximal accuracy. It may be noted that the coefficients of the l eigenvectors (suitably called as basis vectors (BV)) corresponding to l large eigenvalues are retained if these l eigenvalues collectively satisfy the given threshold.

4.4 Classification based on IJaya-ELM

In this section, we first discuss the preliminaries of extreme learning machine (ELM) and Jaya algorithm, and thereafter present the proposed IJaya-ELM learning algorithm for single-hidden layer feedforward neural networks (SLFNs) in oder to classify the MR brain as healthy or pathological.

4.4.1 Extreme Learning Machine (ELM)

Single-hidden layer feedforward neural networks (SLFNs) have been shown to be used in many applications as they successfully approximate any continuous function and classify any disjoint region. To train the SLFNs, gradient-based learning algorithms such as Levenberg-Marquardt (LM) and backpropagation (BP) algorithm have been widely used. However, despite their popularity, these learning algorithms face various issues such as poor learning speed due to improper learning steps, getting trapped at local minima, requiring large number of iterations to obtain better learning performance, and overfitting [21]. A recently developed learning algorithm called extreme learning machine (ELM) avoids the limitations of gradient based learning schemes. ELM has also the potential for solving multi-class classification and regression tasks [19, 20]. In contrast to other conventional learning algorithms such as BP, SVM and LS-SVM, ELM learns faster with better generalization performance. In ELM, the hidden node parameters (the input weights and hidden biases) are randomly assigned, while the output weights of SLFNs are analytically determined by simple inverse operation of the hidden layer output matrix. ELM is discussed below mathematically.

Given N distinct training samples (x_i, t_i), where $x_i = [x_{i1}, x_{i2}, \ldots, x_{il}]^T \in R^l$ and $t_i = [t_{i1}, t_{i2}, \ldots, t_{iC}]^T \in R^C$, the SLFNs having n_h hidden nodes and activation function $\phi(.)$ can be represented as

$$\sum_{i=1}^{n_h} w_i^o \phi(x_j) = \sum_{i=1}^{n_h} w_i^o \phi(w_i^h \cdot x_j + b_i) = o_j, \quad j = 1, 2, \ldots, N \tag{12}$$

Here, $w_i^h = \left[w_{i1}^h, w_{i2}^h, \ldots, w_{il}^h\right]^T$ represents the weight vector that links between i^{th} hidden neuron and the input neurons, $w_i^o = \left[w_{i1}^o, w_{i2}^o, \ldots, w_{iC}^o\right]^T$ 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. The SLFNs can approximate these N samples with zero error, i.e., $\exists\ w_i^h$, w_i^o, and b_i such that

$$\sum_{i=1}^{n_h} w_i^o \phi(w_i^h \cdot x_j + b_i) = t_j, \quad j = 1, 2, \ldots, N \tag{13}$$

Now, (13) can be represented in matrix form as

$$\mathbf{H}w^o = \mathbf{T} \tag{14}$$

where,

$$\mathbf{H}(w_1^h, w_2^h, \ldots, w_{n_h}^h, b_1, b_2, \ldots, b_{n_h}, x_1, x_2, \ldots, x_N) = \begin{bmatrix} \phi(w_1^h \cdot x_1 + b_1) & \ldots & \phi(w_{n_h}^h \cdot x_1 + b_{n_h}) \\ \vdots & \ldots & \vdots \\ \phi(w_1^h \cdot x_N + b_1) & \ldots & \phi(w_{n_h}^h \cdot x_N + b_{n_h}) \end{bmatrix}_{N \times n_h},$$

$$w^o = \begin{bmatrix} w_1^{oT} \\ \vdots \\ w_{n_h}^{o\ T} \end{bmatrix}_{n_h \times C} \quad and\ \mathbf{T} = \begin{bmatrix} t_1^T \\ \vdots \\ t_N^T \end{bmatrix}_{N \times C}$$

Here, $\mathbf{H}$ denotes the hidden layer output matrix. Now, the output weights w^o can be analytically determined by finding the smallest norm least square (LS) solution of the above linear system (12) as

$$\hat{w^o} = \mathbf{H}^{\dagger}\mathbf{T} \tag{15}$$

where, $\mathbf{H}^{\dagger}$ indicates the Moore-Penrose (MP) generalized inverse of matrix $\mathbf{H}$ and with this method ELM leads better generalization performance [68]. 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.

4.4.2 Jaya algorithm

Jaya algorithm is a recent optimization algorithm developed by Rao [33] and has been gaining attractions of the researchers for its simplicity and robustness. Jaya algorithm is shown to provide better results than other optimization algorithms [34, 41]. Unlike other population-based optimization algorithms, it does not need any algorithm-specific parameters; however, it needs the common control parameters like population size, generation number, etc. The conceptual idea behind this scheme is that it always moves the obtained solution toward the best solution and avoids the worst solution.

Suppose $f(s)$ is the objective function to be minimized or maximized. At any iteration k, let there are n number of candidate solutions (i.e., $j = 1, 2, \ldots, n$) each having dimension (or number of variables) d (i.e., $d = 1, 2, \ldots, m$). If $s_{jd}(k)$ denotes the value for j^{th} solution in d^{th} dimension during iteration k, then its modified value can be obtained as

$$s'_{jd}(k) = s_{jd}(k) + r_{1d}(k)(s_{bestd}(k) - |s_{jd}(k)|) - r_{2d}(k)(s_{worstd}(k) - |s_{jd}(k)|) \tag{16}$$

where $s_{bestd}(k)$ represents the value for the *best* candidate solution in d^{th} dimension and $s_{worstd}(k)$ represents the value for the *worst* candidate solution in d^{th} dimension during iteration k. It is worth mentioning that the candidate *best* and *worst* are the best and worst solution having best and worst fitness values in the entire population of an iteration. $r_{1d}(k)$ and $r_{2d}(k)$ indicate two random numbers in dimension d during k^{th} iteration which lie in the interval [0, 1]. $s'_{jd}(k)$ denotes the updated value of $s_{jd}(k)$. The term

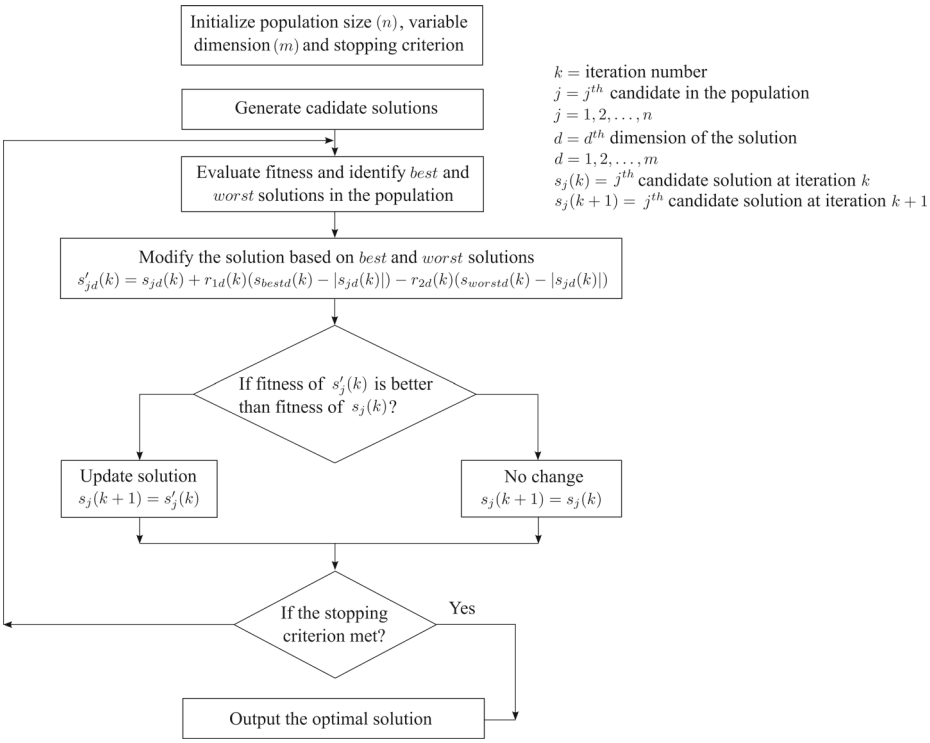

Fig. 3 Flow diagram of the Jaya algorithm

"$r_{1d}(k)(s_{bestd}(k) - |s_{jd}(k)|)$" defines that the solution tries to move toward the best solution and the term "$-r_{2d}(k)(s_{worstd}(k) - |s_{jd}(k)|)$" indicates that the solution tries avoid the worst solution. The modified $s'_{jd}(k)$ value is accepted if the functional value generated by it is better. The overall steps involved in Jaya algorithm are shown in Fig. 3.

4.4.3 Proposed improved extreme learning machine

Since ELM randomly chooses the input weights and hidden biases, it leads to two crucial problems [47, 66, 68]: (i) ELM needs more number of hidden neurons than conventional gradient based methods which make ELM respond slowly to unknown testing data, and (ii) ELM prompts to an ill-conditioned hidden layer output matrix **H** in presence of more hidden neurons which induces poor generalization performance. Condition number was found to be a good qualitative measure to find the conditioning of a matrix [66]. It indicates how close a system is to be ill-conditioned. It may be noted that an ill-conditioned system holds large condition number while a well-conditioned system holds small condition number. The 2-norm condition number of the matrix **H** can be calculated as,

$$\mathcal{K}_2(\mathbf{H}) = \sqrt{\frac{\lambda_{max}(\mathbf{H}^T\mathbf{H})}{\lambda_{min}(\mathbf{H}^T\mathbf{H})}} \tag{17}$$

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

In order to tackle these issues, few efforts have been made by researchers in the last decade using evolutionary algorithms (EAs) and swarm intelligence based algorithms since these algorithms have the benefits of global searching for optimization problems [18]. Zhu et al. [68] suggested a hybrid algorithm called evolutionary ELM (E-ELM), where a modified differential evolution (DE) algorithm is utilized to optimize hidden node parameters and MP generalized inverse is utilized to find the solution. They have shown that E-ELM provides faster learning speed and better generalization performance than other traditional algorithms, while it obtains much more compact network than ELM. However, E-ELM demands two additional parameters to tune, namely, the mutation factor and the crossover factor. Xu and Shu [47] introduced another evolutionary ELM based on PSO (PSO-ELM) to select the hidden node parameters which requires only one parameter to tune. They have added boundary conditions into conventional PSO to enhance the performance of ELM. Later, in [14], an improved PSO based ELM (IPSO-ELM) is proposed to find optimal SLFNs. In this, IPSO considers both the root mean squared error (RMSE) and the norm of output weights of validation set to obtain better convergence performance. Suresh et al. [37] have proposed a hybrid learning algorithm using real-coded genetic algorithm and ELM (RCGA-ELM) for no-reference image quality assessment. But, RCGA requires two genetic parameters such as crossover and mutation. While, Zhao et al. [66] have offered an input weight selection technique for improving the conditioning of ELM with the help of linear hidden neurons. With this technique, they have achieved numerical stability without degrading accuracy.

From the above literature, it is observed that different researchers have utilized optimization algorithms like GA and its variants, PSO and its variants, DE, etc., to find the optimal hidden node parameters. These techniques have their own advantages. However, they need proper tuning of their algorithm-specific parameters as it significantly influences the performance of the algorithms. Therefore, in order to resolve the problem of improper tuning, in this paper a parameter less based scheme known as Jaya algorithm is used. In addition, we propose a new scheme IJaya-ELM by combining the improved Jaya (IJaya) algorithm with the ELM. IJaya-ELM avoids the issues faced by existing methods in the recent literature. In this scheme, IJaya is harnessed to optimize the hidden node parameters and MP generalized inverse to analytically find the solution. It is worth mentioning here that the improved Jaya algorithm searches global optima considering both RMSE and norm of the output weights of SLFNs which on the other hand improve the generalization performance and conditioning of the SLFN. The main goal of IJaya is to minimize the norm of the output weights and to bound the hidden node parameters within a specific range in order to enhance the convergence performance of ELM. The steps of the proposed IJaya-ELM is delineated as follows:

(a) At first, initialize randomly all the candidate solutions in the population such that each candidate solution consists of a set of input weights and hidden biases as

$$s_j = \left[w^h_{11}, w^h_{12}, \ldots, w^h_{1l}, w^h_{21}, w^h_{22}, \ldots, w^h_{2l}, w^h_{n_h 1}, w^h_{n_h 2}, \ldots, w^h_{n_h l}, b_1, b_2, \ldots, b_{n_h} \right] \tag{18}$$

It may be noted that all the input weights and hidden biases are randomly initialized within a range of [-1,1].

(b) For each solution, evaluate the output weights and fitness. Here, we set the root-mean squared error (RMSE) on the validation set as the fitness rather than the whole training set in order to avoid the overfitting. The fitness can be defined as

$$f() = \sqrt{\frac{\sum_{j=1}^{N_v} || \sum_{i=1}^{n_h} w_i^o \phi(w_i^h \cdot x_j + b_i) - t_j||_2^2}{N_v}} \tag{19}$$

where, N_v indicates the number of validation samples.

(c) Find s_{best} and s_{worst} of all the solutions in the population and modify the solutions using (16).

(d) Update the solutions using the fitness value and the norm of the output weights and generate new population as follows:

$$s_j(k+1) = \begin{cases} s'_j(k) \text{ if } f(s_j(k)) - f(s'_j(k)) > \epsilon f(s_j(k)) \\ \quad \text{or } \left(|f(s_j(k)) - f(s'_j(k))| < \epsilon f(s_j(k)) \text{ and } ||w^o_{s'_j}|| < ||w^o_{s_j}|| \right) \\ s_j(k) \text{ otherwise} \end{cases} \tag{20}$$

where, $f(s_j(k))$ and $f(s'_j(k))$ denotes the fitness value of the candidate solution j and its corresponding modified solution during iteration k, respectively. $w^o_{s_j}$ and $w^o_{s'_j}$ represents the output weights generated by MP generalized inverse for candidate solution j and its corresponding modified solution, respectively. $\epsilon > 0$ is a user-defined tolerance rate.

(e) As given in the literature, all the input weights and biases should lie in the range of [-1, 1]. Therefore, the following equation is followed in the IJaya-ELM in order to deal with the solution out-of-bound issue

$$s_{jd}(k+1) = \begin{cases} -1 \text{ if } s_{jd}(k+1) < -1 \\ 1 \quad \text{if } s_{jd}(k+1) > 1 \end{cases}, \quad 1 \le j \le N_p,\ 1 \le d \le D \tag{21}$$

(f) Repeat (c)-(e) until the maximum number of iterations are over. Finally, the optimal input weights and hidden biases are obtained, and are employed on the testing data to find the performance of the system.

As the proposed scheme uses (20) to find the optimal input weights and hidden biases, it tends to provide the smaller 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. In general, the proposed IJaya-ELM has the following advantages: it has no algorithm-specific parameters, it improves the conditioning, and it produces better generalization performance with a much more compact network. Compared to other gradient based methods and classical ELM, the proposed approach does not need activation function to be differentiable.

The proposed PBDS involves techniques like O-DR2T, PCA+LDA, and IJaya-ELM, and hence it is referred to as O-DR2T + PCA+LDA + IJaya-ELM. The overall steps followed is articulated in Algorithm 2.

Algorithm 2 Implementation steps of the proposed PBDS

Offline learning:

1: **for** each ground truth image **do**
2: Enhance the contrast using CLAHE.
3: Apply O-DR2T with degree 2 on the enhanced image.
4: Obtain the O-DR2T coefficients and form a feature vector set of dimension D.
5: **end for**
6: Apply PCA+LDA approach to reduce the dimension of feature vector from D to l, where l is calculated from $NCSV$ measure. Retain the corresponding l basis vector (BV) coefficients.
7: Perform K-fold stratified cross validation on all the whole dataset and generate the training, validation and testing data
8: Train the ELM algorithm using improved Jaya algorithm and find the optimized input weights and hidden biases. In IJaya, evaluate the fitness RMSE on the validation set.
9: Calculate the output weights using the optimized input weights and hidden biases.
10: Evaluate the classification performance on the testing set.

Online prediction:

1: Load the unknown MR image into the proposed system.
2: Preprocess the query image with CLAHE.
3: Employ O-DR2T with degree 2 over the enhanced image.
4: Obtain the O-DR2T features and store it in a feature vector.
5: Find reduced feature set by multiplying the feature vector with the retained BV coefficients.
6: Feed the reduced feature set to the SLFN classifer trained by IJaya-ELM and predict the output label as healthy or pathological.

5 Experimental design and evaluation

In order to validate the proposed PBDS, simulation has been carried out on three different datasets, namely, DS-66, DS-160, and DS-255. For statistical analysis, cross-validation (CV) has been employed to avoid over-fitting problems. CV makes the classifier to generalize on independent datasets. In this work, we incorporate stratification into CV which splits the folds in such a manner that each fold will have a similar class distributions. Figure 4 depicts the setting of a 5-fold CV for a single run. In each trial, one fold is used for testing,

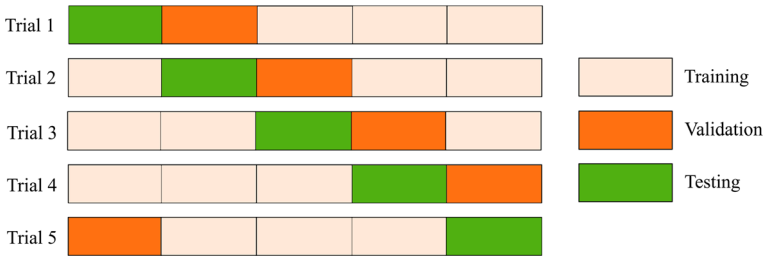

Fig. 4 Illustration of 5-fold cross validation setting for a single run

Table 1 Statistical setting of K-fold SCV for three benchmark datasets [8, 27, 55]

Dataset	K-fold SCV	Total samples		Training		Validation		Testing	
		H	P	H	P	H	P	H	P
DS-66	6	18	48	12	32	3	8	3	8
DS-160	5	20	140	12	84	4	28	4	28
DS-255	5	35	220	21	132	7	44	7	44

one for validation and the rests for training. The validation set is used to find the parameters of the IJaya-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. For DS-160 and DS-255, we select 5-fold SCV. But for DS-66 (18 healthy and 48 pathological) if we select 5-fold stratified cross validation (SCV), then each fold will have the different number of samples from two classes. Hence, for DS-66, we employ 6-fold (SCV). It is worth mentioning here that the statistical setting for all the three datasets is kept similar to the literatures as shown in Table 1. It may be noted that we run the SCV procedure 10 times on three datasets to avoid randomness.

Four different measures, namely, sensitivity (S_e), specificity (S_p), precision (P_r) and accuracy are used to evaluate the proposed system. S_e is the fraction of pathological MR samples correctly predicted by the model, while S_p is the fraction of healthy MR samples correctly predicted by the model. 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 the proposed IJaya-ELM scheme with other schemes such as Jaya-ELM, PSO-ELM, APSO-ELM, E-ELM and GA-ELM, two additional parameters, namely, condition number and norm of output weights are used.

6 Experimental results and analysis

The proposed system was implemented using MATLAB toolbox on a machine with 3.4 GHz processor, 8 GB RAM, and windows 10 OS. The parameters used and the statistical set up were kept similar to other competent schemes to derive relative comparisons.

6.1 Preprocessing and feature extraction results

The quality features of an MR image is dependent on the quality of input image. To enhance the original MR images CLAHE is utilized, which relies on the proper setting of its parameters. In the present case, the original MR image is divided into 64 contextual regions. The number of bins and the clip limit (β) are set as 256 and 0.01 respectively. It may be noted that uniform distribution scheme is selected for each region to obtain a flat histogram shape. The representative enhanced images corresponding to four original MR images are shown in Fig. 5. The affected regions in the enhanced images are more clear compared to the original images.

O-DR2T is applied to each of the preprocessed images and the features are extracted as the transform coefficients of O-DR2T. In this case, the number of levels in 2D DWT is set as 2 with Haar wavelet as the basis. As the images are of size 256×256, therefore the total number of features extracted from a single image are $256 * 256 = 65536$ which is huge in size.

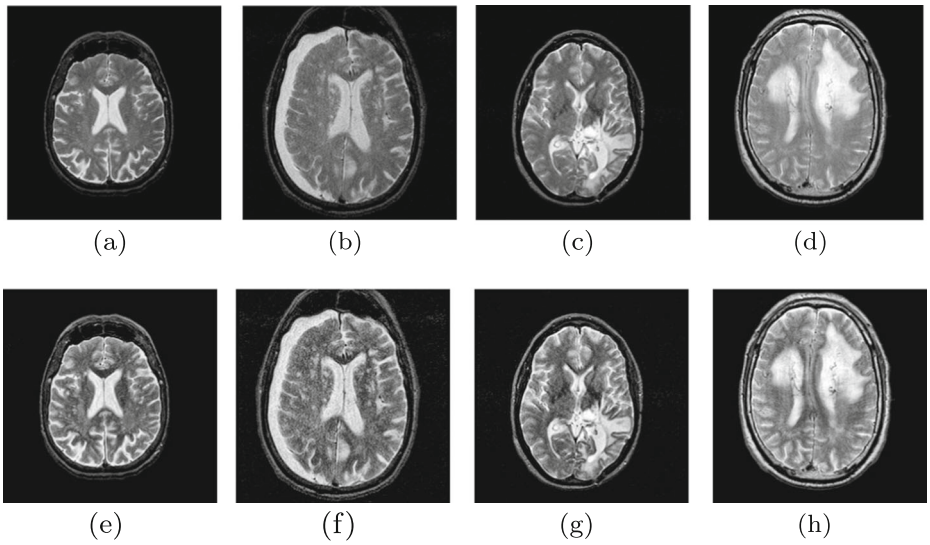

Fig. 5 Preprocessing using CLAHE. Row 1 lists the original MR samples. Row 2 lists the corresponding contrast enhancement using CLAHE

6.2 Feature reduction results

In order to attain better performance and make the classifier's job easier, the high dimensional O-DR2T features (65536 features) are reduced using PCA+LDA. 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, however, PCA+LDA, requires relatively less number of features. In particular, setting the threshold value for NCSV as 0.95, two features are considered from PCA+LDA and 15 features are considered from PCA separately. Additionally, the classification accuracy with respect to the number of features for both PCA and PCA+LDA on three datasets are shown in Fig. 6. From the figures, it is clear that PCA with 15 features and PCA+LDA with two features are providing higher results on all the three datasets. Therefore, it can be concluded that PCA+LDA approach is more suitable than only PCA.

6.3 Classification results

For classification of MR images as healthy or pathological, we employ a combined learning algorithm called IJaya-ELM for SLFN. In this section, first we have compared the performance of the proposed IJaya-ELM with other learning algorithms, namely, Jaya-ELM, GA-ELM, PSO-ELM, adaptive PSO-ELM (APSO-ELM), E-ELM, ELM and BPNN. The basic PSO-ELM along with a time varying inertia weight parameter is referred to as APSO-ELM. In IJaya-ELM, we use sigmoidal function as the activation function and normalize all the inputs to the network into the range [-1,1]. It may be noted that the population size and the maximum number of iterations for IJaya-ELM, Jaya-ELM, GA-ELM, PSO-ELM, APSO-ELM, and E-ELM algorithm is kept same i.e., 20 and 30 respectively. The ϵ value in IJaya-ELM is experimentally determined as 0.05. The parameters involved in different algorithms are experimentally determined and their values are listed below. In

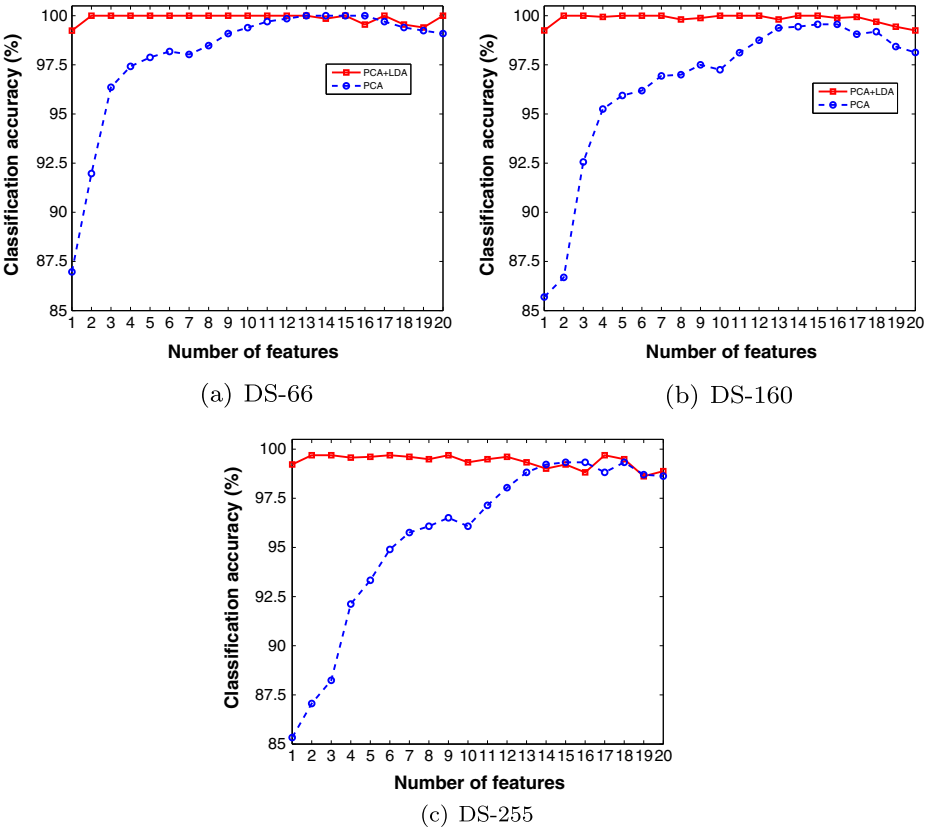

Fig. 6 Classification accuracy with respect to number of features for three datasets

case of PSO-ELM, the value of acceleration coefficients c_1 and c_2 are set as 2, while in E-ELM, the crossover rate (CR) and scaling factor (F) are set as 0.9 and 0.8 respectively. In APSO-ELM, the initial and final inertia parameters ω_1 and ω_2 are chosen as 0.4 and 0.9 respectively. For GA-ELM, we select the crossover rate and mutation rate as 0.7 and 0.1 respectively.

Table 2 Performance comparison of different classifiers on DS-66

Classifiers	ACC (%)	Hidden neurons (n_h)	Norm	Condition number ($\mathcal{K}_2$)
BPNN	100.00	4	–	–
ELM	100.00	5	42.4354	3.5919e+03
GA-ELM	100.00	3	32.7753	120.6437
E-ELM	100.00	3	22.3902	54.6665
PSO-ELM	99.70	3	30.3307	82.0004
APSO-ELM	100.00	3	20.7755	49.3035
Jaya-ELM	**99.85**	**3**	**24.2312**	**62.5218**
IJaya-ELM	**100.00**	**3**	**16.7405**	**31.4217**

Our proposed schemes are highlighted in bold

Table 3 Performance comparison of different classifiers on DS-160

Classifiers	ACC (%)	Hidden neurons (n_h)	Norm	Condition number ($\mathcal{K}_2$)
BPNN	99.88	5	–	–
ELM	99.94	6	89.1218	5.1528e+03
GA-ELM	99.88	3	27.7148	118.4492
E-ELM	100.00	3	12.9528	56.0894
PSO-ELM	100.00	3	17.2419	81.2390
APSO-ELM	100.00	3	14.2694	57.1167
Jaya-ELM	**100.00**	**3**	**14.4950**	**63.1756**
IJaya-ELM	**100.00**	**3**	**11.7418**	**44.3138**

Our proposed schemes are highlighted in bold

The performance of IJaya-ELM, Jaya-ELM, GA-PSO, PSO-ELM, APSO-ELM, E-ELM, ELM and BPNN on three benchmark datasets are reported in Tables 2, 3 and 4. From the tables, it is seen that IJaya-ELM obtains higher accuracy than others with less hidden neurons on all the datasets. Jaya-ELM achieves ideal accuracy on DS-160, while it earns smaller accuracy than IJaya-ELM on DS-66 and DS-255. Further, E-ELM earns better performance compared to others except IJaya-ELM and APSO-ELM outperforms PSO-ELM. It can also be seen that standard ELM demands more hidden neurons than other algorithms. Furthermore, it is observed that the condition value of the matrix $\mathbf{H}$ obtained by IJaya-ELM algorithm is smaller compared to others on all the datasets. The norm value of IJaya-ELM is also found to be less than others and therefore, it can have better generalization performance than traditional ELM and its variants. It is proved that the smaller norm value of w^o results in a smaller condition value of matrix $\mathbf{H}$. Further, among Jaya-ELM and IJaya-ELM, IJaya-ELM obtains smaller condition and norm values, and higher accuracy. Therefore, it can be concluded that the proposed algorithm (IJaya-ELM) can achieve better generalization performance with compact networks than others. The results reported in the tables are the average values of 50 trials.

Moreover, to demonstrate the effectiveness of the proposed IJaya-ELM algorithm with two features, accuracy comparison has been made with k-NN, random forest (RF), and SVM classifier along with BPNN, ELM, and Jaya-ELM on all the three datasets and the results are shown in Fig. 7. For DS-66, the accuracies earned by k-NN, BPNN, RF, SVM, ELM

Table 4 Performance comparison of different classifiers on DS-255

Classifiers	ACC (%)	Hidden neurons (n_h)	Norm	Condition number ($\mathcal{K}_2$)
BPNN	99.29	5	–	–
ELM	99.37	6	102.6683	4.0175e+03
GA-ELM	99.53	3	26.0816	137.5234
E-ELM	99.65	3	15.6426	74.2324
PSO-ELM	99.57	3	24.9681	113.3067
APSO-ELM	99.65	3	19.4705	98.0824
Jaya-ELM	**99.61**	**3**	**18.3876**	**78.2147**
IJaya-ELM	**99.69**	**3**	**12.8941**	**54.0704**

Our proposed schemes are highlighted in bold

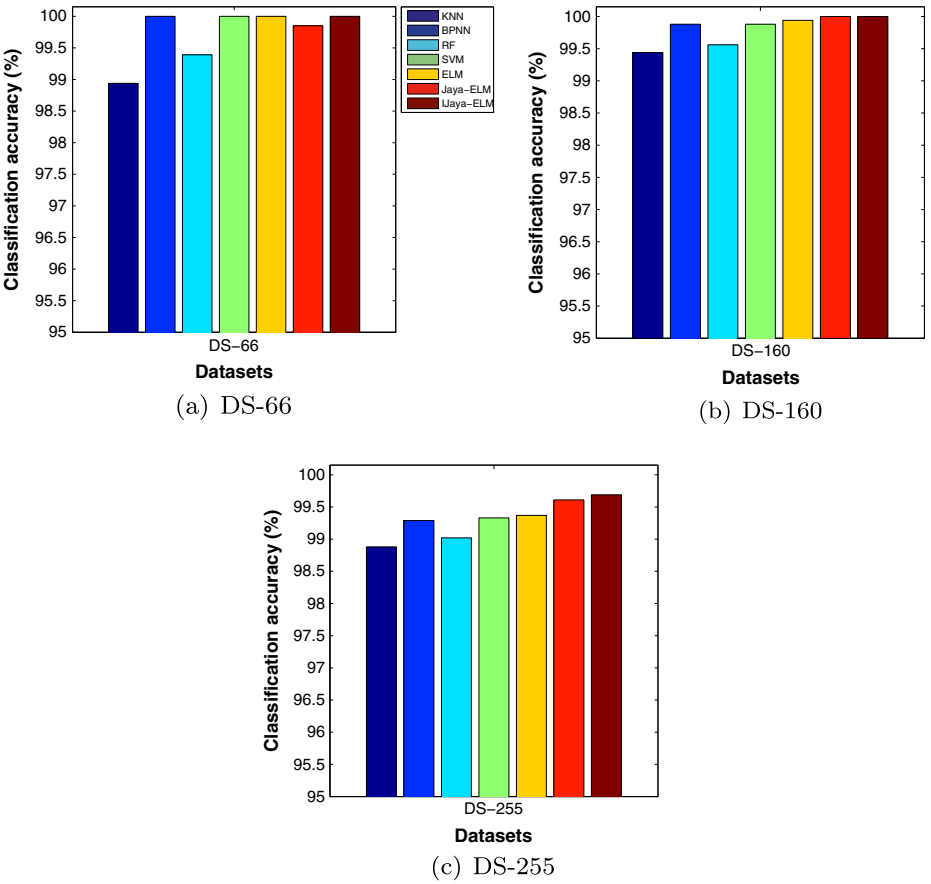

Fig. 7 Classification accuracy achieved by different classifiers on three standard datasets

and Jaya-ELM are 98.94%, 100.00%, 99.39%, 100.00%, 100.00%, and 99.85% respectively. The accuracies obtained by k-NN, BPNN, RF, SVM, ELM and Jaya-ELM for DS-160 are 99.44%, 99.88%, 99.56%, 99.88%, 99.94% and 100.00%, respectively; while they are 98.88%, 99.29%, 99.02%, 99.33%, 99.37%, and 99.61% respectively for DS-255. From these results, it can be concluded that IJaya-ELM earns ideal classification on DS-66 and DS-160 datasets and an accuracy of 99.69% on DS-255 dataset which is superior to all other classifiers. Therefore, the proposed learning algorithm is found to be the most suitable algorithm among all other learning algorithms.

Table 5 lists the correctly classified samples and the corresponding accuracies obtained by O-DR2T+ PCA+LDA + IJaya-ELM on DS-255 during each trial of a $10 \times K$-fold SCV process. The results in the table indicate that the proposed scheme can correctly classify 2542 samples out of 2550 samples (2200 pathological and 350 healthy samples). Further, among 2200 pathological samples, 2195 are correctly classified by our scheme and the rest five samples are misclassified to healthy class. While among 350 healthy samples, 347 samples are correctly classified by our scheme and rest three samples are misclassified to pathological class. Considering these results, the sensitivity (S_e), specificity (S_p), and precision values (P_r) of the proposed scheme are computed as 99.77%, 99.14%, and 99.86%, respectively which are listed in Table 6.

Table 5 10 × 5-fold SCV result of O-DR2T + PCA+LDA + IJaya-ELM method on DS-255

Run	Fold-1	Fold-2	Fold-3	Fold-4	Fold-5	Total
1	51 (100.00)	51 (100.00)	51 (100.00)	50 (98.04)	50 (98.04)	253 (99.22)
2	51 (100.00)	51 (100.00)	51 (100.00)	51 (100.00)	51 (100.00)	255 (100.00)
3	51 (100.00)	51 (100.00)	51 (100.00)	51 (100.00)	51 (100.00)	255 (100.00)
4	51 (100.00)	51 (100.00)	51 (100.00)	51 (100.00)	51 (100.00)	255 (100.00)
5	50 (98.04)	51 (100.00)	51 (100.00)	51 (100.00)	51 (100.00)	254 (99.61)
6	51 (100.00)	51 (100.00)	51 (100.00)	50 (98.04)	51 (100.00)	254 (99.61)
7	51 (100.00)	51 (100.00)	50 (98.04)	51 (100.00)	51 (100.00)	254 (99.61)
8	50 (98.04)	51 (100.00)	51 (100.00)	51 (100.00)	51 (100.00)	254 (99.61)
9	51 (100.00)	51 (100.00)	51 (100.00)	51 (100.00)	50 (98.04)	254 (99.61)
10	51 (100.00)	51 (100.00)	50 (98.04)	51 (100.00)	51 (100.00)	254 (99.61)
Sum						2542
Average						254.2 (99.69)

To compare the efficacy of PCA+LDA over PCA, another experiment has been carried out on all the three datasets. The performance of both the schemes, namely, O-DR2T+ PCA + IJaya-ELM and O-DR2T+ PCA+LDA + IJaya-ELM are shown in Table 6. It may be noticed that the proposed O-DR2T+ PCA+LDA + IJaya-ELM scheme earns better performances than O-DR2T+ PCA + IJaya-ELM on all the datasets with relatively less number of features. Moreover, O-DR2T+ PCA + IJaya-ELM obtains slightly lesser sensitivity, specificity and precision values than O-DR2T+ PCA+LDA + IJaya-ELM. However, the higher the sensitivity value of a CAD system, the better is the performance of the CAD system. Therefore, the proposed O-DR2T+ PCA+LDA + IJaya-ELM scheme holds greater potential in making correct clinical decisions.

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

Dataset	Schemes	O-DR2T+PCA+IJaya-ELM	O-DR2T + PCA+LDA + IJaya-ELM
	No. of features	15	2
DS-66	S_e	100.00	100.00
	S_p	100.00	100.00
	P_r	100.00	100.00
	ACC	100.00	100.00
DS-160	S_e	99.64	100.00
	S_p	99.00	100.00
	P_r	99.86	100.00
	ACC	99.56	100.00
DS-255	S_e	99.50	99.77
	S_p	98.29	99.14
	P_r	99.73	99.86
	ACC	99.33	99.69

Table 7 Classification accuracy (%) comparison of the proposed method with wavelet based method

Schemes	No. of features	DS-66	DS-160	DS-255
DWT + PCA + IJaya-ELM	15	100.00	99.38	99.17
DWT + PCA+LDA + IJaya-ELM	2	100.00	99.94	99.41
O-DR2T + PCA+LDA + IJaya-ELM	**2**	**100.00**	**100.00**	**99.69**

Our proposed scheme is highlighted in bold

Further, in order to support the effectiveness of O-DR2T features over DWT features, we have conducted an experiment where DWT features are used in place of O-DR2T features and record the results with the same number of features as shown in Table 7. It may be seen that the proposed scheme achieves better performance than DWT based schemes on all the datasets. Here, the DWT features are extracted from all the sub-bands of 3-level decomposition. Additionally, the DWT features used in literature [10, 27, 53] are also tested which results in smaller accuracy than the proposed scheme. Finally, it is concluded that with O-DR2T features the proposed scheme brings potential improvements in the performance of the PBDS system.

6.4 Comparison with other PBDSs

An extensive comparison with twenty-two existing competent PBDSs has been made on three datasets in the context of feature size, run size, and the classification accuracy as given in Table 8. It can be seen that a large number of the PBDSs yield perfect classification on DS-66, but merely two schemes, such as RT + PCA + LS-SVM [8] and DWPT + TE + GEPSVM [52] offer ideal classification on DS-160. It can also be noticed that no existing PBDSs can achieve perfect classification, but the suggested system earns higher classification accuracy i.e., 99.69% than others with a minimum number of features. Though the improvement in accuracy is marginal and comparable with some of the existing schemes, the result is obtained over a number of runs of a K-fold SCV procedure. This reflects the improvement in proposed scheme to be robust and reliable. The use of IJaya-ELM in the proposed scheme leads to have better generalization performance and faster response on unknown testing data.

From the experimental results, it is clear that the suggested scheme yields superior performance in the context of classification accuracy and number of features used compared to other existing schemes over all the three datasets. The proposed system employs O-DR2T, and IJaya-ELM which possesses several advantages. O-DR2T helps in capturing edge and texture features effectively from MR images. IJaya-ELM obtains compact network structure, faster learning speed and better generalization performance in contrast to other traditional learning algorithms that are frequently employed in existing PBDSs. These methods collectively increase the strength of the system. However, the proposed system has the following loopholes. The proposed system has been validated on three available datasets which accommodate images from patients during the late and middle stages of diseases, but a larger dataset with images from all stages of diseases can be tested in order to achieve better generalization performance. The current work deals with solving a two-class classification problem, however solving a multi-class brain disease classification problem is highly in demand.

Table 8 Comparative analysis with other competent PBDSs on three standard datasets

Existing PBDSs	Feature size	Run	*ACC* (%)		
			DS-66	DS-160	DS-255
DWT + SVM + POLY [5]	4761	5	98.00	97.15	96.37
DWT + PCA + BPNN + SCG [53]	19	5	100.00	98.29	97.14
DWT + PCA + FNN + SCABC [61]	19	5	100.00	98.93	97.81
DWT + PCA + FNN + ACPSO [59]	19	5	100.00	98.75	97.38
DWT + PCA + KSVM [60]	19	5	100.00	99.38	98.82
DWPT + SE + GEPSVM [52]	16	10	99.85	99.62	98.78
DWPT + TE + GEPSVM [52]	16	10	100.00	100.00	99.33
WPTE + FNN + RCBBO [38]	16	10	100.00	100.00	99.49
WE + HMI + GEPSVM [58]	14	10	100.00	99.56	98.63
DWT + PCA + ADBRF [27]	13	10	100.00	99.18	98.35
FRFE + WTT + TSVM [43]	12	10	100.00	99.69	98.98
DTCWT + VE + GEPSVM [39]	12	10	100.00	99.75	99.25
FRFE + WTT + DP-MLP + ARCBBO [55]	12	10	100.00	99.19	98.24
RT + PCA + LS-SVM [8]	9	5	100.00	100.00	99.39
DWT + PCA + *k*-NN [11]	7	5	98.00	97.54	96.79
FPCNN + DWT + PCA + FNN [10]	7	10	100.00	98.88	98.43
SWT + PCA + IABAP-FNN [42]	7	10	100.00	99.44	99.18
SWT + PCA + ABC-SPSO-FNN [42]	7	10	100.00	99.75	99.02
SWT + PCA + GEPSVM [51]	7	10	100.00	99.62	99.02
WE + NBC [67]	7	10	92.58	91.87	90.51
DWT + PCA + LDA + RF [25]	7	10	100.00	99.75	99.14
MBD + SLFN + PSO-TTC [63]	5	10	100.00	98.19	98.08
O-DR2T + PCA + IJaya-ELM	**15**	**10**	**100.00**	**99.56**	**99.33**
O-DR2T + PCA+LDA + IJaya-ELM (Proposed)	**2**	**10**	**100.00**	**100.00**	**99.69**

Our proposed schemes are highlighted in bold

7 Conclusions and future work

In this paper, an attempt has been made to develop an efficient pathological brain detection system. The proposed scheme initially uses O-DR2T to extract the features from the enhanced brain MR images. Subsequently, a PCA+LDA approach has been employed to reduce the feature dimensionality. Finally, a novel learning algorithm called IJaya-ELM is proposed to train the SLFN. The proposed scheme inherits the advantages of O-DR2T and ELM for detection of pathological brain in MR images. The experimental results on three standard datasets demonstrate that the proposed scheme yields higher accuracy than other competent schemes with a minimum number of features. Moreover, it has been shown that the proposed IJaya-ELM leaning algorithm holds several advantages over other learning algorithms.

The proposed IJaya-ELM can be applied to regression problems as well as multi-label classification problems. Other advanced machine learning techniques such as dictionary

learning and deep learning could be investigated as potential alternatives to the proposed IJaya-ELM in future. However, our proposed system has following limitations. The proposed PBDS has been validated on small datasets, however, a larger dataset collected online will further prove its effectiveness. Further, the images in the chosen datasets are collected from the late and middle stage of the diseases, images collected during the all stages need to be validated. In future, interactive machine learning (iML) algorithms [15–17] can also be studied to overcome the issues of automatic machine learning algorithms.

References

1. Candes E, Demanet L, Donoho D, Ying L (2006) Fast discrete curvelet transforms. Multiscale Model Simul 5(3):861–899
2. Candès EJ, Donoho DL (1999) Ridgelets: a key to higher-dimensional intermittency? Philos Trans R Soc Lond A Math Phys Eng Sci 357(1760):2495–2509
3. Candès EJ, Donoho DL (2000) Curvelets- a surprisingly effective nonadaptive representation for objects with edges. Vanderbilt University Press, Nashville, pp 105–120
4. Candès EJ, Donoho DL (2004) New tight frames of curvelets and optimal representations of objects with piecewise C^2 singularities. Commun Pure Appl Math 57(2):219–266
5. Chaplot S, Patnaik LM, Jagannathan NR (2006) Classification of magnetic resonance brain images using wavelets as input to support vector machine and neural network. Biomed Signal Process Control 1(1):86–92
6. Cormack A (1981) The radon transform on a family of curves in the plane. Proc Am Math Soc 83(2):325–330
7. Cormack A (1982) The radon transform on a family of curves in the plane. ii. Proc Am Math Soc 86(2):293–298
8. Das S, Chowdhury M, Kundu K (2013) Brain MR image classification using multiscale geometric analysis of ripplet. Prog Electromagn Res 137:1–17
9. Do MN, Vetterli M (2003) The finite ridgelet transform for image representation. IEEE Trans Image Process 12(1):16–28
10. El-Dahshan EA, Mohsen HM, Revett K, Salem ABM (2014) Computer-aided diagnosis of human brain tumor through MRI: a survey and a new algorithm. Expert Syst Appl 41(11):5526–5545
11. El-Dahshan ESA, Honsy T, Salem ABM (2010) Hybrid intelligent techniques for MRI brain images classification. Digital Signal Process 20(2):433–441
12. Ghahremani M, Ghassemian H (2015) Remote sensing image fusion using ripplet transform and compressed sensing. IEEE Geosci Remote Sens Lett 12(3):502–506
13. Hamza R, Muhammad K, Lv Z, Titouna F (2017) Secure video summarization framework for personalized wireless capsule endoscopy. Pervasive and Mobile Computing
14. Han F, Yao HF, Ling QH (2013) An improved evolutionary extreme learning machine based on particle swarm optimization. Neurocomputing 116:87–93
15. Holzinger A (2016) Interactive machine learning for health informatics: when do we need the human-in-the-loop? Brain Infor 3(2):119–131
16. Holzinger A, Plass M, Holzinger K, Crišan GC, Pintea CM, Palade V (2016) Towards interactive machine learning (iml): applying ant colony algorithms to solve the traveling salesman problem with the human-in-the-loop approach. In: International conference on availability, reliability, and security. Springer, pp 81–95
17. Holzinger A, Plass M, Holzinger K, Crisan GC, Pintea CM, Palade V (2017) A glass-box interactive machine learning approach for solving np-hard problems with the human-in-the-loop. arXiv preprint. arXiv:1708.01104
18. Holzinger K, Palade V, Rabadan R, Holzinger A (2014) Darwin or lamarck? Future challenges in evolutionary algorithms for knowledge discovery and data mining. In: Interactive knowledge discovery and data mining in biomedical informatics. Springer, pp 35–56
19. Huang GB, Wang DH, Lan Y (2011) Extreme learning machines: a survey. Int J Mach Learn Cybern 2(2):107–122
20. Huang GB, Zhu QY, Siew CK (2004) Extreme learning machine: a new learning scheme of feedforward neural networks. In: IEEE international joint conference on neural networks, vol 2. IEEE, pp 985–990
21. Huang GB, Zhu QY, Siew CK (2006) Extreme learning machine: theory and applications. Neurocomputing 70(1):489–501

22. Johnson KA, Becker JA (1999) The whole brain atlas. http://www.med.harvard.edu/AANLIB/
23. Maitra M, Chatterjee A (2006) A Slantlet transform based intelligent system for magnetic resonance brain image classification. Biomed Signal Process Control 1(4):299–306
24. Martínez AM, Kak AC (2001) PCA versus LDA. IEEE Trans Pattern Anal Mach Intell 23(2):228–233
25. Nayak DR, Dash R, Majhi B (2015) Classification of brain MR images using discrete wavelet transform and random forests. In: Fifth national conference on computer vision, pattern recognition, image processing and graphics (NCVPRIPG). IEEE, pp 1–4
26. Nayak DR, Dash R, Majhi B (2015) Least squares svm approach for abnormal brain detection in mri using multiresolution analysis. In: 2015 international conference on computing, communication and security (ICCCS). IEEE, pp 1–6
27. Nayak DR, Dash R, Majhi B (2016) Brain MR image classification using two-dimensional discrete wavelet transform and AdaBoost with random forests. Neurocomputing 177:188–197
28. Nayak DR, Dash R, Majhi B (2016) Pathological brain detection using curvelet features and least squares SVM. Multimed Tool Appl 75:1–24
29. Nayak DR, Dash R, Majhi B (2017) Stationary wavelet transform and adaboost with SVM based pathological brain detection in MRI scanning. CNS Neurol Disord Drug Targets 16:137–149
30. Nayak DR, Dash R, Majhi B, Mohammed J (2016) Non-linear cellular automata based edge detector for optical character images. Simulation 92:1–11
31. Nayak DR, Dash R, Majhi B, Prasad V (2017) Automated pathological brain detection system: a fast discrete curvelet transform and probabilistic neural network based approach. Expert Systems with Applications
32. Pizer SM, Johnston RE, Ericksen JP, Yankaskas BC, Muller KE (1990) Contrast-limited adaptive histogram equalization: speed and effectiveness. In: Proceedings of the first conference on visualization in biomedical computing. IEEE, pp 337–345
33. Rao R (2016) Jaya: a simple and new optimization algorithm for solving constrained and unconstrained optimization problems. Int J Ind Eng Comput 7(1):19–34
34. Rao R, More K, Taler J, Ocłoń P (2016) Dimensional optimization of a micro-channel heat sink using jaya algorithm. Appl Therm Eng 103:572–582
35. Sajjad M, Khan S, Jan Z, Muhammad K, Moon H, Kwak JT, Rho S, Baik SW, Mehmood I (2017) Leukocytes classification and segmentation in microscopic blood smear: a resource-aware healthcare service in smart cities. IEEE Access 5:3475–3489
36. Saritha M, Joseph KP, Mathew AT (2013) Classification of MRI brain images using combined wavelet entropy based spider web plots and probabilistic neural network. Pattern Recogn Lett 34(16):2151–2156
37. Suresh S, Babu RV, Kim H (2009) No-reference image quality assessment using modified extreme learning machine classifier. Appl Soft Comput 9(2):541–552
38. Wang S, Li P, Chen P, Phillips P, Liu G, Du S, Zhang Y (2017) Pathological brain detection via wavelet packet tsallis entropy and real-coded biogeography-based optimization. Fundamenta Informaticae 151(1-4):275–291
39. Wang S, Lu S, Dong Z, Yang J, Yang M, Zhang Y (2016) Dual-tree complex wavelet transform and twin support vector machine for pathological brain detection. Appl Sci 6(6):169
40. Wang S, Phillips P, Yang J, Sun P, Zhang Y (2016) Magnetic resonance brain classification by a novel binary particle swarm optimization with mutation and time-varying acceleration coefficients. Biomed Eng/Biomedizinische Technik 61:431–441
41. Wang S, Rao RV, Chen P, Zhang Y, Liu A, Wei L (2017) Abnormal breast detection in mammogram images by feed-forward neural network trained by jaya algorithm. Fundamenta Informaticae 151(1-4):191–211
42. Wang S, Zhang Y, Dong Z, Du S, Ji G, Yan J, Yang J, Wang Q, Feng C, Phillips P (2015) Feed-forward neural network optimized by hybridization of PSO and ABC for abnormal brain detection. Int J Imaging Syst Technol 25(2):153–164
43. Wang S, Zhang Y, Yang X, Sun P, Dong Z, Liu A, Yuan TF (2015) Pathological brain detection by a novel image feature—fractional Fourier entropy. Entropy 17(12):8278–8296
44. Westbrook C (2014) Handbook of MRI technique. Wiley, Oxford
45. Xu J, Wu D (2012) Ripplet transform type II transform for feature extraction. IET Image Process 6(4):374–385
46. Xu J, Yang L, Wu D (2010) Ripplet: a new transform for image processing. J Vis Commun Image Represent 21(7):627–639
47. Xu Y, Shu Y (2006) Evolutionary extreme learning machine based on particle swarm optimization. In: International symposium on neural networks. Springer, pp 644–652
48. Yang G, Zhang Y, Yang J, Ji G, Dong Z, Wang S, Feng C, Wang Q (2015) Automated classification of brain images using wavelet-energy and biogeography-based optimization. Multimed Tool Appl 75:1–17

49. Yang J, Yang J (2003) Why can LDA be performed in PCA transformed space? Pattern Recogn 36(2):563–566
50. Zhang G, Wang Q, Feng C, Lee E, Ji G, Wang S, Zhang Y, Yan J (2015) Automated classification of brain MR images using wavelet-energy and support vector machines. In: 2015 international conference on mechatronics, electronic, industrial and control engineering (MEICx-15), pp 683–686
51. Zhang Y, Dong Z, Liu A, Wang S, Ji G, Zhang Z, Yang J (2015) Magnetic resonance brain image classification via stationary wavelet transform and generalized eigenvalue proximal support vector machine. J Med Imaging Health Inform 5(7):1395–1403
52. Zhang Y, Dong Z, Wang S, Ji G, Yang J (2015) 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
53. Zhang Y, Dong Z, Wu L, Wang S (2011) A hybrid method for MRI brain image classification. Expert Syst with Appl 38(8):10,049–10,053
54. Zhang Y, Ranjan Nayak D, Yang M, Yuan TF, Liu B, Lu H, Wang S (2017) Detection of unilateral hearing loss by stationary wavelet entropy. CNS Neurol Disord Drug Targets (Formerly Current Drug Targets-CNS & Neurological Disorders) 16(2):122–128
55. Zhang Y, Sun Y, Phillips P, Liu G, Zhou X, Wang S (2016) A multilayer perceptron based smart pathological brain detection system by fractional Fourier entropy. J Med Syst 40(7):1–11
56. Zhang Y, Wang S, Dong Z, Phillip P, Ji G, Yang J (2015) Pathological brain detection in magnetic resonance imaging scanning by wavelet entropy and hybridization of biogeography-based optimization and particle swarm optimization. Prog Electromagn Res 152:41–58
57. Zhang Y, Wang S, Ji G, Dong Z (2013) An MR brain images classifier system via particle swarm optimization and kernel support vector machine. Sci World J 2013:1–9
58. Zhang Y, Wang S, Sun P, Phillips P (2015) Pathological brain detection based on wavelet entropy and Hu moment invariants. Bio-Med Mater Eng 26(s1):S1283–S1290
59. Zhang Y, Wang S, Wu L (2010) A novel method for magnetic resonance brain image classification based on adaptive chaotic PSO. Prog Electromagn Res 109:325–343
60. Zhang Y, Wu L (2012) An MR brain images classifier via principal component analysis and kernel support vector machine. Prog Electromagn Res 130:369–388
61. Zhang Y, Wu L, Wang S (2011) Magnetic resonance brain image classification by an improved artificial bee colony algorithm. Prog Electromagn Res 116:65–79
62. Zhang Y, Chen S, Wang S, Yang JF, Phillips P (2015) 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
63. Zhang Y, Chen XQ, Zhan TM, Jiao ZQ, Sun Y, Chen ZM, Yao Y, Fang LT, Lv YD, Wang S (2016) Fractal dimension estimation for developing pathological brain detection system based on minkowski-Bouligand method. IEEE Access 4:5937–5947
64. Zhang Y, Wang S, Yang XJ, Dong ZC, Liu G, Phillips P, Yuan TF (2015) Pathological brain detection in MRI scanning by wavelet packet Tsallis entropy and fuzzy support vector machine. SpringerPlus 4(1):1–16
65. Zhang Y, Zhang Y, Lv YD, Hou XX, Liu FY, Jia WJ, Yang MM, Phillips P, Wang S (2017) Alcoholism detection by medical robots based on Hu moment invariants and predator–prey adaptive-inertia chaotic particle swarm optimization. Computers & Electrical Engineering
66. Zhao G, Shen Z, Miao C, Man Z (2009) On improving the conditioning of extreme learning machine: a linear case. In: 7th international conference on information, communications and signal processing, ICICS. IEEE, pp 1–5
67. Zhou X, Wang S, Xu W, Ji G, Phillips P, Sun P, Zhang Y (2015) Detection of pathological brain in MRI scanning based on wavelet-entropy and naive Bayes classifier. In: Bioinformatics and biomedical engineering, pp 201–209
68. Zhu QY, Qin AK, Suganthan PN, Huang GB (2005) Evolutionary extreme learning machine. Pattern Recogn 38(10):1759–1763

Deepak Ranjan Nayak is currently pursuing PhD in Computer Science and Engineering at National Institute of Technology, Rourkela, India. His current research interests include medical image analysis, pattern recognition and cellular automata. He is currently serving as the reviewer of many SCI indexed journals such as IET Image Processing, Multimedia Tools and Applications, Computer Vision and Image Understanding, Computer and Electrical Engineering, Fractals, Journal of Medical Imaging and Health Informatics, IEEE Access, etc. He is a student member of IEEE.

Ratnakar Dash received his PhD degree from National Institute of Technology, Rourkela, India, in 2013. He is currently working as Assistant Professor in the Department of Computer Science and Engineering at National Institute of Technology, Rourkela, India. His field of interests include signal processing, image processing, intrusion detection system, steganography, etc. He is a professional member of IEEE, IE, and CSI. He has published forty research papers in journals and conferences of international repute.

Banshidhar Majhi received his PhD degree from Sambalpur University, Odisha, India, in 2001. He is currently working as a Professor in the Department of Computer Science and Engineering at National Institute of Technology, Rourkela, India. His field of interests include image processing, data compression, cryptography and security, parallel computing, soft computing, and biometrics. He is a professional member of MIEEE, FIETE, LMCSI, IUPRAI, and FIE. He served as reviewer of many international journals and conferences. He is the author and co-author of over 70 journal papers of international repute. Besides, he has 100 conference papers and he holds 2 patents on his name. He has received "Samanta Chandra Sekhar Award" for the year 2016 by Odisha Bigyan Academy for his outstanding contributions to Engineering and Technology.