

## Accepted Manuscript

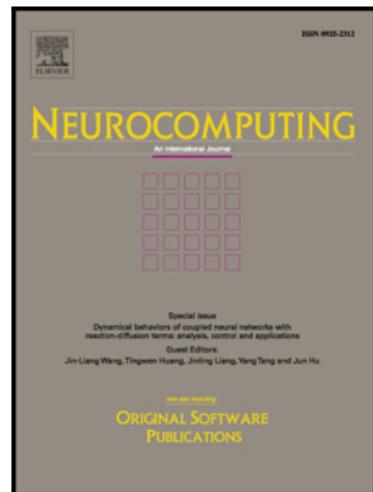
Discrete ripplet-II transform and modified PSO based improved evolutionary extreme learning machine for pathological brain detection

Deepak Ranjan Nayak, Ratnakar Dash, Banshidhar Majhi

PII: S0925-2312(17)31865-9

DOI: [10.1016/j.neucom.2017.12.030](https://doi.org/10.1016/j.neucom.2017.12.030)

Reference: NEUCOM 19164



To appear in: *Neurocomputing*

Received date: 25 April 2017

Revised date: 20 September 2017

Accepted date: 5 December 2017

Please cite this article as: Deepak Ranjan Nayak, Ratnakar Dash, Banshidhar Majhi, Discrete ripplet-II transform and modified PSO based improved evolutionary extreme learning machine for pathological brain detection, *Neurocomputing* (2017), doi: [10.1016/j.neucom.2017.12.030](https://doi.org/10.1016/j.neucom.2017.12.030)

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

# Discrete ripplet-II transform and modified PSO based improved evolutionary extreme learning machine for pathological brain detection

Deepak Ranjan Nayak<sup>a,\*</sup>, Ratnakar Dash<sup>a</sup>, Banshidhar Majhi<sup>a</sup>

<sup>a</sup>*Pattern Recognition Lab, Department of Computer Science and Engineering, National Institute of Technology, Rourkela, India 769 008*

## Abstract

Recently there has been remarkable advances in computer-aided diagnosis (CAD) system development for detection of the pathological brain through MR images. Feature extractors like wavelet and its variants, and classifiers like feed-forward neural network (FNN) and support vector machine (SVM) are very often used in these systems despite the fact that they suffer from many limitations. This paper presents an efficient and improved pathological brain detection system (PBDS) that overcomes the problems faced by other PBDSs in the recent literature. First, we support the use of contrast limited adaptive histogram equalization (CLAHE) to enhance the quality of the input MR images. Second, we use discrete ripplet-II transform (DR2T) with degree 2 as the feature extractor. Third, in order to reduce the huge number of coefficients obtained from DR2T, we employ PCA+LDA approach. Finally, an improved hybrid learning algorithm called MPSO-ELM has been proposed that combines modified particle swarm optimization (MPSO) and extreme learning machine (ELM) for segregation of MR images as pathological or healthy. In MPSO-ELM, MPSO is utilized to optimize the hidden node parameters (input weights and hidden biases) of single-hidden-layer feedforward neural networks (SLFN) and the output weights are determined analytically. The proposed method is contrasted with the current state-of-the-art methods on three benchmark datasets. Experimental results indicate that our proposed scheme brings potential improvements in terms of classification accuracy and number of features. Additionally, it is observed that the proposed MPSO-ELM algorithm achieves higher accuracy and obtains compact network architecture compared to conventional ELM and BPNN classifier.

**Keywords:** Computer-aided diagnosis (CAD), Magnetic resonance imaging (MRI), Discrete ripplet-II transform (DR2T), Extreme learning machine (ELM), Modified PSO (MPSO)

\*Corresponding author

Email addresses: [depakranjannayak@gmail.com](mailto:depakranjannayak@gmail.com) (Deepak Ranjan Nayak), [ratnakar@nitrk.ac.in](mailto:ratnakar@nitrk.ac.in) (Ratnakar Dash), [bmajhi@nitrk.ac.in](mailto:bmajhi@nitrk.ac.in) (Banshidhar Majhi)

## 1. Introduction

Brain disease is one of the leading cause of death in people with different age groups across the globe. Many types of brain diseases exist such as neoplastic diseases (brain tumor), cerebrovascular diseases (stroke), degenerative diseases, and infectious diseases. Some of these diseases cause minor problems in the human brain and some prompt to death. Therefore, development of the automated pathological brain detection system (PBDS) which is a type of computer-aided diagnosis (CAD) system specifically designed for the diagnosis of the human brain is of great importance. PBDS plays an important role in arriving at correct and quick clinical decisions. An advanced medical imaging modality known as magnetic resonance imaging (MRI) has been commonly used as the input to the PBDSs because of its advantage of providing huge information about the soft tissues of the human brain [1, 2, 3]. In addition, MRI is a non-invasive and radiation-free imaging modality as compared to other modalities like X-ray and CT scan. However, due to the enormous data storage, it is hard to interpret MR images manually. In particular, manual interpretation is a costly, troublesome and time-consuming task [4, 5]. To overcome such issues, automated PBDSs need to be developed by using dedicated computer systems which can assist radiologists in taking accurate and faster decisions. PBDS utilizes various image processing and pattern recognition algorithms in its different stages.

Many attempts have been made toward the development of various PBDSs in the past decade [6]. However, the development of ideal PBDS is still in its infancy because of the difficulty in selecting proper algorithms for feature extraction, feature selection and classification which will combinedly work in all cases regardless of the type of image modalities and the dataset size. Hence, PBDS remains an open challenging problem in front of researchers. Our objective in this study is to enhance the performance of the PBDS for abnormality detection in the human brain.

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 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 pre-

dicting continuous functions. To train FNN, conventional gradient-based learning algorithms such as back-propagation (BP) and Levenberg-Marquardt (LM) are used which have many limitations such as trapping at local minima, slower learning speed, and learning epochs. Traditional SVM classifier encounters higher computational complexity and it performs poorly on large datasets. Furthermore, the number of features required in some PBDSs is also higher which makes the classifier's task more difficult and costly.

This paper aims at developing an efficient PBDS which can overcome the issues faced by existing PBDSs. The proposed system markedly improves the recent results with the help of ripplet-II transform and a new variant of extreme learning machine (ELM). In more detail, the major contributions are

- (a) Use of discrete ripplet-II transform (DR2T) for feature extraction as it is effective in capturing 2D singularities along with a group of curves from MR images.
- (b) Use of a recently proposed learning algorithm known as ELM in order to overcome the problems of traditional learning algorithms. In particular, ELM provides faster learning speed and better generalization performance than other conventional learning algorithms.
- (c) Combining modified particle swarm optimization (MPSO) and ELM (MPSO-ELM) to avoid the difficulties faced by basic ELM such as local minima issue, slow response speed on testing data, high requirement of hidden neurons and ill-conditioned problem.
- (d) Comparison with other competent methods in terms of classification accuracy and number of features on three well-known datasets.

The remaining part of the article is organized as follows. Related works are summarized in Section 2. Section 3 offers the materials used in the experiments. The proposed methodology is discussed in Section 4. The statistical setting and pseudocode of the proposed scheme are presented in Section 5. In Section 6, the experimental results are analyzed. Conclusions based on this study and the future research directions are highlighted in Section 7.

## 2. Related Work

During the past decade, a number of pathological brain detection systems (PBDSs) have been reported in the literature for detecting various brain diseases. In almost all PBDSs, MRI has been used as the imaging modality. The PBDSs can be categorized into two classes, i.e., direct feature based PBDS and indirect feature based PBDS depending on the type of features used. In the first

category, the coefficients from several image transforms are directly utilized as the key features, while in the later one, statistical descriptors like energy, entropy, mean and standard deviation, evaluated from the direct coefficients are served as the features. The systems falling under the first category mostly require feature transformation or selection techniques in order to reduce the high dimensional feature space. However, these techniques are optional in case of the second category.

Chaplot et al. [1] were the forebears who proposed PBDS with the help of 2D DWT features and two classifiers such as self-organizing map (SOM) and SVM. The authors in [7] 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. [8] introduced a hybrid approach with the assistance of 2D DWT and two classifiers such as  $k$ -nearest neighbor ( $k$ -NN) and feed forward back-propagation artificial neural network (FP-ANN). Moreover, to reduce the feature dimensionality, they applied principal component analysis (PCA) on the feature vectors generated from 2D DWT. Further, with same features the authors in [9, 3, 10, 11], proposed new PBDSs. In these cases, gradient-based schemes and population based optimization schemes such as scaled conjugate gradient (SCG), 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). Zhang et al. [12] have suggested a PBDS where DWT plus PCA based features are given to a KSVM classifier. In [4], the authors proposed a PBDS where the features are derived from Ripple transform (RT) and then subjected to PCA for dimensionality reduction. Thereafter, they applied least squares SVM (LS-SVM) in order to get significant results. Later on, in [6], the authors have offered a scheme where DWT is used for feature extraction after the employment of feedback pulse coupled neural network (FPCNN). Finally, they have employed FP-ANN classifier. Afterward, Wang et al. [13] offered a new PBDS based on stationary wavelet transform (SWT) which is translation invariant in nature [14]. In this, PCA is used to derive a low dimensional feature vector. Additionally, in order to optimize the parameter of FNN classifier, they made use of two evolutionary schemes, namely, artificial bee colony (ABC) and PSO. These schemes were coined as IABAP-FNN, ABC-SPSO-FNN, and HPA-FNN. In another work, Zhang et al. [15] have deployed weighted-type fractional Fourier transform (WFRFT) and PCA for feature extraction and reduction, respectively. For classification, they have used two variants of SVM, namely, generalized eigenvalue proximal SVM (GEPSVM) and twin SVM (TSVM). Later, Nayak et al. [5] have proposed a PBDS with the support of 2D DWT and probabilistic PCA (PPCA). In this, AdaBoost with random forests (ADBRF) scheme was employed for classification. In [16], authors

have offered a PBDS which uses SWT, PCA, and GEPSVM for feature extraction, reduction, and classification. The authors in [17] have proposed a PBDS where the features are extracted from the 95 HL<sub>3</sub> sub-band of 2D DWT. PCA+LDA technique was applied instead of PCA in order to get more relevant features. Chen et al. [18] offered a new PBDS in which Minkowski-Bouligand dimension (MBD) features are computed after detection of edges from brain images using Canny edge detector. Thereafter, they proposed an improved PSO based on three-segment particle representation, time-varying acceleration coefficient, and chaos theory (PSO-TTC) to train the single-hidden layer 100 feedforward neural network. Dash et al. [19] proposed a PBDS based on curvelet features. In this, LS-SVM was employed for detection of pathological brain.

Many recent articles on PBDS have been used various feature descriptors like energy, entropy, mean and standard deviation etc., in the feature extraction stage. For example, Saritha et al. [20] have offered a PBDS which uses the wavelet entropy (WE) values as the features. The spider 105 web plot and *t*-test scheme are subsequently used to select the significant features. Finally, a probabilistic neural network (PNN) is employed for classification. Later, Yang et al. [21] have proposed a PBDS where the energy values of a level-3 DWT are used as features. For classification, they applied biogeography-based optimization (BBO) method with SVM in order to achieve better generalization performance. In [22], a discrete wavelet packet transform (DWPT) based PBDS is 110 proposed. Two different types of entropies namely, Shannon entropy (SE) [23] and Tsallis entropy (TE) were calculated from the sub-bands. GEPSVM is utilized to assign the class label as healthy or pathological. Furthermore, in [24], a PBDS in which FNN with a hybrid BBO and PSO based method known as HBP is proposed. In this work, wavelet entropy has been used as the features. In [25], WE and a Naive Bayes classifier (NBC) based PBDS is proposed, while in [26], a wavelet 115 energy and SVM based PBDS is introduced. Zhang et al. [27] have used Tsallis entropy values of DWPT as the features and fuzzy support vector machine (FSVM) as the classifier. In [28], seven WE features and seven Hu moment invariants (HMI) features are used followed by GEPSVM+RBF classifier. Wang et al. [29] have proposed a PBDS based on a novel feature called fractional Fourier 120 entropy (FRFE) which is the combination of FRFT and SE. Welch's *t*-test (WTT) and Mahalanobis distance (MD) was applied separately to select the relevant features and subsequently, twin SVM (TSVM) classifier is employed for classification. Later, in [30], a PBDS based on FRFE features and multilayer perceptron (MLP) is proposed. 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, adaptive real coded BBO (ARCBBO) approach is

125 employed to update the weights of MLP. In [31], the authors have employed three varieties of binary  
 PSO (BPSO) to select significant features from the 25 entropy values (primary features) of a 8-level  
 DWT. PNN was deployed for classification. While in [32], the variance and entropy (VE) values  
 relative to the sub-bands of a dual-tree complex wavelet transform (DTCWT) are used as features.  
 In this, GEPSVM and TSVM are employed as the classifier. Later, for feature extraction Nayak  
 130 et al.[33] have computed energy and entropy values from the sub-bands of 2D-SWT. They have  
 employed a symmetric uncertainty ranking (SUR) filter for feature selection. Finally, AdaBoost  
 with support vector machine (ADBSVM) is used for classification.

The literature study shows that in most PBDSSs wavelet and its variants (like SWT, DWPT,  
 DTCWT, etc.) are frequently used as the feature extractor. However, traditional DWT suffers from  
 135 many drawbacks such as limited directional selectivity and translation variance. SWT can resolve  
 translation variance issue; however, it leads to redundancy and it is not able to capture higher  
 dimensional singularities. Further, DTCWT is efficient and less redundant which offers more di-  
 rectional selectivities (i.e., six) as compared to SWT and DWT. Here, it can be concluded that all  
 140 these transforms are less capable of handling 2D singularities. Therefore, further improvements in  
 directional selectivity need to be explored to capture curve like features from MR images. Further-  
 more, it has been observed that FNN and SVM are commonly used in many PBDSSs despite they  
 require more parameters to tune and are time-consuming. Additionally, most of the schemes have  
 been validated on small datasets and shown higher accuracies; however, they perform poorly when  
 145 evaluated on large datasets. Thus, there exists a scope to mitigate the shortcomings of the existing  
 schemes in terms of number of features needed and improvement in accuracy on large datasets.

Keeping this in mind, we have proposed an efficient PBDS to classify the MR image as healthy  
 or pathological more efficiently. The proposed PBDS uses discrete ripplet-II transform (DR2T) for  
 feature extraction due to its ability in capturing directional features (edges and curves). Thereafter,  
 PCA+LDA approach is employed in order to decide the most significant feature set. Eventually,  
 150 an improved learning algorithm called MPSO-ELM for SLFN is introduced which offers advan-  
 tages such as local minima avoidance, better generalization capability, faster learning rate, and  
 well-conditioned compared to standard classifiers like FNN, SVM, LS-SVM, ELM, etc. These im-  
 provements lead the proposed PBDS to a more robust and accurate system over other existing  
 schemes.

155 **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  $256 \times 256$  in axial view plane which is available in Medical School of Harvard University website [34]. Along with the healthy brain samples, the datasets DS-66 and DS-160 have samples from seven classes of diseases, namely, sarcoma, AD, AD plus visual agnosia (VA), glioma, meningioma, Huntington's disease (HD), and Pick's disease (PD). However, 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 Figure 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.

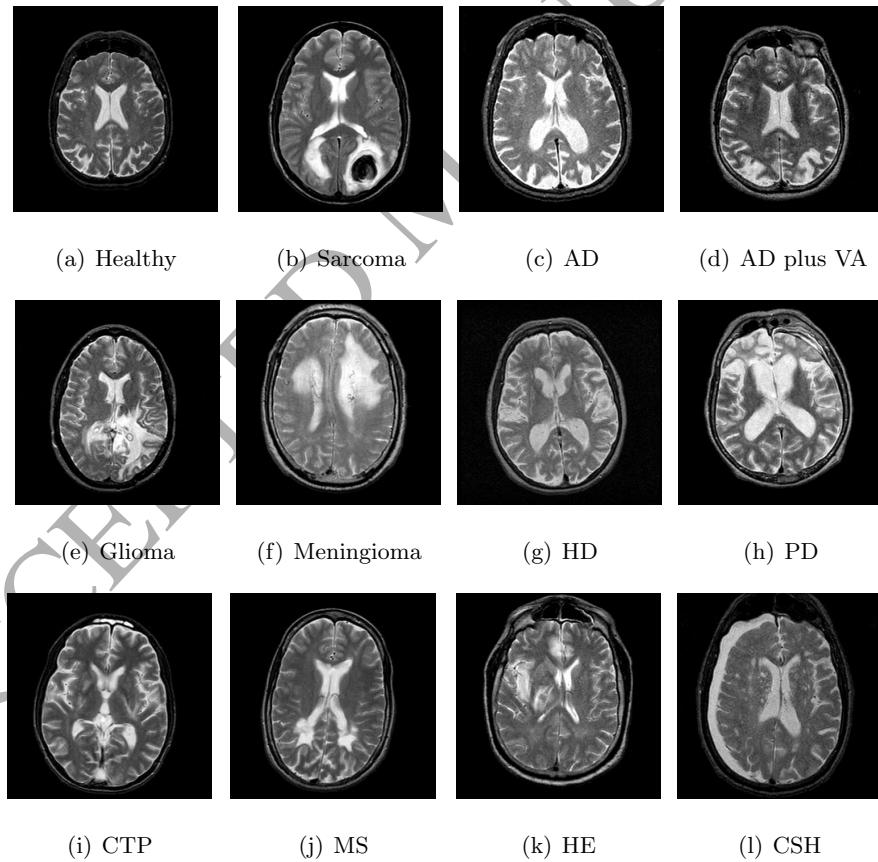


Figure 1: Samples of T2-weighted brain MR images [4]

The proposed work is a two-class classification problem (healthy or pathological), where the pathological class includes images from all kinds of diseases. It is worth mentioning here that predicting pathological samples as healthy (cost sensitivity problem) is more severe than the converse one.  
<sup>170</sup> This inaccurate prediction may cause patient's death. To alleviate this issue, a greater degree of pathological samples are deliberately taken into account as compared to healthy samples with the aim of making the system more biased towards the pathological samples.

#### 4. Proposed methodology

This section describes the methods involved in the proposed PBDS. The overall architecture of  
<sup>175</sup> the proposed PBDS is shown in Figure 2. As shown in the figure, the proposed PBDS has four major steps, namely, preprocessing, feature extraction, feature reduction, and classification. The input of the system is a brain MR image and the output is the class label (healthy or pathological). In the preprocessing step, we use contrast limited adaptive histogram equalization (CLAHE). In feature extraction step, we employ ripplet-II transform to derive features and in feature reduction  
<sup>180</sup> step, we harness PCA+LDA approach. Subsequently, for classification, an improved hybrid learning method MPSO-ELM is utilized, where MPSO 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 first part includes the training and evaluation process with the reduced feature sets, whereas,  
<sup>185</sup> the second part predicts a class label for a query MR image. In the following, all the steps are delineated in detail.

##### 4.1. Preprocessing based on CLAHE

Image preprocessing is one of the most vital and rudimentary steps in image analysis which leads to improvement in the quality of the images. It has been observed that some images in the selected datasets are low-contrast in nature. Therefore, to enhance the images, a well-known methodology  
<sup>190</sup> named contrast limited adaptive histogram equalization (CLAHE) has been employed. CLAHE is a variation of adaptive histogram equalization (AHE) which at first computes a histogram of gray values in a contextual region centered around each pixel and thereafter, it assigns a value to each pixel intensity within the display range according to the pixel intensity rank in its local histogram [35]. Dissimilar to AHE, CLAHE has the benefits of preventing over-enhancement of  
<sup>195</sup> noise and diminishing the edge-shadowing effect [36]. It uses a fixed value dubbed as clip limit which helps in clipping the histogram prior to the computation of cumulative distribution function

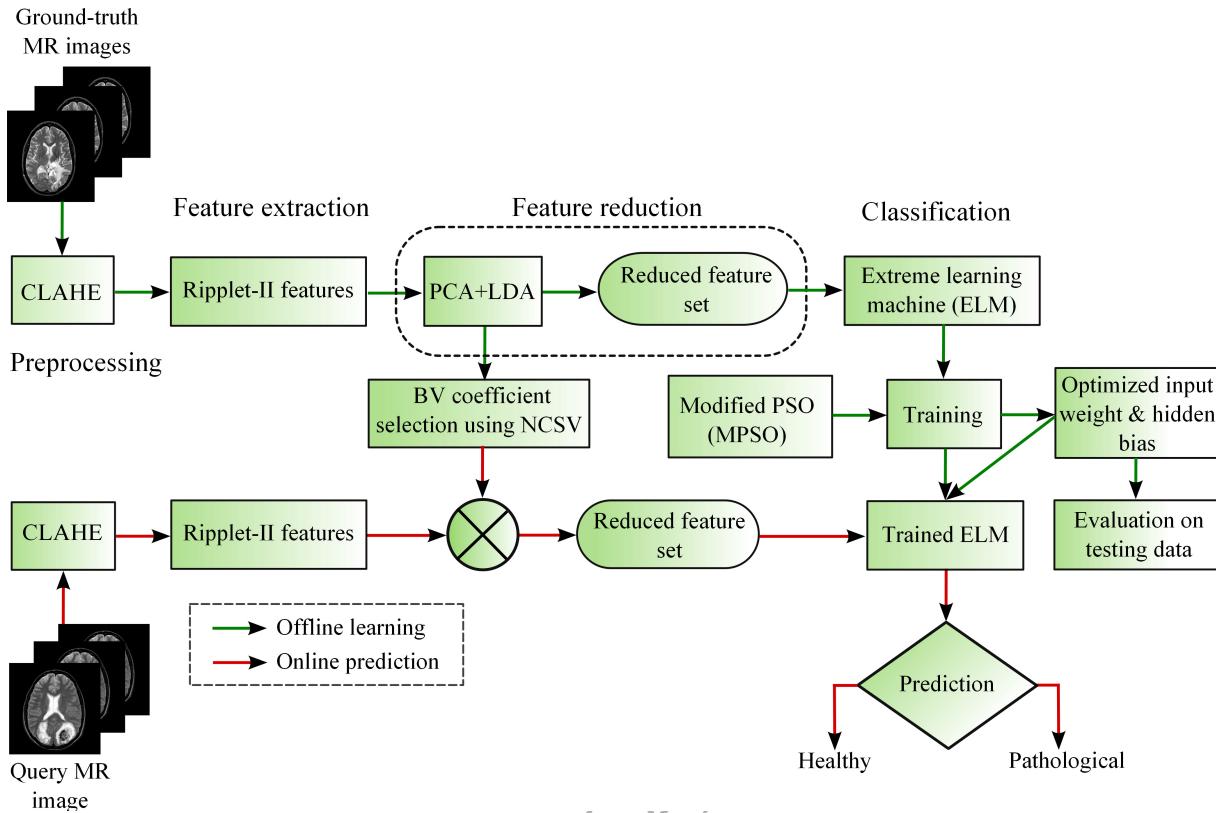


Figure 2: Architecture of the proposed PBD system

(CDF). However, CLAHE redistributes those parts of the histogram that surpass the clip limit equally among all histogram bins.

#### 4.2. Feature extraction based on DR2T

Fourier transform has been found to be less suitable for feature extraction from images since it loses the time information and it can not handle 1D singularities. Hence, it fails to provide efficient representation of images that contains edges, however, it only works well for 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 [37, 38]. Ridgelet holds great potential in representing line singularities (i.e., it is capable of extracting lines of arbitrary orientation), but it can not handle 2D singularities effectively. Thereafter, first generation curvelet transform based on multiscale ridgelet was proposed by Candes and

210 Dohono to resolve the 2D singularities along smooth curves [39]. Later on, they proposed the second  
 generation curvelet transform called fast discrete curvelet transform (FDCT) which is simple, fast,  
 and less redundant than the former one [40]. Because of the capabilities like multiresolution, more  
 directional selectivity, anisotropy, localization, etc., it has drawn attentions over last decades. The  
 anisotropic property guarantees solving 2D singularities along  $C^2$  curves and to accomplish this,  
 215 curvelet utilizes a parabolic scaling law [41]. However, the reason behind the selection of parabolic  
 scaling is not clear. In order to combat this issue, a new transform called as ripplet-I transform  
 is proposed which generalizes the scaling law [42, 43]. 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-  
 220 I transform with anisotropy capability of representing 2D singularities along arbitrarily shaped  
 curves. Then, they have proposed ripplet-II transform [44] which is based on generalized Radon  
 transform (GRT) [45, 46] to further improve the capability of representing 2D singularities. It sat-  
 isfies the properties like multiresolution, localization, good directionality, and flexibility. Compared  
 to wavelet and ridgelet transform, ripplet-II has the fastest decay in coefficients and for which the  
 225 sparser representation of the images having edges is possible. Further, it enjoys rotation invariant  
 property. In summary, it is capable of giving rotation invariant features and sparse feature vectors  
 which is crucial for classification task. Therefore, it has been leveraged in applications like texture  
 classification and image retrieval [44]. As the ripplet-II transform is efficient in representing edges  
 230 and textures than other conventional transforms and the affected regions in MR images contain  
 edges and textures of arbitrary shapes, hence it is used as the feature extraction tool in this study.

#### 4.2.1. Ripplet-II transform

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

$$RT^2_g(s, t, d, \theta) = \int \int \bar{\psi}_{s,t,d,\theta}(\rho, \alpha) g(\rho, \alpha) \rho \, d\rho \, d\alpha \quad (1)$$

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

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

where  $\varphi : \mathbb{R} \rightarrow \mathbb{R}$  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, ripplet-II transform can capture structural information along arbitrary curves.

Using (1) and (2), we have

$$RT^2_g(s, t, d, \theta) = \langle \varphi_{s,t}(r), GR_d[g] \rangle \quad (3)$$

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

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

The GRT can also be evaluated using Fourier transform [44]. Eq.(3) indicates that ripplet-II transform is the inner product between GRT and 1D wavelet. It can also be represented as

$$g(\rho, \alpha) \xrightarrow{GRT} GR_d[g](r, \theta) \xrightarrow{1D-WT} RT^2_g(s, t, d, \theta) \quad (5)$$

which defines that ripplet-II transform works 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

$$g(\rho, \alpha) \xrightarrow{DGRT} GR_d[g](r, \theta) \xrightarrow{1D-DWT} RT^2_g(s, t, d, \theta) \quad (6)$$

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 [44]

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

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

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

where  $T_n(.)$  denotes the Chebyshev polynomial of degree  $n$ .

<sup>235</sup> 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  $(\rho, \alpha)$  by  $(\rho'^2, 2\alpha')$  in  $g(\rho, \alpha)$ . Subsequently, generate a new image  $g'(x, y)$  by interpolation after converting polar coordinates  $(\rho', \alpha')$  to Cartesian coordinates  $(x, y)$ . The variables  $x$  and  $y$  hold integer values.

- 240 (ii) Employ discrete CRT on  $g'(x, y)$  that produces  $R(r', \theta')$  and then substitute  $(r', \theta')$  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  $(r', \theta')$  to  $(\sqrt{r}, \theta/2)$  makes DR2T coefficients more sparser than others.

#### 4.2.2. Feature generation

245 In the proposed system, DR2T is used as the feature extractor. For each training input MR image, we apply 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 derived 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 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  $(\rho, \alpha)$  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  $g'(x, y)$  i.e.,  $G'(u, v)$  along  $\theta$  (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  $\theta$  (columns)
  - 7:     Apply 1D DWT on  $g_{inv}$  along  $r$  and obtain the coefficients
  - 8:     Arrange the coefficients in a vector of size  $1 \times D$ , where  $D$  is the total number of features and store it in a matrix
  - 9: **end for**
  - 10: Obtain a feature matrix  $F_M$  containing all vectors
- 

250 4.3. Feature reduction based on PCA+LDA

It has been noticed that the features derived using DR2T are of high dimension which prompts to high computational overhead and high storage space requirement. Therefore, the application of different dimensionality reduction techniques is of great importance to obtain the most relevant candidate features. Principal component analysis (PCA) has been found to be effective in reducing

255 feature dimension. PCA transforms high-dimensional input data to a lower-dimensional space termed as principal subspace while keeping maximum variations of the data. That is, PCA always seeks a direction that best represents the data by excluding the class labels and hence unsupervised in nature.

In contrast to PCA, linear discriminant analysis (LDA), a supervised approach, attempts to  
 260 find a feature subspace that best discriminates between the classes and therefore has drawn the attention of researchers in the past decades. More formally, LDA always searches those vectors over which the samples of dissimilar classes are far from each other, whereas the samples of similar classes are close to each other. However, traditional LDA leads to degradation in performance while dealing with high dimensional and small sample size problem as in these cases the within-scatter  
 265 matrix ( $S_w$ ) becomes singular [47]. Further, to make sure that  $S_w$  does not become singular, at least  $D + C$  (where  $D$ =dimension of the feature vector and  $C$ =number of classes) number of samples are needed which in general is practically not possible [48]. To tackle this problem, a popular approach called PCA+LDA has been applied in our proposed work, where a  $D$ -dimensional data is first reduced using PCA to an  $M$ -dimensional data and thereafter reduced to a  $l$ -dimensional data  
 270 using LDA,  $l << M < D$ .

The process of finding an optimal or relevant feature set is described as follows. Initially, the eigenvalues of different features are arranged in decreasing order. Subsequently, a measure called the normalized cumulative sum of variances (NCSV) corresponding to each feature is calculated and the NCSV value for  $j^{th}$  feature is defined as

$$NCSV(j) = \frac{\sum_{u=1}^j \alpha(u)}{\sum_{u=1}^D \alpha(u)} ; 1 \leq j \leq D \quad (9)$$

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 features for which the NCSV value surpasses the threshold are selected. Relevant features selected are determined experimentally to have a maximal accuracy. It is worth mentioning here that the coefficients of the first  $l$  eigenvectors  
 275 (satisfying the threshold) are retained to perform feature reduction during online prediction. Here, the selected eigenvectors are suitably called as basis vectors (BV).

#### 4.4. Classification based on MPSO-ELM

In this section, we first discuss the preliminaries of extreme learning machine (ELM) and particle swarm optimization (PSO), and thereafter we describe the proposed MPSO-ELM algorithm in 280 detail.

##### 4.4.1. Extreme learning machine (ELM)

In the past decades, single-hidden layer feedforward neural networks (SLFNs) have been shown to be used in many applications as it could approximate any continuous function and classify any disjoint region. To train the SLFNs, gradient-based learning algorithms such as Levenberg-285 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, get trapped at local minima, require large number of iterations to obtain better learning performance, and overfitting [49]. A recently developed learning algorithm known as extreme learning machine (ELM) avoids the limitations faced by gradient based learning 290 schemes. ELM has also the potential for high performance classification and solving regression tasks [50, 51]. Different from 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 295 has been discussed below mathematically.

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

$$\sum_{i=1}^{n_h} w_i^o \phi(x_j) = \sum_{i=1}^{n_h} w_i^o \phi(w_i^h \cdot x_j + b_i) = o_j, \quad j = 1, 2, \dots, N \quad (10)$$

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. The SLFNs can approximate these  $N$  samples with zero error, i.e.,  $\exists w_i^h, w_i^o$ , and  $b_i$  such that

$$\sum_{i=1}^{n_h} w_i^o \phi(w_i^h \cdot x_j + b_i) = t_j, \quad j = 1, 2, \dots, N \quad (11)$$

Now, Eq. (11) can be represented in matrix form as

$$\mathbf{H}w^o = \mathbf{T} \quad (12)$$

where,

$$\mathbf{H}(w_1^h, w_2^h, \dots, w_{n_h}^h, b_1, b_2, \dots, b_{n_h}, x_1, x_2, \dots, x_N) \\ = \begin{bmatrix} \phi(w_1^h \cdot x_1 + b_1) & \dots & \phi(w_{n_h}^h \cdot x_1 + b_{n_h}) \\ \vdots & \dots & \vdots \\ \phi(w_1^h \cdot x_N + b_1) & \dots & \phi(w_{n_h}^h \cdot x_N + b_{n_h}) \end{bmatrix}_{N \times n_h}, w^o = \begin{bmatrix} w_1^{oT} \\ \vdots \\ w_{n_h}^{oT} \end{bmatrix}_{n_h \times C} \text{ and } \mathbf{T} = \begin{bmatrix} t_1^T \\ \vdots \\ t_N^T \end{bmatrix}_{N \times C}$$

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 (Eq. (12)) as

$$\hat{w}^o = \mathbf{H}^\dagger \mathbf{T} \quad (13)$$

300 where,  $\mathbf{H}^\dagger$  indicates the Moore-Penrose (MP) generalized inverse of matrix  $\mathbf{H}$  and this method helps ELM to have better generalization performance [52]. 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.

#### 305 4.4.2. Particle swarm optimization (PSO)

Particle swarm optimization (PSO), a population-based evolutionary algorithm introduced by Eberhart and Kennedy [53, 54], has shown great potential in a wide variety of search and optimization problems. This algorithm is influenced by the social behavior of bird flocking or fish schooling. Unlike other evolutionary algorithms such as genetic algorithms (GA) and differential evolution 310 (DE), PSO does not require any evolution parameters (crossover and mutation) and is easy to implement. In PSO, each bird of a flock represents a solution in the search space which is also called as a particle. Each particle is characterized by its own position and velocity. At first, PSO initializes a set of particles (or solutions) randomly known as swarm and thereafter looks for the best solution (optima) by updating generations. In each iteration, the velocity of each particle is 315 updated by two best values, namely,  $pbest$  and  $gbest$ , where  $pbest$  denotes the position of the best solution found so far by a particle and  $gbest$  indicates the best value found so far by all the particles in the population. Then, the position of the particle is adjusted using the updated velocity.

For a  $D$ -dimensional search space, the position and velocity of  $j^{th}$  particle can be expressed as  $S_j = (s_{j1}, s_{j2}, \dots, s_{jD}) \in \Re^D$  and  $V_j = (v_{j1}, v_{j2}, \dots, v_{jD}) \in \Re^D$ , respectively. The  $pbest$  and  $gbest$  of the  $j^{th}$  particle is denoted as  $P_{jpbest} = (p_{j1}, p_{j2}, \dots, p_{jD})$  and  $P_{gbest} = (p_{gbest1}, p_{gbest2}, \dots, p_{gbestD})$

respectively. Then the traditional PSO is stated as [53, 54]

$$v_{jd}(k+1) = v_{jd}(k) + c_1 * rand1() * (p_{jd}(k) - s_{jd}(k)) + c_2 * rand2() * (p_{gbestd}(k) - s_{jd}(k)) \quad (14)$$

$$s_{jd}(k+1) = s_{jd}(k) + v_{jd}(k+1), \quad 1 \leq j \leq N_p, \quad 1 \leq d \leq D \quad (15)$$

where,  $N_p$  indicates the total number of particles,  $c_1$  and  $c_2$  are the acceleration coefficients, and  $rand1()$  and  $rand2()$  denote two separate uniform distributed random numbers in the range [0,1].

Later, Shi and Eberhart [55] proposed an adaptive PSO (APSO) by introducing a new parameter called inertia weight into traditional PSO and is defined as

$$v_{jd}(k+1) = \omega * v_{jd}(k) + c_1 * rand1() * (p_{jd}(k) - s_{jd}(k)) + c_2 * rand2() * (p_{gbestd}(k) - s_{jd}(k)) \quad (16)$$

where,  $\omega$  denotes the inertia weight which helps in balancing the local and global search.  $\omega$  can be defined as a decreasing function of time instead of a fixed positive value,

$$\omega(k) = \omega_f - [k * (\omega_f - \omega_i)]/maxIter \quad (17)$$

320 where,  $\omega_i$ ,  $\omega_f$ ,  $k$ , and  $maxIter$  represent initial weight, final weight, current iteration number and maximum iteration number, respectively. Contrasted with classical PSO, this variant of PSO is more efficient since it plays a role in finding global optimum with a reasonable number of iterations.

#### 4.4.3. Proposed MPSO-ELM method

Since ELM randomly chooses the input weights and hidden biases, it leads to two crucial problems [52, 56, 57]: (i) it needs more hidden neurons than conventional gradient based methods which make ELM respond slowly to unknown testing data, and (ii) it causes an ill-conditioned hidden layer output matrix  $\mathbf{H}$  in presence of more hidden neurons for which ELM induces poor generalization performance. Condition number has been found to be a good qualitative measure to calculate the conditioning of a matrix [57]. It indicates how close a system is to be ill-conditioned. 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,

$$\mathcal{K}_2(\mathbf{H}) = \sqrt{\frac{\lambda_{max}(\mathbf{H}^T \mathbf{H})}{\lambda_{min}(\mathbf{H}^T \mathbf{H})}} \quad (18)$$

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

325 In order to tackle these issues, few efforts have been made 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. Zhu et al. [52] 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  
330 have shown that E-ELM provides faster learning speed and better generalization performance than  
other traditional algorithms. Additionally, it obtains much more compact network than basic ELM.  
However, E-ELM demands two additional parameters to tune, namely, the mutation factor and the  
crossover factor. Because of the potency of PSO to search global optimum, it can be expected  
that hybridization of PSO and ELM should be promising for training SLFNs. Xu and Shu [56]  
335 introduced another E-ELM based on PSO to select the hidden node parameters which require only  
one parameter to tune manually. They have added boundary conditions into conventional PSO to  
enhance the performance of ELM. Later, in [58], an improved PSO based ELM (IPSO-ELM) is  
proposed to find optimal SLFNs. During searching, IPSO considers both the root mean squared  
340 error (RMSE) and the norm of output weights of the validation set in order to obtain better  
convergence performance. Suresh et al. [59] 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. Zhao et al. [57] 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  
345 accuracy.

In this paper, a new approach combining modified PSO (MPSO) and ELM is suggested which  
avoids the issues faced by existing methods in the recent literature. In this method, we use MPSO to  
optimize the hidden node parameters and MP generalized inverse to analytically find the solution.  
Since the two acceleration components ( $c_1$  : cognitive component and  $c_2$  : social component) in  
PSO strongly influences the convergence to the global optimum solution, proper selection of these  
components are very important. In order to select these components efficiently, we incorporate a  
strategy called time-varying acceleration coefficients (TVAC) [60] in addition to the time-varying  
inertia weight factor in classical PSO and hence we termed it as MPSO. TVAC improves the global  
search ability at the beginning stage of the optimization and encourages the particles to converge  
toward the global optima at the end of the search. Here, we select a large value of  $c_1$  and small  
value of  $c_2$  at the beginning of optimization, whereas we select a small value of  $c_1$  and large value

of  $c_2$  at the final stage. Mathematically,  $c_1$  and  $c_2$  can be defined as

$$c_1(k) = c_{1i} + [k * (c_{1f} - c_{1i})] / \text{maxIter} \quad (19)$$

$$c_2(k) = c_{2i} + [k * (c_{2f} - c_{2i})] / \text{maxIter} \quad (20)$$

where,  $c_{1i}$  and  $c_{1f}$  indicate the initial and final value of  $c_1$ , respectively. Similarly,  $c_{2i}$  and  $c_{2f}$  represent the initial and final value of  $c_2$ , respectively.  $k$  and  $\text{maxIter}$  are the current iteration number and maximum number of allowable iterations.

The main goal of MPSO is to minimize the norm of the output weights and to bound the hidden node parameters within a specific range which on the other hand enhances the convergence performance of ELM. This idea was conceived by Bartlett [61] where it is stated that neural networks tend to have better generalization performance with the weights of smaller norm. Therefore, we consider both RMSE and norm of the output weights of SLFNs to search the global optima in MPSO. The steps of the proposed MPSO-ELM are delineated as follows:

- (a) At first, initialize randomly all the particles in the swarm such that each particle consists of a set of input weights and hidden biases as

$$P_j = \left[ w_{11}^h, w_{12}^h, \dots, w_{1l}^h, w_{21}^h, w_{22}^h, \dots, w_{2l}^h, w_{n_h 1}^h, w_{n_h 2}^h, \dots, w_{n_h l}^h, b_1, b_2, \dots, b_{n_h} \right] \quad (21)$$

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

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

$$f() = \sqrt{\frac{\sum_{j=1}^{N_v} \left\| \sum_{i=1}^{n_h} w_i^o \phi(w_i^h \cdot x_j + b_i) - t_j \right\|_2^2}{N_v}} \quad (22)$$

where,  $N_v$  indicates the number of validation samples.

- (c) Update the  $P_{jbest}$  of all the particles and  $P_{gbest}$  of the swarm using the fitness value and the norm of the output weights as follows:

$$P_{jbest} = \begin{cases} P_j & \text{if } f(P_j) < f(P_{jbest}) \text{ and } \|w_{P_j}^o\| < \|w_{P_{jbest}}^o\| \\ P_{jbest} & \text{otherwise} \end{cases} \quad (23)$$

$$P_{gbest} = \begin{cases} P_j & \text{if } f(P_j) < f(P_{gbest}) \text{ and } \|w_{P_j}^o\| < \|w_{P_{gbest}}^o\| \\ P_{gbest} & \text{otherwise} \end{cases} \quad (24)$$

where,  $f(P_j)$ ,  $f(P_{jbest})$ , and  $f(P_{gbest})$  denotes the fitness value of the current particle  $j$ , the best position of the particle  $j$  so far, and the global best position in the swarm, respectively.

360  $w_{P_j}^o$ ,  $w_{P_{jbest}}^o$ , and  $w_{P_{gbest}}^o$  represents the output weights generated by MP generalized inverse for the the current particle  $j$ , the best position of the particle  $j$  so far, and the global best position in the swarm, respectively.

- (d) Calculate the value of time-varying parameters,  $\omega$ ,  $c_1$  and  $c_2$  using Eqs. (17), (19) and (20) respectively. Subsequently, update the velocity and position of all the particles in the swarm based on Eqs. (15) and (16) and generate the new population.
- 365 (e) Based on the literature [49, 50], all the input weights and biases should lie in the range of [-1,1]. Thus, check the following condition to deal with the position out-of-bound issue

$$s_{jd}(k+1) = \begin{cases} -1 & \text{if } s_{jd}(k+1) < -1 \\ 1 & \text{if } s_{jd}(k+1) > 1 \end{cases}, \quad 1 \leq j \leq N_p, \quad 1 \leq d \leq D \quad (25)$$

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

As the proposed scheme uses Eq. (23) and (24) to find the optimal input weights and hidden 370 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 MPSO-ELM has the following advantages: it has no evolution parameters, it improves the conditioning, and it provides better generalization performance with a much more compact network. Compared to other gradient based methods and classical ELM, 375 this approach does not need activation function to be continuously differentiable. The pseudocode of the MPSO-ELM algorithm is given in Algorithm 2. Since the proposed scheme involves major techniques such as DR2T, PCA+LDA and MPSO-ELM, the scheme is referred to as “DR2T + PCA+LDA + MPSO-ELM”. The overall steps followed in the proposed system are articulated in Algorithm 3.

**Algorithm 2** Pseudocode of the proposed MPSO-ELM scheme

---

```

1: Initializing MPSO Initialize all particles in the swarm where each particle correspond to
   input weights and biases. Initialize  $\omega_i$ ,  $\omega_f$ ,  $c_{1i}$ ,  $c_{1f}$ ,  $c_{2i}$  and  $c_{2f}$ 
2: Evaluate the output weights for each particle according to Eq. (13)
3: Calculate the validation error for each particle according to Eq. (22) as fitness
4: Initialize the  $P_{pbest}$  and  $P_{gbest}$  using the fitness values
5: while  $k <$  Max number of iterations do
6:   Calculate the time-varying parameters  $\omega$ ,  $c_1$  and  $c_2$  according to Eqs. (17), (19) and (20)
7:   for each particle do
8:     Update the velocity and position according to Eq. 15 and (16)
9:     Check if any particles go beyond the search space, amend it according to Eq. (25)
10:    Update the  $P_{jpbest}$  and  $P_{gbest}$  according to Eq. (23) and Eq. (24) respectively
11:   end for
12:    $k = k + 1$ 
13: end while
14: Output the global best position

```

---

380 **5. Experimental design and evaluation**

In this section, we discuss about the experimental design and performance measures used. In order to validate the proposed scheme “DR2T + PCA+LDA + MPSO-ELM”, 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 which avoids over-fitting problem. Also, it makes 385 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 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 are used for training. The validation set is used to find the parameters of the MPSO-ELM i.e., it helps us to know when to stop training. The test set is used 390 to evaluate the performance in a run of five trials. For DS-66, we employ 6-fold stratified cross validation (SCV) while for other two datasets, we select 5-fold SCV. Additionally, we run the SCV procedure 10 times on three datasets to avoid randomness. It is worth mentioning here that the statistical setting for all the three datasets is kept similar to the literatures [4, 5, 30] as shown in Table 1.

**Algorithm 3** Implementation steps of the proposed PBDS**Offline learning:**

- 1: **for** each ground truth image **do**
- 2:     Enhance the contrast using CLAHE
- 3:     Apply DR2T with degree 2 on the enhanced image.
- 4:     Obtain the 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 MPSO and find the optimized input weights and hidden biases. Calculate the RMSE on the validation set as fitness
- 9: Compute 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 as input to the system.
- 2: Preprocess the query image with CLAHE
- 3: Employ DR2T with degree 2 over the enhanced image
- 4: Obtain the 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 classifier trained by MPSO-ELM and predict the output label as healthy or pathological

395 Four different measures, namely, sensitivity ( $S_e$ ), specificity ( $S_p$ ), precision ( $P_r$ ) and accuracy ( $ACC$ ) are used to test the effectiveness of the proposed scheme.  $S_e$  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,  $ACC$  determines the fraction of the correctly predicted samples (both pathological and healthy) in the total number of testing samples. Moreover, to  
400 compare the proposed MPSO-ELM method with other methods such as DE-ELM, PSO-ELM, ELM, and BPNN, two additional measures (condition number and norm of output weights) are used.

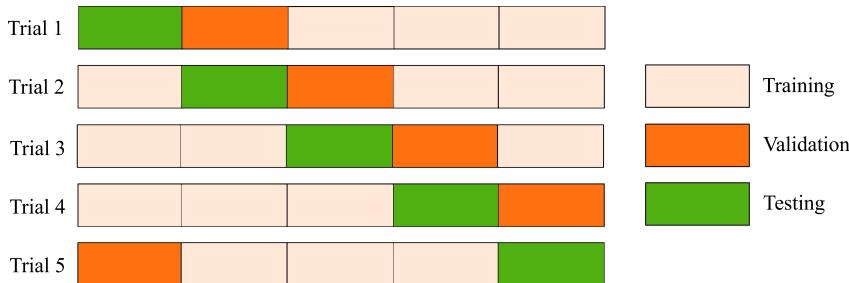


Figure 3: Illustration of 5-fold cross validation setting for a single run

Table 1: Statistical setting of  $K$ -fold SCV for three benchmark datasets [4, 5, 30]

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

## 6. Experimental results and analysis

The proposed system was implemented using MATLAB toolbox on a machine with 3.7 GHz processor, 8 GB RAM, and windows 8 OS. The parameters used and the statistical set up was kept similar to other competent schemes to derive relative comparisons.

### 6.1. Preprocessing and feature extraction results

To enhance the contrast of original MR image, 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 ( $\beta$ ) are selected as 256 and 0.01. 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 Figure 4. From the figure, it is evident that the affected regions in the enhanced images are more clear as compared to the original images.

Next, we apply DR2T to each of the preprocessed images and extract the features as the transform coefficients of DR2T. In DR2T, we set the number of decomposition levels for 1D DWT as 2 and use Haar wavelet as the basis because of its simplicity. As the images are of size  $256 \times 256$ ,

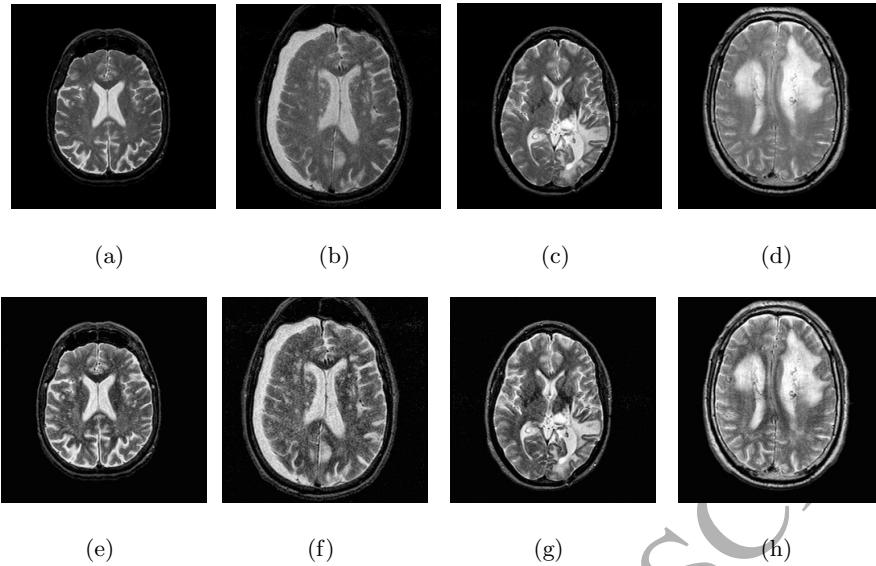


Figure 4: Preprocessing using CLAHE ( $\beta=0.01$  and region size= $8\times 8$ ). Row 1 lists the original MR samples. Row 2 lists the corresponding contrast enhancement using CLAHE

therefore the total number of features extracted by DR2T from a single image are  $256*256 = 65536$  which is much larger in size.

420 Further, 2D DWT and ridgelet transform are applied in place of DR2T and their coefficients are stored. The magnitude of the coefficients of each transform is computed and then the coefficients of the corresponding transforms are normalized by their largest coefficient. Finally, the normalized magnitude of coefficients is sorted in decreasing order to check the rate of decay in coefficients (Figure 5). It may be observed that DR2T provides the fastest decay in coefficients as compared to  
 425 2D DWT and ridgelet. Therefore, DR2T generates sparse feature vectors which strongly influences the classification performance.

### 6.2. Feature reduction results

We employ PCA+LDA to reduce the high dimensional feature space (65536 features) derived from DR2T. The significant feature set is obtained according to the NCSV values of the features.  
 430 The NCSV values relating to various number of features obtained by PCA+LDA and standard PCA are shown in Figure 6. It can be observed that PCA+LDA preserves maximum information with only three features, whereas standard PCA requires more number of features. For a chosen threshold value of 0.95, standard PCA and PCA+LDA result in 13 and three features respectively. Therefore, PCA+LDA approach is found to be more suitable for finding the significant features.

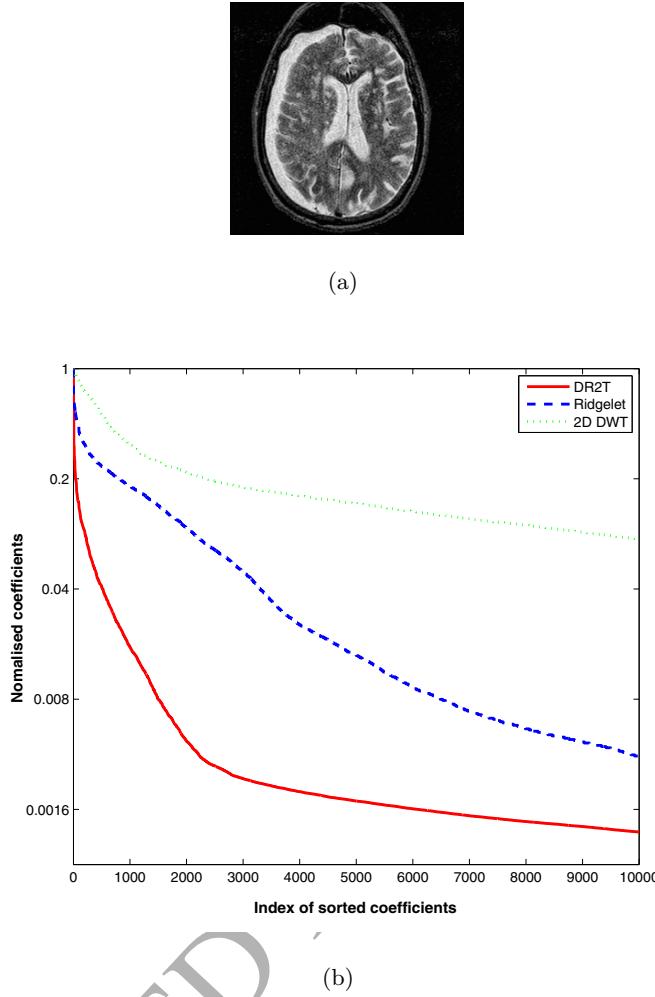


Figure 5: Coefficient decaying comparison among different image transforms

435 6.3. Classification results

For classification of MR images as healthy or pathological, we employ an improved learning algorithm called MPSO-ELM for SLFN classifier. In this section, first we have compared the performance of the proposed MPSO-ELM with other learning algorithms, namely, differential evolution with ELM (DE-ELM), PSO-ELM, ELM and BPNN. It may be noted that the population size and the maximum number of iterations for MPSO-ELM, PSO-ELM and DE-ELM algorithm are kept same i.e., 20 and 30 respectively. The parameters used in the MPSO-ELM algorithm are obtained by experiences and are listed in Table 2. For PSO-ELM, the value of acceleration coefficients  $c_1$  and  $c_2$  are set as 2, while for DE-ELM, the crossover rate and scaling factor ( $F$ ) are set as 0.9 and 0.8 respectively.

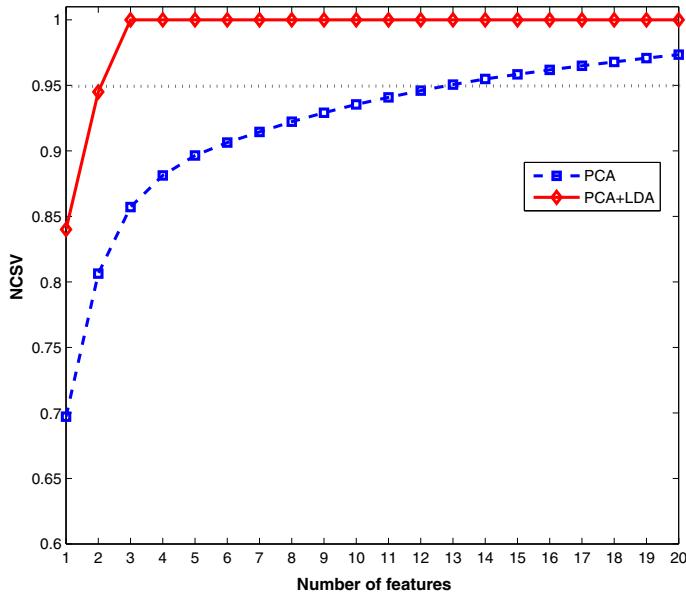


Figure 6: NCSV values with respect to different number of features on the combination of three datasets

Table 2: Parameters list in MPSO-ELM

Parameters	Values	Descriptions
$\omega_i$	0.4	initial inertial weight
$\omega_f$	0.9	final inertial weight
$c_{1i}$	2.5	initial value of cognitive component $c_1$
$c_{1f}$	0.5	final value of cognitive component $c_1$
$c_{2i}$	0.5	initial value of social component $c_2$
$c_{2f}$	2.5	final value of social component $c_2$
$P_j$	[-1,1]	particle initialization

The performance of all algorithms such as MPSO-ELM, PSO-ELM, DE-ELM, ELM, and BPNN are tested on the three benchmark datasets and is reported in Tables 3-5. From the tables, it can be observed that DE-ELM, PSO-ELM, and MPSO-ELM obtain higher accuracy on all the three datasets with comparatively less hidden neurons than basic ELM and BPNN. DE-ELM and PSO-ELM achieve similar accuracy with MPSO-ELM on DS-66 and DS-160, while they earn smaller accuracy compared to MPSO-ELM on DS-255. It can also be seen that classical ELM demands more hidden neurons than BPNN.

Table 3: Performance comparison of different classifiers on DS-66

Classifiers	ACC (%)	Hidden neurons ( $n_h$ )	Norm	Condition number ( $\mathcal{K}_2$ )
BPNN	100.00	5	-	-
ELM	100.00	7	57.3199	6.0980e+03
DE-ELM	100.00	4	21.4401	119.9567
PSO-ELM	100.00	4	26.5115	135.9442
<b>MPSO-ELM</b>	<b>100.00</b>	<b>4</b>	<b>19.1374</b>	<b>106.3816</b>

Table 4: Performance comparison of different classifiers on DS-160

Classifiers	ACC (%)	Hidden neurons ( $n_h$ )	Norm	Condition number ( $\mathcal{K}_2$ )
BPNN	100.00	5	-	-
ELM	99.94	7	167.9688	1.4007e+04
DE-ELM	100.00	3	15.6922	57.1863
PSO-ELM	100.00	3	16.2196	71.1975
<b>MPSO-ELM</b>	<b>100.00</b>	<b>3</b>	<b>12.8221</b>	<b>51.9365</b>

Table 5: Performance comparison of different classifiers on DS-255

Classifiers	ACC (%)	Hidden neurons ( $n_h$ )	Norm	Condition number ( $\mathcal{K}_2$ )
BPNN	99.22	7	-	-
ELM	99.18	10	56.2456	3.6126e+03
DE-ELM	99.57	5	11.1868	104.6756
PSO-ELM	99.49	5	13.0537	392.9668
<b>MPSO-ELM</b>	<b>99.69</b>	<b>5</b>	<b>10.5511</b>	<b>96.4639</b>

Moreover, it is observed that the condition value of the matrix  $\mathbf{H}$  obtained by DE-ELM, PSO-ELM and MPSO-ELM algorithm is smaller as compared to conventional ELM over all the datasets. The norm value of MPSO-ELM, PSO-ELM and DE-ELM is also found to be less than ELM and therefore they can have better generalization performance than basic ELM. Further, it is seen that the smaller norm value of  $w^o$  results in a smaller condition value of matrix  $\mathbf{H}$ . Among DE-ELM, PSO-ELM and MPSO-ELM, MPSO-ELM obtains smaller condition value, smaller norm value and higher accuracy. Therefore, it can be concluded that the proposed MPSO-ELM can achieve better generalization performance with compact networks. The results reported in the tables are the average values of 50 trials and the parameters of all the schemes are determined

through experimental evaluation.

To demonstrate the efficiency of the proposed MPSO-ELM algorithm with three features, accuracy comparison has been made with  $k$ -NN, random forest (RF), and SVM classifier along with BPNN, ELM, and PSO-ELM on all the three datasets, and the results are shown in Figure 7.

- 465 For DS-66, the accuracies earned by  $k$ -NN, BPNN, RF, SVM, ELM, and PSO-ELM are 99.55%, 100.00%, 99.39%, 99.70%, 100.00%, and 100% respectively. The accuracies obtained by  $k$ -NN, BPNN, RF, SVM, ELM, and PSO-ELM are 99.50%, 100.00%, 99.38%, 100.00%, 99.94%, and 100.0% respectively on DS-160; while the accuracies are 99.02%, 99.22%, 99.14%, 99.37%, 99.18%, and 99.49% respectively on DS-255. However, MPSO-ELM earns ideal classification on DS-66 and 470 DS-160, and obtains an accuracy of 99.69% on DS-255 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.

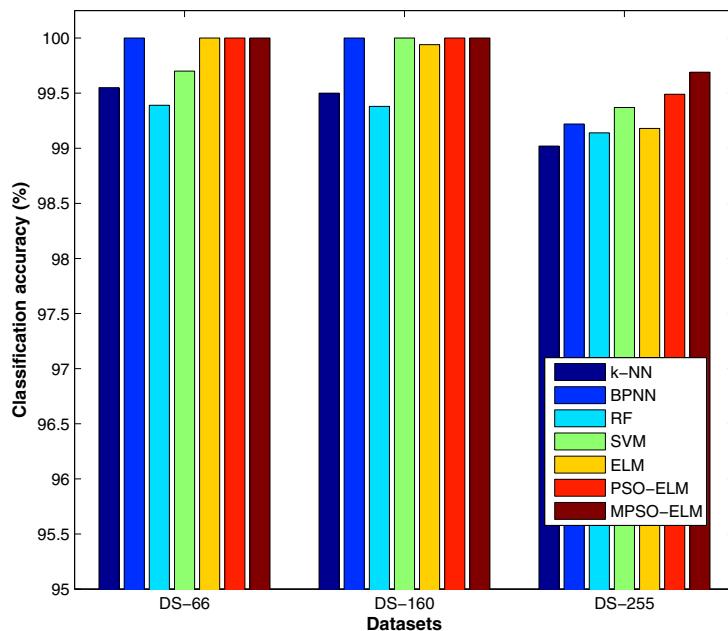


Figure 7: Classification accuracy of different classifiers over three standard datasets

- Table 6 lists the correctly classified samples and the corresponding accuracies obtained by “DR2T+ PCA+LDA + MPSO-ELM” on DS-255 during each trial of a  $10 \times 5$ -fold SCV process. 475 The results in the table indicate that the proposed “DR2T+ PCA+LDA + MPSO-ELM” scheme can correctly classify 2542 samples out of 2550 samples (2200 pathological and 350 healthy samples).

Table 6:  $10 \times 5$ -fold SCV result of DR2T + PCA+LDA + MPSO-ELM scheme over 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)	51 (100.00)	254 (99.61)
2	51(100.00)	50 (98.04)	51 (100.00)	51 (100.00)	50 (98.04)	253 (99.22)
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	51 (100.00)	51 (100.00)	51 (100.00)	51 (100.00)	51 (100.00)	255 (100.00)
6	51 (100.00)	50 (98.04)	51 (100.00)	51 (100.00)	51 (100.00)	254 (99.61)
7	51 (100.00)	51 (100.00)	50 (98.04)	51 (100.00)	50 (98.04)	253 (99.22)
8	51 (100.00)	51 (100.00)	51 (100.00)	51 (100.00)	51 (100.00)	255 (100.00)
9	51 (100.00)	51 (100.00)	51 (100.00)	51 (100.00)	51 (100.00)	255 (100.00)
10	51 (100.00)	49 (96.08)	51 (100.00)	51 (100.00)	51 (100.00)	253 (99.22)
Sum						2542
Average						254.2 (99.69)

Further, among 2200 pathological samples, 2196 are correctly classified by our scheme and the rest four samples are misclassified to healthy class. While among 350 healthy samples, 346 samples are correctly classified by our scheme and rest four samples are misclassified to pathological class.

480 Considering these results, the sensitivity ( $S_e$ ), specificity ( $S_p$ ), and precision values of “DR2T+ PCA+LDA + MPSO-ELM” scheme are computed as 99.82%, 98.86%, and 99.82%, respectively which are listed in Table 7.

To compare the efficacy of PCA+LDA over PCA, we employ both of them separately in the proposed system and named as “DR2T+ PCA + MPSO-ELM” and “DR2T+ PCA+LDA + MPSO-ELM”. The performances of both the schemes over three datasets are shown in Table 7. It may be noticed that the proposed “DR2T+ PCA+LDA + MPSO-ELM” scheme earns better performance than “DR2T+ PCA + MPSO-ELM” on all the datasets with less number of features. For DS-255, “DR2T+ PCA + MPSO-ELM” obtains slightly higher specificity and precision values than “DR2T+ PCA+LDA + MPSO-ELM”. However, the higher the sensitivity value of a CAD system, 490 the better is the performance of the CAD system. Therefore, the proposed “DR2T+ PCA+LDA + MPSO-ELM” scheme holds greater potential in making correct clinical decisions.

Further, in order to support the effectiveness of DR2T features over DWT and curvelet (FDCT) features, we have conducted an experiment where both DWT and FDCT features are separately used in the proposed system and the results are reported in Table 8. It may be seen that the

Table 7: Classification performance (%) of the proposed scheme with PCA and PCA+LDA over three datasets

Dataset	Schemes	DR2T+PCA+MPSO-ELM	DR2T+PCA+LDA+MPSO-ELM
		No. of feature	3
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
	$S_p$	99.50	100.00
	$P_r$	99.92	100.00
	$ACC$	99.62	100.00
DS-255	$S_e$	99.14	99.82
	$S_p$	99.43	98.86
	$P_r$	99.91	99.82
	$ACC$	99.18	99.69

Table 8: Classification accuracy (%) comparison with wavelet and curvelet based method over three datasets

Schemes	No. of features	DS-66	DS-160	DS-255
DWT + PCA + MPSO-ELM	13	100.00	99.31	99.06
DWT + PCA+LDA + MPSO-ELM	3	100.00	99.56	99.22
FDCT + PCA+LDA + MPSO-ELM	3	100.00	99.94	99.61
DR2T + PCA+LDA + MPSO-ELM	3	100.00	100.00	99.69

495 proposed scheme with DR2T features achieves better performance than DWT and FDCT features  
on all the datasets. Here, the DWT features are derived from all the sub-bands of 2-level decomposition.  
In addition, the DWT features used in literature [3, 6, 5] are also tested which results in smaller accuracy compared to the proposed scheme. It may be noted that the FDCT features  
are similar to the literature [62]. It can be concluded here that using DR2T features the proposed  
500 scheme brings potential improvements in the performance. Furthermore, to test the effectiveness  
of CLAHE in the proposed system, we study the results of the system with CLAHE and without  
CLAHE. The results demonstrate that the system without CLAHE attains slightly smaller  
accuracy (i.e., 99.61%).

#### 6.4. Comparison with other PBDSs

505 An extensive comparison with 21 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 9. 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” [4] and “DWPT + TE + GEPSVM” [22] along with our proposed “DR2T + PCA+LDA + MPSO-ELM” scheme offer ideal classification on DS-  
 510 160. Further, it is noticed that no existing PBDSs can earn perfect classification on DS-255, but the suggested system earns higher classification accuracy i.e., 99.69% with a minimum number of features than others. 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 MSPO-  
 515 ELM in the proposed scheme leads to have better generalization performance and faster response on unknown testing data.

#### 6.5. Strengths and weaknesses

From the experiments, it has been shown that the proposed system markedly improves the recent results. The suggested PBDS makes use of discrete ripplet-II transform (DR2T) for feature  
 520 extraction. Unlike other transforms like Fourier transform, DWT, DTCWT, ridgelet, etc., DR2T has the capability of representing 2D singularities along arbitrarily shaped curves which is inherent in MR images. In addition, it provides rotation invariant and sparse features which is essential for performing classification task. Further, MPSO-ELM is proposed to classify the MR images as it possesses interesting properties which enable our system to provide faster response to probe MR  
 525 images. Additionally, it improves the conditioning and produces better generalization performance with a much more compact network. This makes the system to achieve better results as compared to its counterparts.

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  
 530 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. Further, MPSO requires more parameter to tune, so a less parameter based optimization scheme can be investigated in future.

Table 9: 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 [1]	4761	5	98.00	97.15	96.37
DWT + PCA + BPNN + SCG [3]	19	5	100.00	98.29	97.14
DWT + PCA + FNN + SCABC [10]	19	5	100.00	98.93	97.81
DWT + PCA + FNN + ACPSO [9]	19	5	100.00	98.75	97.38
DWT + PCA + KSVM [12]	19	5	100.00	99.38	98.82
DWPT + SE + GEPSVM [22]	16	10	99.85	99.62	98.78
DWPT + TE + GEPSVM [22]	16	10	100.00	100.00	99.33
WE + HMI + GEPSVM [28]	14	10	100.00	99.56	98.63
DWT + PCA + ADBRF [5]	13	10	100.00	99.18	98.35
FRFE + WTT + TSVM [29]	12	10	100.00	99.69	98.98
DTCWT + VE + GEPSVM [32]	12	10	100.00	99.75	99.25
FRFE + WTT + DP-MLP + ARCBBO [30]	12	10	100.00	99.19	98.24
RT + PCA + LS-SVM [4]	9	5	100.00	100.00	99.39
DWT + PCA + $k$ -NN [8]	7	5	98.00	97.54	96.79
FPCNN + DWT + PCA + FNN [6]	7	10	100.00	98.88	98.43
SWT + PCA + IABAP-FNN [13]	7	10	100.00	99.44	99.18
SWT + PCA + ABC-SPSO-FNN [13]	7	10	100.00	99.75	99.02
SWT + PCA + GEPSVM [16]	7	10	100.00	99.62	99.02
WE + NBC [25]	7	10	92.58	91.87	90.51
DWT + PCA + LDA + RF [17]	7	10	100.00	99.75	99.14
MBD + SLFN + PSO-TTC [18]	5	10	100.00	98.19	98.08
<b>DR2T + PCA + MPSO-ELM</b>	<b>13</b>	<b>10</b>	<b>100.00</b>	<b>99.62</b>	<b>99.18</b>
<b>DR2T + PCA+LDA + MPSO-ELM</b>	<b>3</b>	<b>10</b>	<b>100.00</b>	<b>100.00</b>	<b>99.69</b>
<b>(Proposed)</b>					

535 **7. Conclusions and future work**

In this paper, an attempt has been made to develop an improved pathological brain detection system. The proposed scheme initially uses DR2T features to extract the relevant features from the enhanced brain MR images. A PCA+LDA approach has been employed to reduce the feature dimension. Finally, a new learning algorithm called MPSO-ELM is proposed to train the SLFN.

540 The proposed scheme inherits the advantages of DR2T and ELM for detection of the pathological brain from 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 MPSO-ELM method offers several advantages over other methods such as BPNN, SVM, and conventional ELM.

545 This work opens up many research directions. The proposed PBDS has been validated on various accessible datasets which are smaller of size, but a bigger dataset collected online will further prove its potency. To improve the generalization behavior of the proposed scheme, images from various imaging modalities such as CT, MRSI, and PET can be considered. Hybridizing a less parameter-based optimization algorithm with ELM is another possible future work. Deep learning algorithms could be investigated as potential alternatives to the proposed MPSO-ELM. In addition, 550 other popular transforms such as nonsubsampled contourlet transform (NSCT) and contourlet can be tested as the feature extractor.

**References****References**

- 555 [1] S. Chaplot, L. M. Patnaik, N. R. Jagannathan, Classification of magnetic resonance brain images using wavelets as input to support vector machine and neural network, *Biomedical Signal Processing and Control* 1 (1) (2006) 86–92.
- [2] C. Westbrook, *Handbook of MRI technique*, John Wiley & Sons, Oxford, 2014.
- 560 [3] Y. Zhang, Z. Dong, L. Wu, S. Wang, A hybrid method for MRI brain image classification, *Expert Systems with Applications* 38 (8) (2011) 10049–10053.
- [4] S. Das, M. Chowdhury, K. Kundu, Brain MR image classification using multiscale geometric analysis of ripplet, *Progress In Electromagnetics Research* 137 (2013) 1–17.

- [5] D. R. Nayak, R. Dash, B. Majhi, Brain MR image classification using two-dimensional discrete wavelet transform and AdaBoost with random forests, *Neurocomputing* 177 (2016) 188–197.
- 565 [6] E. A. El-Dahshan, H. M. Mohsen, K. Revett, A. B. M. Salem, Computer-aided diagnosis of human brain tumor through MRI: A survey and a new algorithm, *Expert Systems with Applications* 41 (11) (2014) 5526–5545.
- [7] M. Maitra, A. Chatterjee, A Slantlet transform based intelligent system for magnetic resonance brain image classification, *Biomedical Signal Processing and Control* 1 (4) (2006) 299–306.
- 570 [8] E. S. A. El-Dahshan, T. Honsy, A. B. M. Salem, Hybrid intelligent techniques for MRI brain images classification, *Digital Signal Processing* 20 (2) (2010) 433–441.
- [9] Y. Zhang, S. Wang, L. Wu, A novel method for magnetic resonance brain image classification based on adaptive chaotic PSO, *Progress In Electromagnetics Research* 109 (2010) 325–343.
- 575 [10] Y. Zhang, L. Wu, S. Wang, Magnetic resonance brain image classification by an improved artificial bee colony algorithm, *Progress In Electromagnetics Research* 116 (2011) 65–79.
- [11] Y. Zhang, S. Wang, G. Ji, Z. Dong, An MR brain images classifier system via particle swarm optimization and kernel support vector machine, *The Scientific World Journal* 2013 (2013) 1–9.
- 580 [12] Y. Zhang, L. Wu, An MR brain images classifier via principal component analysis and kernel support vector machine, *Progress In Electromagnetics Research* 130 (2012) 369–388.
- [13] S. Wang, Y. Zhang, Z. Dong, S. Du, G. Ji, J. Yan, J. Yang, Q. Wang, C. Feng, P. Phillips, Feed-forward neural network optimized by hybridization of PSO and ABC for abnormal brain detection, *International Journal of Imaging Systems and Technology* 25 (2) (2015) 153–164.
- 585 [14] Y. Chen, L. Shi, Q. Feng, J. Yang, H. Shu, L. Luo, J.-L. Coatrieux, W. Chen, Artifact suppressed dictionary learning for low-dose CT image processing, *IEEE Transactions on Medical Imaging* 33 (12) (2014) 2271–2292.
- [15] Y.-D. Zhang, S. Chen, S.-H. Wang, J.-F. Yang, P. Phillips, Magnetic resonance brain image classification based on weighted-type fractional Fourier transform and nonparallel support vector machine, *International Journal of Imaging Systems and Technology* 25 (4) (2015) 317–590 327.

- [16] Y. Zhang, Z. Dong, A. Liu, S. Wang, G. Ji, Z. Zhang, J. Yang, 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) (2015) 1395–1403.
- [17] D. R. Nayak, R. Dash, B. Majhi, 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, 2015, pp. 1–4.
- [18] Y.-D. Zhang, X.-Q. Chen, T.-M. Zhan, Z.-Q. Jiao, Y. Sun, Z.-M. Chen, Y. Yao, L.-T. Fang, Y.-D. Lv, S.-H. Wang, Fractal dimension estimation for developing pathological brain detection system based on Minkowski-Bouligand method, *IEEE Access* 4 (2016) 5937–5947.
- [19] D. R. Nayak, R. Dash, B. Majhi, Pathological brain detection using curvelet features and least squares SVM, *Multimedia Tools and Applications* (2016) 1–24.
- [20] M. Saritha, K. P. Joseph, A. T. Mathew, Classification of MRI brain images using combined wavelet entropy based spider web plots and probabilistic neural network, *Pattern Recognition Letters* 34 (16) (2013) 2151–2156.
- [21] G. Yang, Y. Zhang, J. Yang, G. Ji, Z. Dong, S. Wang, C. Feng, Q. Wang, Automated classification of brain images using wavelet-energy and biogeography-based optimization, *Multimedia Tools and Applications* (2015) 1–17.
- [22] Y. Zhang, Z. Dong, S. Wang, G. Ji, J. Yang, 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) (2015) 1795–1813.
- [23] D. R. Nayak, R. Dash, B. Majhi, J. Mohammed, Non-linear cellular automata based edge detector for optical character images, *Simulation* (2016) 1–11.
- [24] Y. Zhang, S. Wang, Z. Dong, P. Phillip, G. Ji, J. Yang, Pathological brain detection in magnetic resonance imaging scanning by wavelet entropy and hybridization of biogeography-based optimization and particle swarm optimization, *Progress in Electromagnetics Research* 152 (2015) 41–58.
- [25] X. Zhou, S. Wang, W. Xu, G. Ji, P. Phillips, P. Sun, Y. Zhang, Detection of pathological brain in MRI scanning based on wavelet-entropy and naive Bayes classifier, in: *Bioinformatics and Biomedical Engineering*, 2015, pp. 201–209.

- 620 [26] G. Zhang, Q. Wang, C. Feng, E. Lee, G. Ji, S. Wang, Y. Zhang, J. Yan, 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), 2015, pp. 683–686.
- 625 [27] Y.-D. Zhang, S.-H. Wang, X.-J. Yang, Z.-C. Dong, G. Liu, P. Phillips, T.-F. Yuan, Pathological brain detection in MRI scanning by wavelet packet Tsallis entropy and fuzzy support vector machine, SpringerPlus 4 (1) (2015) 1–16.
- [28] Y. Zhang, S. Wang, P. Sun, P. Phillips, Pathological brain detection based on wavelet entropy and Hu moment invariants, Bio-medical Materials and Engineering 26 (s1) (2015) S1283–S1290.
- 630 [29] S. Wang, Y. Zhang, X. Yang, P. Sun, Z. Dong, A. Liu, T.-F. Yuan, Pathological brain detection by a novel image featurefractional Fourier entropy, Entropy 17 (12) (2015) 8278–8296.
- [30] Y. Zhang, Y. Sun, P. Phillips, G. Liu, X. Zhou, S. Wang, A multilayer perceptron based smart pathological brain detection system by fractional Fourier entropy, Journal of Medical Systems 40 (7) (2016) 1–11.
- 635 [31] S. Wang, P. Phillips, J. Yang, P. Sun, Y. Zhang, Magnetic resonance brain classification by a novel binary particle swarm optimization with mutation and time-varying acceleration coefficients, Biomedical Engineering/Biomedizinische Technik (2016) 1–10.
- [32] S. Wang, S. Lu, Z. Dong, J. Yang, M. Yang, Y. Zhang, Dual-tree complex wavelet transform and twin support vector machine for pathological brain detection, Applied Sciences 6 (6) (2016) 169.
- 640 [33] D. R. Nayak, R. Dash, B. Majhi, Stationary wavelet transform and adaboost with SVM based pathological brain detection in MRI scanning, CNS & Neurological Disorders Drug Targets 16.
- [34] K. A. Johnson, J. A. Becker, The Whole Brain Atlas, <http://www.med.harvard.edu/AANLIB/> (1999).
- 645 [35] S. M. Pizer, R. E. Johnston, J. P. Erickson, B. C. Yankaskas, K. E. Muller, Contrast-limited adaptive histogram equalization: speed and effectiveness, in: Proceedings of the First Conference on Visualization in Biomedical Computing, IEEE, 1990, pp. 337–345.

- [36] E. D. Pisano, S. Zong, B. M. Hemminger, M. DeLuca, R. E. Johnston, K. Muller, M. P. Braeuning, S. M. Pizer, Contrast limited adaptive histogram equalization image processing to improve the detection of simulated spiculations in dense mammograms, Journal of Digital Imaging 11 (4) (1998) 193–200.
- [37] M. N. Do, M. Vetterli, The finite ridgelet transform for image representation, IEEE Transactions on Image Processing 12 (1) (2003) 16–28.
- [38] E. J. Candès, D. L. Donoho, Ridgelets: A key to higher-dimensional intermittency?, Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences 357 (1760) (1999) 2495–2509.
- [39] E. J. Candès, D. L. Donoho, Curvelets- a surprisingly effective nonadaptive representation for objects with edges, Vanderbilt University Press, Nashville, TN (2000) 105–120.
- [40] E. Candes, L. Demanet, D. Donoho, L. Ying, Fast discrete curvelet transforms, Multiscale Modeling & Simulation 5 (3) (2006) 861–899.
- [41] E. J. Candès, D. L. Donoho, New tight frames of curvelets and optimal representations of objects with piecewise  $C^2$  singularities, Communications on Pure and Applied Mathematics 57 (2) (2004) 219–266.
- [42] J. Xu, L. Yang, D. Wu, Ripplet: A new transform for image processing, Journal of Visual Communication and Image Representation 21 (7) (2010) 627–639.
- [43] M. Ghahremani, H. Ghassemian, Remote sensing image fusion using ripplet transform and compressed sensing, IEEE Geoscience and Remote Sensing Letters 12 (3) (2015) 502–506.
- [44] J. Xu, D. Wu, Ripplet transform type II transform for feature extraction, IET Image Processing 6 (4) (2012) 374–385.
- [45] A. Cormack, The radon transform on a family of curves in the plane, Proceedings of the American Mathematical Society 83 (2) (1981) 325–330.
- [46] A. Cormack, The radon transform on a family of curves in the plane. ii, Proceedings of the American Mathematical Society 86 (2) (1982) 293–298.
- [47] J. Yang, J.-y. Yang, Why can LDA be performed in PCA transformed space?, Pattern Recognition 36 (2) (2003) 563–566.

- [48] A. M. Martínez, A. C. Kak, PCA versus LDA, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 23 (2) (2001) 228–233.
- [49] G.-B. Huang, Q.-Y. Zhu, C.-K. Siew, Extreme learning machine: theory and applications, *Neurocomputing* 70 (1) (2006) 489–501.
- 680 [50] G.-B. Huang, Q.-Y. Zhu, C.-K. Siew, Extreme learning machine: a new learning scheme of feedforward neural networks, in: *IEEE International Joint Conference on Neural Networks*, Vol. 2, IEEE, 2004, pp. 985–990.
- [51] G.-B. Huang, D. H. Wang, Y. Lan, Extreme learning machines: a survey, *International Journal of Machine Learning and Cybernetics* 2 (2) (2011) 107–122.
- 685 [52] Q.-Y. Zhu, A. K. Qin, P. N. Suganthan, G.-B. Huang, Evolutionary extreme learning machine, *Pattern Recognition* 38 (10) (2005) 1759–1763.
- [53] R. C. Eberhart, J. Kennedy, Particle swarm optimization, in: *Proceedings of the IEEE conference on Neural Network*, IEEE, 1995, pp. 1942–1948.
- 690 [54] R. C. Eberhart, J. Kennedy, A new optimizer using particle swarm theory, in: *Proceedings of the Sixth International Symposium on Micro Machine and Human Science*, IEEE, 1995, pp. 39–43.
- [55] Y. Shi, R. Eberhart, A modified particle swarm optimizer, in: *IEEE International Conference on Evolutionary Computation Proceedings*, 1998. IEEE World Congress on Computational Intelligence., The 1998, IEEE, 1998, pp. 69–73.
- 695 [56] Y. Xu, Y. Shi, Evolutionary extreme learning machine based on particle swarm optimization, in: *International Symposium on Neural Networks*, Springer, 2006, pp. 644–652.
- [57] G. Zhao, Z. Shen, C. Miao, Z. Man, On improving the conditioning of extreme learning machine: a linear case, in: *7th International Conference on Information, Communications and Signal Processing, ICICS*, IEEE, 2009, pp. 1–5.
- 700 [58] F. Han, H.-F. Yao, Q.-H. Ling, An improved evolutionary extreme learning machine based on particle swarm optimization, *Neurocomputing* 116 (2013) 87–93.
- [59] S. Suresh, R. V. Babu, H. Kim, No-reference image quality assessment using modified extreme learning machine classifier, *Applied Soft Computing* 9 (2) (2009) 541–552.

- [60] A. Ratnaweera, S. K. Halgamuge, H. C. Watson, Self-organizing hierarchical particle swarm optimizer with time-varying acceleration coefficients, *IEEE Transactions on Evolutionary Computation* 8 (3) (2004) 240–255.
- [61] P. L. Bartlett, The sample complexity of pattern classification with neural networks: the size of the weights is more important than the size of the network, *IEEE Transactions on Information Theory* 44 (2) (1998) 525–536.
- [62] D. R. Nayak, R. Dash, B. Majhi, V. Prasad, Automated pathological brain detection system: A fast discrete curvelet transform and probabilistic neural network based approach, *Expert Systems with Applications* 88 (2017) 152–164.



**Deepak Ranjan Nayak** is currently pursuing PhD in Computer Science and Engineering at

715 National Institute of Technology, Rourkela, India. His current research interests include medical image analysis, pattern recognition and cellular automata. He serves 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, and IEEE Access. He is a student member of IEEE.



720

**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  
725 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

730 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  
<sup>735</sup> 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.

ACCEPTED MANUSCRIPT