



Deep extreme learning machine with leaky rectified linear unit for multiclass classification of pathological brain images

Deepak Ranjan Nayak¹ · Dibyasundar Das¹ · Ratnakar Dash¹ · Snehashis Majhi¹ · Banshidhar Majhi¹

Received: 21 September 2018 / Revised: 22 December 2018 / Accepted: 15 January 2019 /

Published online: 27 February 2019

© Springer Science+Business Media, LLC, part of Springer Nature 2019

Abstract

Automatic binary classification of brain magnetic resonance (MR) images has made remarkable progress in the past decade. In comparison, a few pieces of work has been reported on multiclass classification of brain MR images. However, there exist enough scopes for improved automation and accuracy. Most of the existing schemes follow the multi-stage pipeline structure of conventional machine learning framework, where the features are designed manually or hand-crafted. In recent years, deep learning models have attracted great interest from researchers for analyzing medical images that eliminate the traditional steps of machine learning. In this paper, we present an automated method based on deep extreme learning machine (ELM) also termed as multilayer ELM (ML-ELM) for multiclass classification of the pathological brain. ML-ELM is a multilayer architecture stacked with ELM based autoencoders. The effectiveness of leaky rectified linear unit (LReLU) activation function is investigated with ML-ELM. Extensive simulations on a multiclass brain MR image dataset indicate that the ML-ELM with LReLU activation (ML-ELM+LReLU) achieves higher performance with faster training speed compared to its counterparts as well as state-of-the-art schemes. The basic purpose of employing ML-ELM+LReLU algorithm is to eliminate the need for hand-crafted feature extraction and to develop a more stable and generalized system for multiclass brain MR image classification.

Keywords Deep learning · Pathological brain · Magnetic resonance imaging · Leaky rectified linear unit · Extreme learning machine

1 Introduction

Brain diseases are recognized to be the most predominant cause of death among individuals with different age groups across the globe. In general, brain diseases are categorized

✉ Deepak Ranjan Nayak
depakranjannayak@gmail.com; drnayak@iecee.org

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

into various types such as cerebrovascular disease (stroke), degenerative disease, infectious disease and neoplastic disease (brain tumor). These diseases are progressive, and their occurrence increase with age. Early diagnosis is hence of great importance to prevent the severity of these diseases and improve the patient's quality life. Magnetic resonance imaging (MRI), a non-invasive neuroimaging technique, has been profoundly used in biomedical and clinical research, particularly in pathological brain detection [28, 32, 42]. MRI dispenses better resolution of brain tissues as opposed to other imaging techniques such as CT, SPECT and X-ray [12, 19]. However, manual inspection of MR images is onerous, time-consuming and requires skilled supervision. Therefore, automatic computer-aided medical diagnosis (CAMD) system is essential to facilitate fast, reliable, and accurate decisions [29, 37]. The research on automatic pathological brain detection has been extensively studied in the past decade. These studies can be roughly categorized into two groups: (1) binary classification based detection system that segregates pathological MR images from normal MR images, and (2) multiclass classification based system that classifies brain into several categories (different types of brain diseases along with healthy brain).

Quite a large number of automated systems have been proposed for binary classification of brain MR images in the past years. In the following, we enumerate the evolution of these automatic models. Chaplot et al. [2] developed an automatic diagnosis model that derives features from the low-frequency component of discrete wavelet transform (DWT) and employs support vector machine (SVM) for classification. In [4], the DWT features are first subjected to principal component analysis (PCA) and then the classification is carried out using k-NN and feed-forward neural network (FNN) classifiers. Later, Zhang et al. [41–44] have developed various hybrid automatic models via DWT features and classifiers such as FNN and kernel SVM whose parameters are tuned by different optimization techniques. Then, a model based on ripplelet transform (RT) features and least squares SVM (LS-SVM) classifier is proposed in [3]. Later on, Nayak et al. have presented a model in light of DWT features and AdaBoost with random forest classifier. In [27] and [45], the entropy features of low-frequency components at each scale of 8-level DWT decomposition (DWT-AE) are fed to the probabilistic neural network (PNN) for pathological brain detection. While Wang et al. [35] have extracted entropy features from all the components of 8-level DWT decomposition. The potency of wavelet packet Tsallis entropy and Shannon entropy features are individually evaluated with generalized eigenvalue proximal SVM (GEPSSVM) classifier in [46]. Yang et al. [40] have proposed to use DWT energy features and SVM classifier for pathological brain detection. Later, stationary wavelet transform (SWT) based energy and entropy features have been introduced in [21] for achieving improved classification accuracy. The curvelet-based features have been studied in [20, 25] to obtain significant classification results. In [48], a model based on pseudo-Zernike moment and kernel SVM is proposed. Nayak et al. [24] have presented an improved model based on two-dimensional PCA features and evolutionary extreme learning machine (ELM). While in [47], wavelet packet Tsallis entropy features (DWPT-TE) and Jaya optimized ELM classifier are used to build the model. Recently, two contributions are reported to analyze the effect of ripplelet-II features (DR2T) along with two individual improved ELM classifiers on detection results [22, 23].

In comparison, there is scant literature on multiclass brain MR image classification. Kalbkhani et al. [14] have extracted features using DWT and generalized autoregressive conditional heteroscedasticity technique. The classification is carried out using SVM classifier that assigns the MR images into eight different categories (one normal and seven brain diseases). Recently, a five-category classification system is developed using deep stacked sparse autoencoder (SSA) in [12].

The literature studies reveal that almost all existing approaches follow a conventional multi-stage pipeline of feature extraction, feature selection, classification. One of the major concerns in these approaches is the choice of proper feature descriptors and classifiers. DWT has been extensively used for feature extraction despite its shortcomings like limited directional selectivity and shift variance. Moreover, the detection accuracy for multiclass brain MR classification is still far from the real-time requirements. Deep neural-network models, on the other hand, have recently obtained remarkable success in medical image analysis [17, 38]. These models automatically learn the high-level features from the input data through their hierarchical structure and eliminate the need for hand-engineered features. The deep SSA used in [12] encounters many problems. The parameters of SSA are optimized using an iterative procedure that elicits poor learning speed. The weights of the hidden layers of SSA are initialized by independent autoencoders in an unsupervised fashion and then, the whole network is fine-tuned using traditional back-propagation (BP)-based learning algorithm.

In this paper, an automated multiclass classification model based on the deep extreme learning machine is developed to counter the above challenges. The main contributions of the current work are summarized as follows.

- ELM is an emerging learning paradigm for single layer feed-forward neural network (SLFN) that produces better generalization capability at faster learning speed [9, 11]. The deep ELM network, also called ML-ELM, is employed for multiclass classification of MR images that involves a stack of ELM autoencoders to provide high-level feature representations and does not require fine-tuning [15].
- Leaky rectified linear unit (LReLU) has had its success in a wide range of applications compared to sigmoid, tanh and ReLU [18, 39] functions. Hence, LReLU function is taken into consideration in ML-ELM that avoids the computationally expensive operation (in particular exponential). The suggested model is referred to as ML-ELM+LReLU in remainder of the paper.
- The efficacy of the proposed model is evaluated on a multiclass brain MR image dataset.

The remainder of this paper is structured as follows. The detail description of the dataset is presented in Section 2. Section 3 presents the basic concepts and theories of ELM along with its practical issues. The proposed method is detailed in Section 4. Section 5 presents the experimental settings and results. Eventually, the concluding remarks of this work are drawn in Section 6.

2 Materials

The multiclass brain MR dataset comprises 200 images (40 normal and 160 pathological brain images) is used to evaluate the proposed model. The pathological brains contain diseases of four categories, namely brain stroke, degenerative, infectious and brain tumor; each category holds 40 images. The images are sourced from the Harvard Medical School website [13] and they are composed of T2-weighted MR scans acquired along the axial view plane. All the images hold a resolution of 256×256 pixels. The dataset is labeled as ‘Multiclass Harvard Dataset (MCHD)’. Figure 1 depicts typical brain MR samples from each of the five categories.

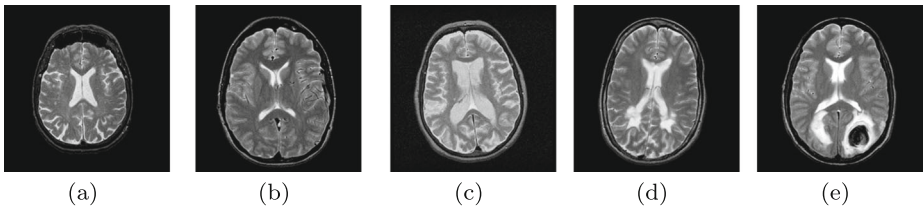


Fig. 1 Typical brain MR images **a** Healthy, **b** Brain stroke, **c** Degenerative, **d** Infectious, and **e** Brain tumor

3 Extreme learning machine (ELM)

ELM developed by Huang et al. [9] is a learning mechanism for SLFN with good generalization capability and fast learning speed. It overcomes the issues of traditional training algorithms and hence, it has been applied in a wide range of classification and regression applications [10, 11]. The principle of ELM is that the hidden node parameters are assigned randomly and kept fixed during training, and the output weights are evaluated analytically by the least square method.

In ELM, the network response at single output node is computed as

$$f_K(\mathbf{x}) = \sum_{i=1}^K w_i^o h_i(\mathbf{x}) = \mathbf{h}(\mathbf{x})w^o$$

where, $w^o = [w_1^o, \dots, w_K^o]^T$ denotes the output weights that links between K hidden nodes and the output node and $\mathbf{h}(\mathbf{x}) = [h_1(\mathbf{x}), \dots, h_K(\mathbf{x})]$ is the output of hidden layer (also called as feature representation) for the input \mathbf{x} which helps in mapping the data from d -dimensional input space to the K -dimensional hidden layer feature space (ELM feature space). ELM is inspired from the Bartlett’s theory for feedforward neural network [1] and it aims to achieve the minimum training error as well as norm of the output weights.

$$\text{Minimize: } \|\mathbf{H}w^o - Y\|^2 \quad \text{and} \quad \|w^o\|$$

where, $Y = [y_1, \dots, y_N]^T$ denotes the target labels and $\mathbf{H} = [\mathbf{h}^T(\mathbf{x}_1), \dots, \mathbf{h}^T(\mathbf{x}_N)]^T$. The output weights w^o can be computed using the Moore-Penrose (MP) generalized inverse of matrix \mathbf{H} as

$$w^o = \mathbf{H}^\dagger Y$$

Another alternative for w^o calculation is reported in [11], which provides a more robust and better generalization performance by introducing a regularization parameter C as

$$w^o = \mathbf{H}^T \left(\frac{I}{C} + \mathbf{H}\mathbf{H}^T \right)^{-1} Y$$

or,

$$w^o = \left(\frac{I}{C} + \mathbf{H}^T \mathbf{H} \right)^{-1} \mathbf{H}^T Y$$

One of the notable characteristics of ELM is that it provides a unified solution to binary-class as well as multiclass classification tasks. However, the basic ELM is shallow in structure and thus, it may not be constructive for feature learning when dealing with natural

signals such as images and videos. The deep learning architecture of ELM called multi-layer ELM (ML-ELM) [15, 31] is shown to be effective for learning meaningful feature representations from images.

4 Deep extreme learning machine with Leaky ReLU

The key challenge remains in traditional machine learning approach is the proper choice of features as they mainly influence the generalization performance. Therefore, careful feature engineering is essential to provide effective representation of the input data. However, designing such engineered features needs domain knowledge and expertise, and thus, takes a lot of time. Multilayer neural networks (MLNN) can be effective in representing the complex data (e.g., images); each layer attempts to learn increasingly high-level features. However, in practice, it is difficult to train MLNN. Hence, recent neural network architectures named as autoencoder (AE) and restricted Boltzmann machine (RBM) [7, 8] have gained significant interests from researchers to perform feature engineering. These networks effectively train MLNNs one layer at a time, and are served as the basic components for building several deep neural networks such as stacked autoencoders (SAE) [34], stacked denoising autoencoders (SDAE) [34], deep belief network (DBN) [8] and deep Boltzmann machine (DBM) [26, 30]. In particular, SAE and SDAE stack the AEs, while DBN and DBM stack the RBMs. These deep networks train their hidden layers individually using either AEs or RBMs in an unsupervised manner and the whole network is then fine-tuned in a supervised fashion using traditional learning method such as BP. Thus, the training of these deep networks is time-consuming and cumbersome. In contrast, the deep ELM (ML-ELM) proposed in [14, 31] facilitates faster and effective learning without the need for fine-tuning. Therefore, ML-ELM is taken into consideration in our work. Besides, we introduce LReLU function in the hidden layer with the aim to improve the learning speed. In the following, we discuss the basic building block of the ML-ELM i.e., ELM autoencoder (ELM-AE) and the deep architecture adopted in the current study.

4.1 ELM autoencoder with Leaky ReLU (ELM-AELR)

ELM theory has been extended to autoencoder known as ELM-AE that serves as the basic building block of ML-ELM. ELM-AE learns to represent powerful features of the input data [15]. Like conventional AE, ELM-AE consists of two parts (i) an encoder and (ii) a decoder as depicted in Fig. 2.

The encoder maps the input $\mathbf{x} = [x^1, x^2, \dots, x^d]$ to a high-level feature representation $\mathbf{h}(\mathbf{x}) = [h_1(x), h_2(x), \dots, h_K(x)]$ using a set of random weights and biases (w^h, b). The (w^h, b) values are made orthogonal according to [15]. Rather than sigmoid, tanh and other non-linear functions that are commonly used in traditional AE and ELM-AE, the leaky ReLU function is taken into consideration in this study (ELM-AELR) for feature mapping because of its potential advantages [18, 39]. Leaky ReLU function, unlike ReLU, maps the negative values to small non-zero slopes, and is mathematically defined as follows

$$\varphi(x) = \begin{cases} x, & x \geq 0 \\ \alpha x, & x < 0 \end{cases}$$

where α is a fixed parameter. LReLU avoids the exponential operation in sigmoid and tanh activation function and hence, helps in faster learning. A clipped ReLU activation can also be considered in place of LReLU [6]. The decoder maps the $\mathbf{h}(\mathbf{x})$ back into input \mathbf{x} through

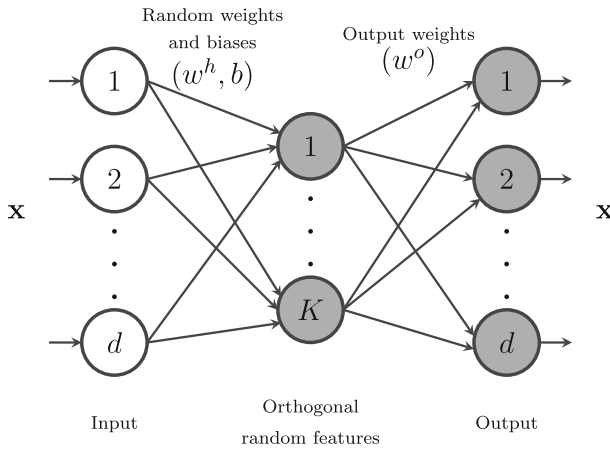


Fig. 2 Network structure of a ELM-AELR. Output is same as input. LReLU activation function is used to compute the hidden output. \mathbf{x} denotes the d -dimensional input, K indicates the number of hidden neurons, w^h indicates the input weights, b represents the bias, and w^o indicates output weights

the output weights w^o that are estimated analytically using $w^o = \mathbf{H}^T (\mathbf{H}\mathbf{H}^T + \frac{I}{C})^{-1} \mathbf{X}$ or $w^o = (\mathbf{H}^T \mathbf{H} + \frac{I}{C})^{-1} \mathbf{H}^T \mathbf{X}$, where $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N]$ denotes the input (or output) data and $\mathbf{H} = [\mathbf{h}_1, \mathbf{h}_2, \dots, \mathbf{h}_N]$ indicates the outputs at each hidden neuron for each input data. It is worth mentioning here that the ELM-AELR can learn three separate representations of the input data similar to ELM-AE –(a) compressed feature representation ($K < d$) (b) equal dimension feature representation ($K = d$), and (c) sparse feature representation ($K > d$).

4.2 Deep ELM with Leaky ReLU (ML-ELM+LReLU)

Similar to other deep networks, ML-ELM+LReLU is a multi-layer network that stacks ELM-AELRs (as shown in Fig. 3). The weights of the hidden layers are assigned by the ELM-AELRs which accomplishes layer-wise unsupervised learning. The output of each hidden layer i in ML-ELM+LReLU can be computed as follows

$$\mathbf{H}_i = \varphi(\mathbf{H}_{i-1} \cdot w_i^{oT}); \quad 1 \leq i \leq L$$

where, $\varphi(\cdot)$ represents the LReLU activation function, \mathbf{H}_i denotes output matrix at the i^{th} hidden layer and, w_i^o is the learned output weights of i^{th} autoencoder and L denotes the number of hidden layers. Note that \mathbf{H}_0 represents the input data \mathbf{X} . Finally, the output weights are evaluated analytically as similar to conventional ELM using the following equation

$$o^w = \mathbf{H}_2^T \left(\frac{I}{C} + \mathbf{H}_2 \mathbf{H}_2^T \right)^{-1} Y$$

or,

$$o^w = \left(\frac{I}{C} + \mathbf{H}_2^T \mathbf{H}_2 \right)^{-1} \mathbf{H}_2^T Y$$

where, \mathbf{H}_2 is the output of last hidden layer and $Y = [y_1, \dots, y_N]^T$ are the target labels. One of the most notable characteristics of the ML-ELM+LReLU is that it does not need

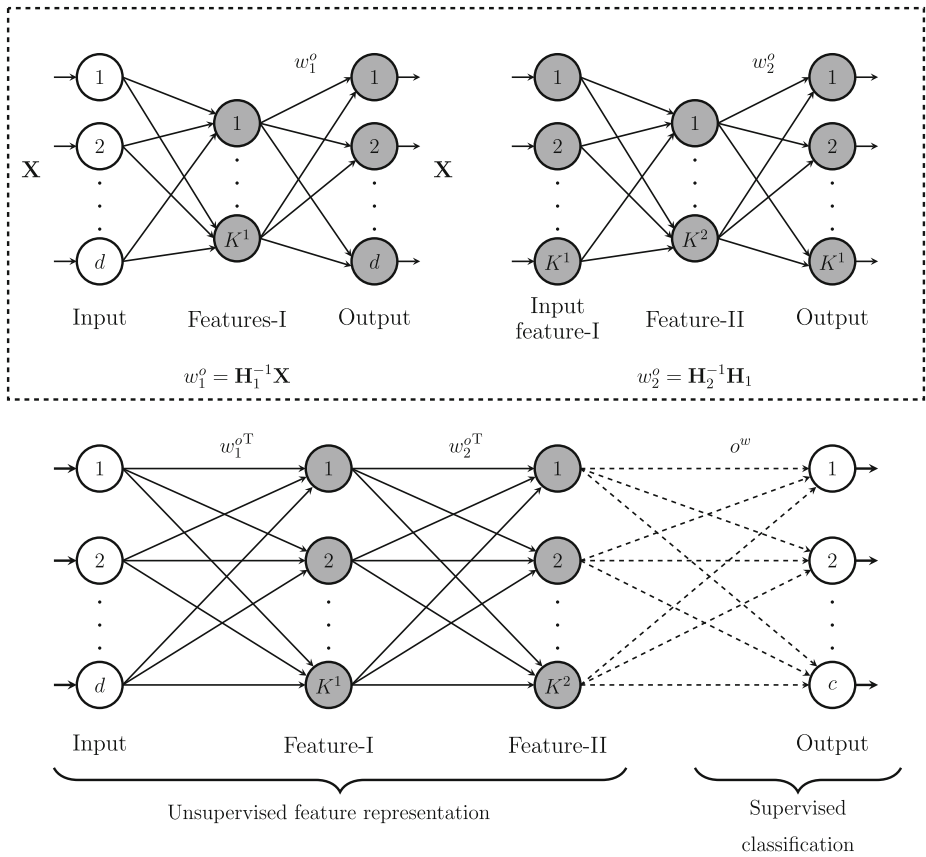


Fig. 3 Architecture of the ML-ELM+LReLU. Figure (above) indicates the two independent ELM-AELRs that performs unsupervised learning. Figure (below) indicates the stacked ELM-AELRs with hidden layers initialized using the pre-trained weights and the output weights (shown in dashed lines) computed using L2-norm regularized least squares. \mathbf{X} denotes the input data, \mathbf{H}_i denotes output matrix at i^{th} hidden layer, K^i indicates the number of hidden neurons at i^{th} hidden layer, w_i^o is the learned output weights of i^{th} autoencoder, and c represents the number of classes

additional fine-tuning as opposed to other traditional deep networks which helps to achieve faster learning.

5 Experimental settings and results

All the programs are developed using MATLAB 2017b environment and are run on a machine with Intel Xeon 2.4 GHz processor and 64 GB RAM. The dataset is divided into two parts– training set and testing set. We have randomly chosen 60% MR samples for training purpose and the rest 40% for testing the model. In particular, 120 and 80 MR samples are chosen for training and testing respectively.

5.1 Experimental setup

We include two hidden layers in the proposed ML-ELM+LReLU network and thus, two individual ELM-AELRs are required to initialize their weights. In ML-ELM+LReLU, we need to set only two hyperparameters, namely, the number of nodes in hidden layers (K) and the regularized parameter (C), while conventional deep neural networks demand more parameters to tune. The hyperparameters K and C need to be chosen meticulously to achieve a good generalization performance. For experiment, we set the value of K and C as $\{50, 100, 150, 200, \dots, 3000\}$ and $\{10^{-10}, 10^{-9}, \dots, 10^9, 10^{10}\}$ respectively. Similar to ML-ELM+LReLU, two-layer architecture is taken into consideration for traditional ML-ELM, SAE and SDAE. The optimal network configuration for ML-ELM+LReLU is experimentally chosen as (256×256) -500-800-5. For fair comparison purpose, we use the similar architecture for original ML-ELM. The regularization parameters for training two individual ELM-AEs are set to 10^{-1} and 10^2 , while it is set as 10^4 for final layer output computation. Further, we choose a (256×256) -100-50-5 network for both SAE and SDAE that demand more user-specified hyperparameters compared to ML-ELM and ML-ELM+LReLU. The number of epochs for these two networks is set to 100 and 200 during pre-training and fine-tuning respectively. The L2 regularization parameters of the two autoencoders are initialized to 0.004 and 0.002. For SDAE, the input corruption rate is assigned to 0.2.

5.2 Data augmentation

The training of deep learning models, in general, requires a very large amount of data in order to provide reliable results. However, obtaining such a large set of medical images is very difficult. One of the effective ways to counter the above issue is data augmentation in which additional images are generated using label-preserving transformations [5]. Moreover, it helps in preventing the network from overfitting issue and thereby, enhancing the performance of the deep network [16]. The number of training samples in the dataset considered is quite less to build a robust deep learning model. Thus, data augmentation is performed over the training samples where each image is subjected to following independent transformations.

- Flipping in horizontal and vertical direction
- Rotation by an angle from $[-45^\circ, 45^\circ]$ with a step size of 5
- Gamma correction with a random r value in the range $[0.7, 1.3]$
- Gaussian noise injection with a variance of 0.01

It is worth noting here that the augmented images that are created by rotation and flipping operation are further subjected to random gamma correction and Gaussian noise injection. The resulting augmented images have the same class label as the original image from which they are obtained. The number of training images is increased by a factor 63, particularly 7560 training samples are generated with the above-mentioned transformations.

Table 1 Performance comparison of ML-ELM+LReLU with its competent methods

Methods	Accuracy (%)	Training time (s)
SAE [8]	88.75	5953.50
SDAE [33]	90.00	6459.27
ML-ELM [15]	91.25	68.72
ELM [11]	82.50	162.50
ML-ELM+LReLU	93.75	46.13

Table 2 Performance evaluation of ML-ELM with different activation functions

Activation function	Accuracy (%)
ML-ELM with sigmoid	91.25
ML-ELM with tanh	88.75
ML-ELM with ReLU	92.50
ML-ELM with LReLU	93.75

5.3 Results and analysis

The classification performance of ML-ELM+LReLU is compared with several relevant methods such as SAE [8], SDAE [33], ML-ELM [15] and ELM [9]. The obtained results over the testing set are listed in Table 1. It is evident that the ML-ELM+LReLU outperforms other deep networks in terms of classification accuracy and training time. Further, an improved accuracy is observed with ML-ELM+LReLU when compare with original ML-ELM and single layer ELM.

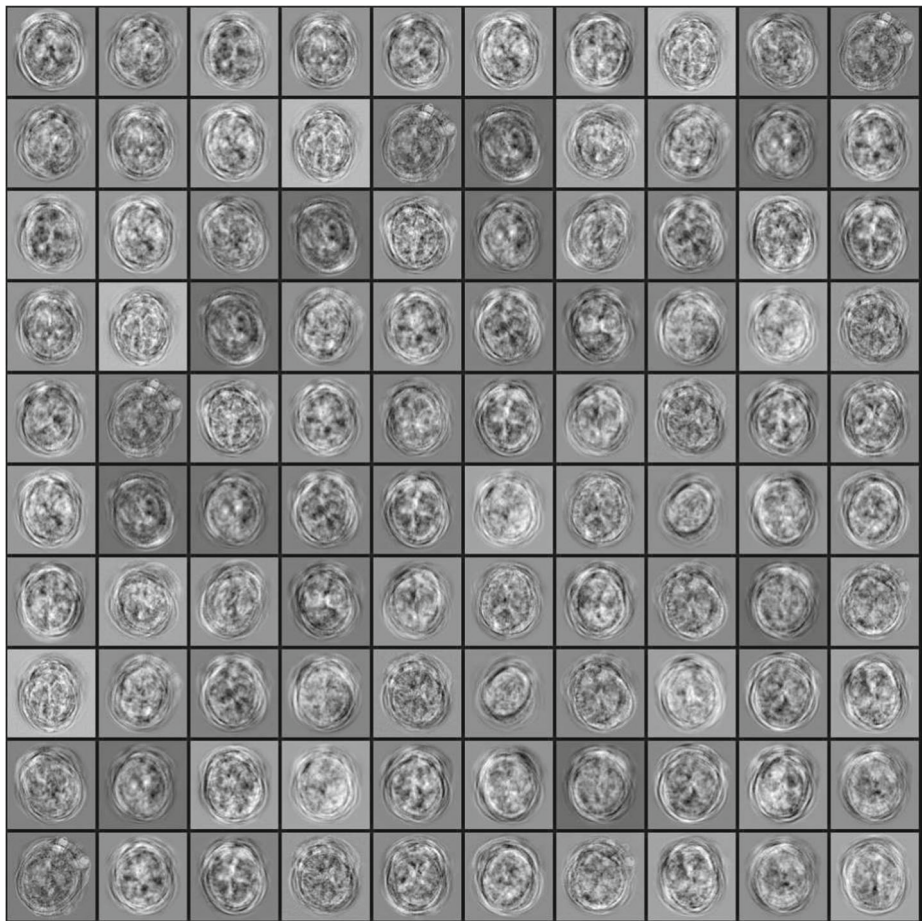


Fig. 4 Visualization of weights learned by the first ELM-AELR

Table 3 Performance comparison of ML-ELM+LReLU with state-of-the-art methods

Method	Accuracy (%)
DWT + PCA + KNN [4]	82.50
DWT–AE + PNN [27, 45]	85.00
DWT + PPCA + ADBRF [19]	87.50
DR2T + PCA+LDA + MPSO–ELM [23]	88.75
DWPT–TE + ELM + Jaya [47]	86.25
SWT–DE + RBF-KSVM [36]	88.75
SSA [12]	90.00
ML–ELM+LReLU (Proposed)	93.75

To demonstrate the effectiveness of LReLU activation function over other non-linear functions in the proposed model, an additional experiment is carried out on the MCHD dataset. The classification results of ML-ELM in presence of various non-linear activation functions are individually tabulated in Table 2. The results show the superiority of LReLU function over sigmoid, tanh and ReLU function.

The representation learned by the encoder of ELM-AELR can be effective in extracting meaningful features from the input brain MR images. Each neuron in the encoder connects to a set of weights that are tuned to represent a specific visual feature. The representation of features learned by the first ELM-AELR of the ML-ELM+LReLU is shown in Fig. 4. In the figure, we have shown the visualization of the weights associated with only 100 neurons (out of 500) in the encoder since the size of the visual weights is significantly large while considering all neurons. The responses of the learned weights demonstrate brain-like structures.

5.4 Comparison with state-of-the-arts

We perform a set of experiments to compare the proposed ML-ELM+LReLU framework with recently published schemes. Table 3 and Fig. 5 show the comparison result among the proposed framework and the state-of-the-art methods over MHCD dataset. It is observed from the table that the proposed framework achieves superior results than other schemes in terms of classification accuracy. It can also be noticed that most of the existing methods except SSA [12] require hand-engineered features. In comparison, the ML-ELM+LReLU learns feature representations directly from the MR image and does not require any engineered features.

Despite the improved classification performance, ML-ELM+LReLU provides faster learning speed due to its following salient features.

1. Compared to traditional autoencoders in which both the input and output weights are trained using an iterative method, the ELM-AELR computes only the output weights using the regularized least squares.
2. A computationally efficient Leaky ReLU function is used in place of the commonly used sigmoid and tanh function.
3. The weights at each hidden layer of ML-ELM+LReLU are initialized by ELM-AELRs, whereas the weights at last layer are computed in a similar fashion to that of single layer ELM.
4. The ML-ELM+LReLU framework does not require additional fine-tuning.

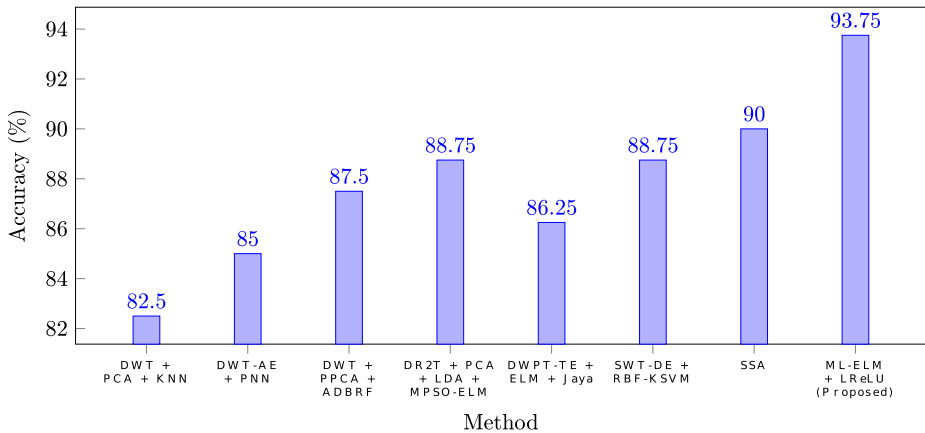


Fig. 5 Accuracy comparison plot between proposed scheme and existing methods

6 Conclusion

In this paper, an automated computer-aided medical diagnosis system is proposed for multiclass classification of brain MR images. The system employs a deep learning model based on multilayer ELM. The leaky ReLU function has been considered for feature mapping that helps in improving the performance as well as the computational speed. The basic purpose of the proposed scheme (ML-ELM+LReLU) is to avoid the manual feature extraction process and achieve good generalization performance with faster training speed. An extensive set of experiments have been performed on a multiclass brain MR dataset to verify the effectiveness of the proposed scheme. The obtained results confirm the superiority of our proposed scheme than its counterparts in terms of classification accuracy and training speed. The proposed ML-ELM+LReLU serves as both a feature extractor and classifier as opposed to the existing schemes.

The efficacy of ML-ELM+LReLU model can be tested on several image classification problems. The application of convolutional neural networks (CNN) could be investigated for multiclass pathological brain detection. At present, our proposed system classifies brain image into a particular brain disorder, but in future, we plan to design a system that can detect more than one disorder simultaneously. In addition, obtaining a larger multiclass brain MR dataset still remains an open challenge.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

1. Bartlett PL (1998) The sample complexity of pattern classification with neural networks: the size of the weights is more important than the size of the network. *IEEE Trans Inf Theory* 44(2):525–536
2. 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

3. Das S, Chowdhury M, Kundu K (2013) Brain MR image classification using multiscale geometric analysis of ripple. *Prog Electromagn Res* 137:1–17
4. El-Dahshan ESA, Honsy T, Salem ABM (2010) Hybrid intelligent techniques for MRI brain images classification. *Digital Signal Process* 20(2):433–441
5. Guo Y, Liu Y, Oerlemans A, Lao S, Wu S, Lew MS (2016) Deep learning for visual understanding: a review. *Neurocomputing* 187:27–48
6. Hannun A, Case C, Casper J, Catanzaro B, Diamos G, Elsen E, Prenger R, Satheesh S, Sengupta S, Coates A et al (2014) Deep speech: Scaling up end-to-end speech recognition. arXiv:1412.5567
7. Hinton GE (2012) A practical guide to training restricted boltzmann machines. In: *Neural networks: tricks of the trade*. Springer, pp 599–619
8. Hinton GE, Salakhutdinov RR (2006) Reducing the dimensionality of data with neural networks. *Science* 313(5786):504–507
9. Huang GB, Zhu QY, Siew CK (2006) Extreme learning machine: theory and applications. *Neurocomputing* 70(1):489–501
10. Huang GB, Wang DH, Lan Y (2011) Extreme learning machines: a survey. *Int J Mach Learn Cybern* 2(2):107–122
11. Huang GB, Zhou H, Ding X, Zhang R (2012) Extreme learning machine for regression and multiclass classification. *IEEE Trans Syst Man Cybern B Cybern* 42 (2):513–529
12. Jia W, Muhammad K, Wang SH, Zhang YD (2017) Five-category classification of pathological brain images based on deep stacked sparse autoencoder. *Multimedia Tools and Applications* pp 1–20
13. Johnson KA, Becker JA The Whole Brain Atlas. <http://www.med.harvard.edu/AANLIB/>
14. Kalbkhani H, Shayesteh MG, Zali-Vargahan B (2013) Robust algorithm for brain magnetic resonance image (MRI) classification based on garch variances series. *Biomed Signal Process Control* 8(6):909–919
15. Kasun LLC, Zhou H, Huang GB, Vong CM (2013) Representational learning with extreme learning machine for big data. *IEEE Intell Syst* 28(6):31–34
16. Krizhevsky A, Sutskever I, Hinton GE (2012) Imagenet classification with deep convolutional neural networks. In: *Advances in neural information processing systems*, pp 1097–1105
17. Litjens G, Kooi T, Bejnordi BE, Setio AAA, Ciompi F, Ghafoorian M, van der Laak JA, Van Ginneken B, Sánchez CI (2017) A survey on deep learning in medical image analysis. *Med Image Anal* 42:60–88
18. Maas AL, Hannun AY, Ng AY (2013) Rectifier nonlinearities improve neural network acoustic models. In: *ICML*, vol 30, p 3
19. 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
20. 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 Syst Appl* 88:152–164
21. 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(2):137–149
22. Nayak DR, Dash R, Majhi B (2018) Development of pathological brain detection system using jaya optimized improved extreme learning machine and orthogonal ripplelet-ii transform. *Multimed Tools Appl* 77(17):22,705–22,733
23. Nayak DR, Dash R, Majhi B (2018) Discrete ripplelet-ii transform and modified PSO based improved evolutionary extreme learning machine for pathological brain detection. *Neurocomputing* 282:232–247
24. Nayak DR, Dash R, Majhi B (2018) An improved pathological brain detection system based on two-dimensional PCA and evolutionary extreme learning machine. *J Med Syst* 42(1):19
25. Nayak DR, Dash R, Majhi B (2018) Pathological brain detection using curvelet features and least squares svm. *Multimed Tools Appl* 77(3):3833–3856
26. Salakhutdinov R, Larochelle H (2010) Efficient learning of deep boltzmann machines. In: *International conference on artificial intelligence and statistics*, pp 693–700
27. 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
28. Shi J, Zheng X, Li Y, Zhang Q, Ying S (2018) Multimodal neuroimaging feature learning with multimodal stacked deep polynomial networks for diagnosis of Alzheimer’s disease. *IEEE J Biomed Health Inform* 22(1):173–183
29. Sinthanayothin C, Boyce JF, Williamson TH, Cook HL, Mensah E, Lal S, Usher D (2002) Automated detection of diabetic retinopathy on digital fundus images. *Diabet Med* 19(2):105–112
30. Srivastava N, Salakhutdinov RR (2012) Multimodal learning with deep boltzmann machines. In: *Advances in neural information processing systems*, pp 2222–2230

31. Tang J, Deng C, Huang GB (2016) Extreme learning machine for multilayer perceptron. *IEEE Trans Neural Netw Learn Syst* 27(4):809–821
32. Turner JA, Potkin SG, Brown GG, Keator DB, McCarthy G, Glover GH (2007) Neuroimaging for the diagnosis and study of psychiatric disorders. *IEEE Signal Proc Mag* 24(4):112–117
33. Vincent P, Larochelle H, Bengio Y, Manzagol PA (2008) Extracting and composing robust features with denoising autoencoders. In: *International conference on machine learning*. ACM, pp 1096–1103
34. Vincent P, Larochelle H, Lajoie I, Bengio Y, Manzagol PA (2010) Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion. *J Mach Learn Res* 11(Dec):3371–3408
35. 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. *Biomedical Engineering/Biomedizinische Technik*, pp 1–10
36. Wang S, Du S, Atangana A, Liu A, Lu Z (2018) Application of stationary wavelet entropy in pathological brain detection. *Multimed Tools Appl* 77(3):3701–3714
37. Wang S, Zhang Y, Zhan T, Phillips P, Zhang Y, Liu G, Lu S, Wu X (2016) Pathological brain detection by artificial intelligence in magnetic resonance imaging scanning. *Prog Electromagn Res* 156:105–133
38. Wong TY, Bressler NM (2016) Artificial intelligence with deep learning technology looks into diabetic retinopathy screening. *Jama* 316(22):2366–2367
39. Xu B, Wang N, Chen T, Li M (2015) Empirical evaluation of rectified activations in convolutional network. [arXiv:1505.00853](https://arxiv.org/abs/1505.00853)
40. Yang G, Zhang Y, Yang J, Ji G, Dong Z, Wang S, Feng C, Wang Q (2016) Automated classification of brain images using wavelet-energy and biogeography-based optimization. *Multimed Tools Appl* 75(23):15,601–15,617
41. 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
42. Zhang Y, Dong Z, Wu L, Wang S (2011) A hybrid method for MRI brain image classification. *Expert Syst Appl* 38(8):10,049–10,053
43. 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
44. 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
45. Zhang Y, Dong Z, Ji G, Wang S (2015) Effect of spider-web-plot in MR brain image classification. *Pattern Recogn Lett* 62:14–16
46. 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. *Entropy* 17(4):1795–1813
47. Zhang YD, Zhao G, Sun J, Wu X, Wang ZH, Liu HM, Govindaraj VV, Zhan T, Li J (2017) Smart pathological brain detection by synthetic minority oversampling technique, extreme learning machine, and jaya algorithm. *Multimedia Tools and Applications*, pp 1–20
48. Zhang YD, Jiang Y, Zhu W, Lu S, Zhao G (2018) Exploring a smart pathological brain detection method on pseudo zernike moment. *Multimed Tools Appl* 77(17):22,589–22,604



Deepak Ranjan Nayak is currently with the Department of Computer Science and Engineering at National Institute of Technology, Rourkela, India. His current research interests include medical image analysis, machine learning, deep learning, and cellular automata. He is currently serving as the reviewer of many peer-reviewed journals and conferences. He has more than twenty publications in journals and conferences. He is a member of IEEE.



Dibyasundar Das is currently pursuing Ph. D in the Computer Science and Engineering at National Institute of Technology, Rourkela, India. He received his B.Tech degree in Information Technology from Biju Patnaik University of Technology, Rourkela, India, in 2011 and his M.Tech degree in Informatics from Siksha 'O' Anusandhan University, India in 2014. His current research interests include optical character recognition, pattern recognition and optimization.



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.



Snehashis Majhi received his B. Tech degree in Computer Science and Engineering from Biju Patnaik University of Technology, Rourkela, India, in 2018. He is currently pursuing M. Tech (R) degree in the Department of Computer Science and Engineering at National Institute of Technology, Rourkela. His major research interests are Machine Learning, Computer Vision and Data mining.



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 serves as reviewer of many international journals and conferences. He is the author and co-author of over 80 journal papers of international repute. Besides, he has 100 conference papers and he holds 2 patents on his name.