



Detection of abnormal brain in MRI via improved AlexNet and ELM optimized by chaotic bat algorithm

Siyuan Lu¹ · Shui-Hua Wang^{2,3,4} · Yu-Dong Zhang^{1,5,6}

Received: 17 February 2020 / Accepted: 3 June 2020 / Published online: 13 June 2020
© Springer-Verlag London Ltd., part of Springer Nature 2020

Abstract

Computer-aided diagnosis system is becoming a more and more important tool in clinical treatment, which can provide a verification of the doctors' decisions. In this paper, we proposed a novel abnormal brain detection method for magnetic resonance image. Firstly, a pre-trained AlexNet was modified with batch normalization layers and trained on our brain images. Then, the last several layers were replaced with an extreme learning machine. A searching method was proposed to find the best number of layers to be replaced. Finally, the extreme learning machine was optimized by chaotic bat algorithm to obtain better classification performance. Experiment results based on 5 × hold-out validation revealed that our method achieved state-of-the-art performance.

Keywords AlexNet · Magnetic resonance image · Deep learning · Extreme learning machine · Computer-aided diagnosis

1 Introduction

Brain is the most sophisticated organ in our body. It is the control center of the nervous system which controls our behavior. So, brain diseases are the most deadly compared with other diseases. Early diagnosis can assist the patients to survive these diseases. Currently, diagnosis of brain disease is dependent on medical imaging. Magnetic resonance image (MRI), computed tomography (CT) and X-ray are common medical imaging modalities in clinical diagnosis. MRI is non-invasion and free of radiation; it can provide clearer imaging results on soft tissue than CT and X-ray. Therefore, it is the first choice for brain disease diagnosis.

Recently, automated medical image analysis is becoming a hot research topic, which requires both medical expertise and machine learning. The developed computer-aided diagnosis (CAD) systems can assist the doctors and physicians to come up with decisions based on medical images. Abnormal brain detection can be regarded as an image recognition and classification problem from the viewpoint of artificial intelligence. A general framework to solve image classification problems often includes feature extraction and classifier training. During the last two

✉ Shui-Hua Wang
sw546@le.ac.uk

✉ Yu-Dong Zhang
yz461@le.ac.uk

Siyuan Lu
sl672@le.ac.uk

- ¹ School of Informatics, University of Leicester, Leicester LE1 7RH, UK
- ² School of Architecture Building and Civil engineering, Loughborough University, Loughborough LE11 3TU, UK
- ³ School of Computer Science and Technology, Henan Polytechnic University, Jiaozuo 454000, Henan, People's Republic of China
- ⁴ School of Mathematics and Actuarial Science, University of Leicester, Leicester LE1 7RH, UK
- ⁵ Department of Information Systems, Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah 21589, Saudi Arabia
- ⁶ Guangxi Key Laboratory of Trusted Software, Guilin University of Electronic Technology, Guilin 541004, China

decades, researchers and practitioners have proposed their methods to detect abnormal brain automatically based on MRI.

These abnormal brain detection methods can be classified toward two groups: traditional machine learning and deep learning. For classical machine learning, the image features are usually handcrafted; much attention is paid to classifier training and optimization. In [1], authors proposed to use discrete wavelet transform (DWT) for feature extraction and employed two classification algorithms for brain MRI classification: neural network self-organizing maps (SOM) and support vector machine (SVM). The SOM and SVM yielded accuracy of 94% and 98%, respectively. El-Dahshan and Hosny [2] proposed a hybrid pathological brain detection method. They firstly employed DWT to extract features from brain MRI. Thereafter, principle component analysis (PCA) was leveraged to reduce the feature dimension. Finally, feedforward back propagation neural network (BPNN) and k nearest neighbors (k -NN) were selected as the classification algorithms. The neural network and k -NN achieved 97% and 98% accuracy, respectively. Kalbkhani and Shayesteh [3] suggested to combine DWT and generalized autoregressive conditional heteroscedasticity (GARCH) for feature generation. PCA and linear discriminant analysis (LDA) were utilized to remove the redundant features. Finally, they trained k -NN and SVM to identify the types of the brain MRIs. Their approach can not only classify abnormal and healthy, but also recognize seven different brain abnormalities. Saritha, Paul Joseph (2013) [4] firstly performed DWT on brain MRIs and extracted entropies from the DWT sub-bands. Then, they proposed to use spider web plots to generate features based on wavelet entropy. Probabilistic neural network (PNN) was trained for image classification which achieved good classification results. El-Dahshan and Mohsen [5] put forward a brain tumor detection system based on brain MRI. The feedback pulse-coupled neural network was trained to segment the brain tumors before classification. Then, DWT with PCA was employed to generate image features. BPNN served as the classification algorithm which recognized the image as abnormal or healthy. The proposed method achieved 99% accuracy on both training and testing samples. Bahadure and Ray [6] put forward their brain tumor segmentation and recognition method. The signal-to-noise ratio of the raw images was improved by pre-processing. Then, they compared the performance of several image segmentation methods including watershed segmentation, fuzzy clustering means, discrete cosine transform and Berkeley wavelet transform, and found out Berkeley wavelet transform performed the best. Then, morphological operation was applied to the segmented images and a set of texture and statistical features were calculated to form the feature

vector. Finally, genetic algorithm (GA) was employed for feature selection and classification. Their system achieved overall accuracy of 92.03%. Gudigar and Raghavendra [7] suggested two image decomposition methods: bidimensional empirical mode decomposition and variational mode decomposition. Then, supervised neighborhood projection embedding and bispectral feature extractor were used to generate feature vector. SVM was trained as the classifier. Experiment results suggested that variational mode decomposition was better than bidimensional empirical mode decomposition with 90.68% accuracy. Acharya and Fernandes [8] proposed a Alzheimer's disease detection system based on brain MRI. A number of different image transforms were employed to extract features including wavelet transform and its variants. Then, Student's t test was utilized for feature selection. Thereafter, k -NN was trained for identification and recognition. There are other reports showing the success of AI and signal processing methods in handling various tasks [9–13].

On the other hand, deep learning techniques usually generate image features in an automated manner. Deep learning is becoming an important tool for image classification in recent five years. Convolutional neural networks (CNNs) have made substantial improvements in image-based machine learning tasks. We no longer need to use image transforms or decomposition methods to extract handcrafted features, because CNN provides a unified framework to implement feature extraction and classification automatically and simultaneously. So, a bunch of deep learning-based abnormal brain detection approaches have been proposed recently. Nayak and Das [14] proposed a multilayer ELM autoencoder with leaky rectified linear unit to classify brain MRIs. The ELM autoencoders were stacked together to form a deep ELM in their experiment of multi-class classification. Deepak and Ameer [15] proposed a brain tumor classification approaches which can distinguish three types of tumors: glioma, meningioma and pituitary tumors. They used pre-trained GoogleNet and transfer learning to implement the classification. The last three layers in the pre-trained GoogleNet were modified and the parameters in early layers remained unchanged. So, the training of the modified GoogleNet was only for determining the weights in the last three layers. Their method achieved good classification performance in experiment. Han and Rundo [16] proposed a data augmentation method for brain tumor detection because the size of medical image datasets is small. A generative adversarial network (GAN)-based brain MRI augmentation algorithm was presented, which improved the classification accuracy. Lu and Lu [17] combined AlexNet and transfer learning for detecting the abnormalities in brain MRI. They used a pre-trained AlexNet modified the last several layers. Then, the whole modified AlexNet was fine-tuned on the

brain MRIs. Their method achieved 100% accuracy on testing set.

From the above analysis, we can find that in abnormal brain detection systems based on classical machine learning, the feature extraction relies on manual image transforms and sometimes requires further feature selection and reduction. But the classifier training is generally faster than deep learning methods because of simple classifier structures and much less parameters. Deep learning methods are capable of generating image features automatically. With convolution and pooling operations in CNN, features can be learned from low level to high level gradually. Nevertheless, training deep CNN models is time-consuming.

The contribution of this study is that we aim to combine the classical machine learning and deep learning technique to obtain the fast training and automated feature learning ability. We improved the performance of pre-trained AlexNet by introducing batch normalization (BN) layers. The improved AlexNet was fine-tuned on our brain MRI dataset. Thereafter, we replaced its last several layers with ELM structure and proposed a searching approach to find the optimal number of layers to be replaced. To obtain better classification results, we optimized the weights and biases in ELM with a novel chaotic bat algorithm. Four different chaotic maps were tested. The evaluation results were all obtained by $5 \times$ hold-out validation. Experimental results suggested that our system achieved good classification performance. Furthermore, our method provided a general framework that can be used in other image classification tasks.

The rest of this paper is as follows. Section 2 presents the brain MRIs in our experiment. Section 3 explains the methods in detail. Section 4 offers the experiment environment and settings, and the experiment research is given in Sect. 5. Finally, Sect. 6 provides the conclusion and future work.

2 Material

The brain MRIs used in this study are obtained from Whole Brain Atlas-Harvard Medical School (website: <http://www.med.harvard.edu/AANLIB/>). The key slices are selected by radiologists of over ten years' experience. Our original dataset contains 177 abnormal samples but merely 28 healthy controls. To balance the ratio of the two classes, we firstly randomly select 14 samples from both classes to form the testing set and the rest samples to form training set. Then, the 14 healthy samples in training set are resampled to get 168 normal samples, which were copied for eleven times. In this way, both training set and testing set are balanced generally. The diseases with abnormal samples include cerebrovascular diseases, neoplastic

diseases, degenerative diseases and inflammatory or infectious diseases. Some samples in our dataset are presented in Fig. 1.

3 Methods

We proposed a novel abnormal brain detection algorithm based on machine learning and deep learning techniques. Firstly, a pre-trained AlexNet was modified and fine-tuned on our brain MRI dataset. Then, we substituted the last several layers in the modified AlexNet with extreme learning machine. Finally, the extreme learning machine was optimized by a novel chaotic bat algorithm to obtain better generalization ability.

3.1 Improved AlexNet

AlexNet is one of the most well-known deep CNN structures proposed by Krizhevsky and Sutskever [18]. AlexNet achieved high classification accuracy on ImageNet dataset, which was a significant breakthrough in machine learning field. Since then, people started to put more time and effort to the research of deep learning models. Various deep CNNs have been invented recently, like ResNet [19], GoogleNet [20], VGG [21], etc., along with numerous training and optimization algorithms.

In this study, we propose to use batch normalization (BN) to improve the robustness of AlexNet for detecting abnormal brain. The distribution of the brain MRIs is complex because of the high variance of human brains. As a result, the distributions of inputs of the layers in AlexNet are different from layer to layer. This can make the parameter training extremely hard and time-consuming, which requires good initialization. In order to overcome this internal covariate shifting, BN is invented. The intuition behind BN is simple. As CNNs are trained in mini-batch mode, BN uses normalization transform on the activations of layers to keep the means and variances fixed. For a random variable x and its values in a mini-batch S :

$$S = [x_1, x_2, x_3, \dots, x_n] \quad (1)$$

The mean μ_S and variance σ_S^2 of x can be obtained by:

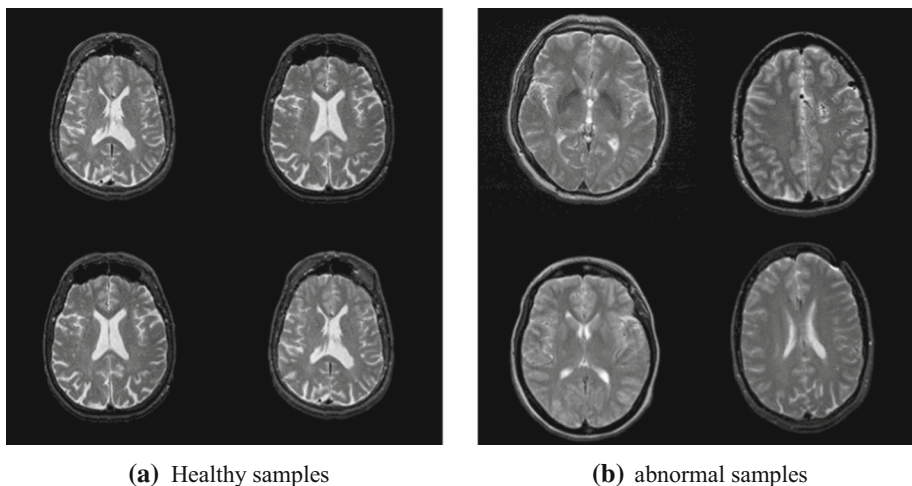
$$\mu_S = \frac{1}{n} \sum_{i=1}^n x_i \quad (2)$$

$$\sigma_S^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \mu_S)^2 \quad (3)$$

So, the normalized values \hat{x}_i can be obtained by:

$$\hat{x}_i = \frac{x_i - \mu_S}{\sqrt{\sigma_S^2 + \varepsilon}} \quad (4)$$

Fig. 1 Some samples of our dataset



Where ε denotes a constant value to increase the numerical stability. Nevertheless, the normalized activations may not be the learning target of the layers in some cases. So, a transformation is added to the result:

$$y_i = \gamma \hat{x}_i + \alpha \tag{5}$$

Where γ and α are two learnable parameters of mini-batch S .

With BN, the training speed of deep CNNs can be accelerated and gradients are less dependent on the initial values of parameters. Furthermore, BN can serve as a regularization, which improves the generalization ability of deep networks.

3.2 ELM

The improved AlexNet can yield good classification performance, but its classification is dependent on the last several layers (mostly fully connected layers). We proposed to replace these layers with a more efficient classifier model: extreme learning machine, to further improve the detection accuracy. ELM is a training algorithm for single-hidden layer feedforward network (SLFN), proposed by Guang-Bin and Qin-Yu [22]. An SLFN contains merely three layers, namely input layer, hidden layer and output layer, shown in Fig. 2. The w and β are the input and output weights, respectively, and b denotes the bias in hidden nodes. The x and o represent the input and output, respectively.

The advantage of ELM is that it is trained without iteration, which makes it converges extremely faster than traditional BPNN [23, 24], and the generalization ability of ELM is also promising [25]. The training algorithm of ELM contains only three steps. Given a training set \mathbf{M} :

$$\mathbf{M} = [(x_1, t_1), (x_2, t_2), (x_3, t_3), \dots, (x_n, t_n)] \tag{6}$$

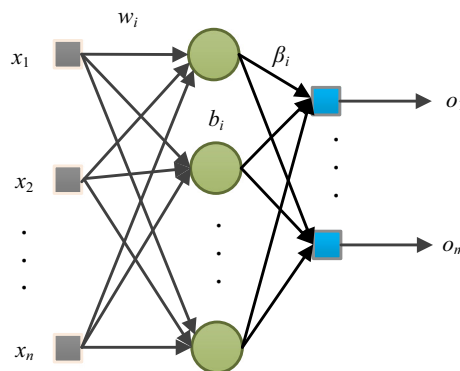


Fig. 2 Architecture of SLFN

Where x_i represents the input vector and t_i denotes the label, ELM firstly initializes the input weight w and bias b with random values. Then, the output matrix of hidden layer \mathbf{H} can be calculated:

$$\mathbf{H} = \sum_{i=1}^{\hat{N}} g_i(w_i x_j + b_i), j = 1, \dots, n \tag{7}$$

Where \hat{N} denotes the number of hidden nodes and $g(\cdot)$ denotes the activation function in hidden layer. Finally, our target is to achieve the ELM output equals to the actual sample labels:

$$\mathbf{H}\beta = \mathbf{T} \tag{8}$$

Where $\mathbf{T} = (t_1, t_2, t_3, \dots, t_n)^T$. So, the β can be obtained by Moore–Penrose pseudo-inverse:

$$\beta = \mathbf{H}^\dagger \mathbf{T} \tag{9}$$

Where \mathbf{H}^\dagger represents the pseudo-inverse of \mathbf{H} . The training algorithm is summarized in Table 1.

From the above analysis, it is clear that the training of ELM is simple to implement. So, ELM is widely applied in practical applications, like recognition [26], prediction [27]

Table 1 Training algorithm of ELM

Pseudocode of ELM	
Input	Training set in Eq. (6)
Step1	Initialize the input weight and bias randomly
Step2	Calculate the output matrix of hidden layer by Eq. (7)
Step3	Obtain the output weight by Eq. (9)
Output	Trained ELM model

and clustering [28]. Therefore, we employ ELM in this study to replace the last several layers for brain MRIs classification. However, the random input weight and bias can have a bad effect on the robustness of the ELM performance; we hope to further optimize these parameters. So, chaotic bat algorithm is proposed to handle this problem.

3.3 SNN

We also employed Schmidt neural network (SNN) and random vector functional link (RVFL) net as classifiers to compare with ELM. SNN and RVFL are both random neural networks, but their structures are different. SNN was proposed by Schmidt and Kraaijveld [29]. There are three layers in SNN, shown in Fig. 3. The weights from input layer to hidden layer were randomly assigned, and there are biases in both hidden layer and output layer. The output of SNN can be expressed as

$$\sum_{i=1}^{\hat{N}} [\beta_i g(\mathbf{w}_i x_j + b_i) + \mathbf{b}] = \mathbf{o}, \quad j = 1, \dots, N \quad (10)$$

Where N denotes the number of hidden nodes. The training of SNN is similar to training of ELM, which can be implemented by pseudo-inverse to get the output weights β .

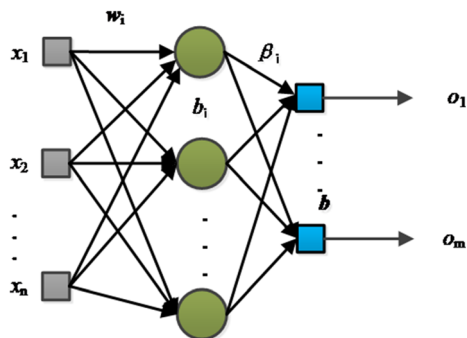


Fig. 3 Architecture of SNN

3.4 RVFL

RVFL was proposed by Pao and Park [30], which is different from ELM and SNN. RVFL firstly maps the input features to enhancement space with random weights and biases. Then, the input features and enhanced features are concatenated to form the feature vector. This structure looks like the shortcut shown in Fig. 4, which is similar to the modules in ResNet. Finally, output weights β are obtained by pseudo-inverse like ELM and SNN.

3.5 Chaotic bat algorithm

Chaotic bat algorithm (CBA) belongs to a swarm intelligent optimization method, which is evolved from bat algorithm [31]. Inspired by the echolocation behavior of bats, CBA uses a set of bats with potential solutions to search the solution space by certain strategies. In every iteration, the parameters of the bats will be updated including the position, velocity and frequency based on the optimal solution found so far. The bat algorithm is better than traditional PSO for optimization, and we introduce chaotic map to improve its searching ability.

Chaotic map is used in updating the positions of bats in our CBA. There are various chaotic maps, and we choose four maps for optimization: sine map, cosine map, Gaussian map and logistic map [32]. The formulae are presented below.

Sine map:

$$x_{k+1} = \mu \sin(\pi x_k) \quad (11)$$

Where k denotes the iteration time and μ the parameter ranging from 0 to 1.

Cosine map:

$$x_{k+1} = \mu \cos(\pi x_k) \quad (12)$$

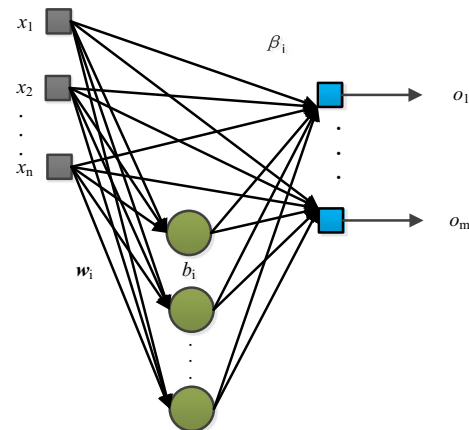


Fig. 4 Architecture of RVFL

Where μ represents the parameter ranging from 0 to 1.

Gaussian map:

$$x_{k+1} = \exp(-\alpha x_k^2) + \beta \tag{13}$$

Where α and β are two parameters of real values.

Logistic map:

$$x_{k+1} = rx_k(1 - x_k) \tag{14}$$

Where r denotes the parameter of positive integer values.

CBA firstly initializes the parameters in the bats with random values. Then, in each iteration, all the bats search the solution space with their velocities and update the solutions using chaotic maps. The best solution in that iteration is obtained by sorting. The iterations will continue until the stop criterion is met. A brief diagram of the CBA is given in Fig. 5.

3.6 BN-AlexNet-ELM-CBA

We propose the abnormal brain detection method based on batch normalized AlexNet, extreme learning machine and chaotic bat algorithm, abbreviated as BN-AlexNet-ELM-CBA. First of all, we employ a pre-trained AlexNet for extracting image feature from brain MRIs. We add the batch normalization layers in the AlexNet model to handle the internal covariate shifting problem. We also modify the last three layers because the original output contains 1000 nodes, but our brain images have only two categories: abnormal and healthy. Totally, six BN layers are added into AlexNet, mainly located after the convolution and pooling

layers. The original fully connected layer ‘fc8’ is also replaced by two new fully connected layers. Because the output matrix dimension of ‘drop7’ is 4096×1 , and the original ‘fc8’ in AlexNet contains 1000 nodes, but our abnormal brain detection is a binary problem. So, we use two layers ‘fc8’ and ‘fc9’ to gradually shrink the dimensions, and the dimensions for ‘fc8’ and ‘fc9’ are 256×1 and 2×1 , respectively. We also construct a transfer-AlexNet (T-AlexNet) for performance comparison. T-AlexNet is built by removing all the batch normalization layers in BN-AlexNet. The three deep CNN structures are given in Fig. 6.

Then, the last several layers in BN-AlexNet are replaced by an ELM classifier. To obtain the optimal layers to be substituted, we proposed to search it by the classification performance. We test the accuracy of our system with n replaced layers based on $5 \times$ hold-out validation and select the best one. The searching algorithm is given in Table 2.

In chaotic bat algorithm optimization for the ELM, the bats contain the input weight w and bias b of the ELM. The fitness function $f(\cdot)$ of CBA is the squared error of predicted labels and actual labels:

$$f(w, b) = \sum_{i=1}^n (o_i - t_i)^2 \tag{15}$$

Where o_i and t_i stand for the output of ELM and the image label, respectively, and n denotes the number of training samples. The solutions in bats are updated with their velocities and chaotic maps:

$$x_i^t = x_i^{t-1} + v_i^t + \lambda \times \text{chaotic}(x_i^{t-1}) \tag{16}$$

Where x_i^t denotes the solution of the i th bat in t th iteration, and λ is the weighting parameter ranging from 0 to 1. In this paper, λ is set as 0.3. All the evaluation is carried out based on $5 \times$ hold-out validation, i.e., we run the systems for five times and calculate the average classification performance for comparison. The pseudocode of our BN-AlexNet-ELM-CBA is presented in Table 3, and a brief diagram is illustrated in Fig. 7. Our method provides a general framework using off-the-shelf deep learning models. The system can be used in other image classification tasks by simple parameter tuning.

4 Experiment

We implemented our BN-AlexNet-ELM-CBA on MATLAB 2018a with deep learning toolbox. The experiment is done on a laptop with i7 7700HQ CPU, GTX 1060 GPU and 16 GB RAM.

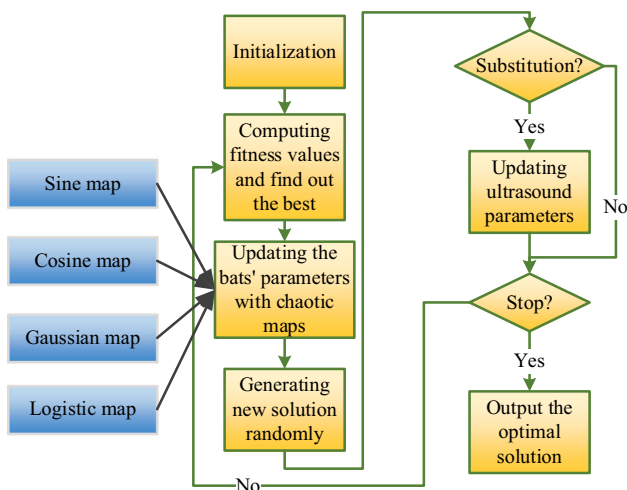


Fig. 5 CBA optimization

Fig. 6 The structures of AlexNet and BN-AlexNet

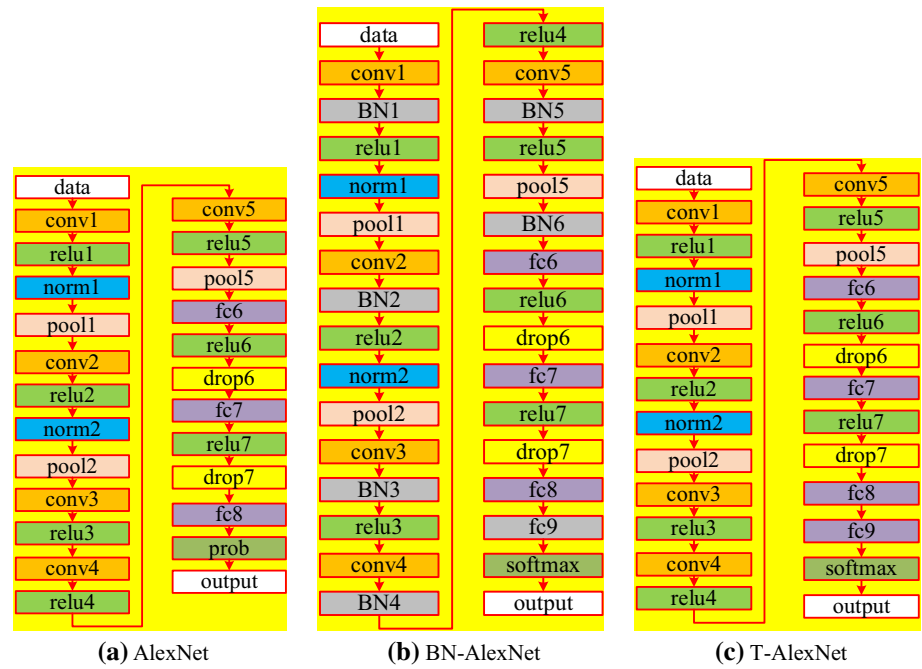


Table 2 Searching algorithm for the optimal layers to be replaced

Pseudocode of searching algorithm for the optimal layers to be replaced	
Input	Training set, testing set and the trained BN-AlexNet
Step1	For $n = 2$ to 6, run Step2 to Step5
Step2	Replace the last n layers in BN-AlexNet with ELM model
Step3	Train the parameters in ELM using chaotic bat algorithm on training set
Step4	Obtain the BN-AlexNet-ELM-CBA classification performance on testing set
Step5	Run from Step2 to Step5 for five times and compute the average accuracy with that n value
Step6	Compare the average performance with different n values and find the best number
Output	The optimal number of layers to be replaced

Table 3 Training of BN-AlexNet-ELM-CBA

Pseudocode of BN-AlexNet-ELM-CBA	
Input	Training set and testing set
Step1	Modify the architecture of a pre-trained AlexNet and obtain the BN-AlexNet
Step2	Train the BN-AlexNet using training set
Step3	Replace the last n layers in BN-AlexNet by ELM structure
Step4	Train the weight and bias in ELM on training set by chaotic bat algorithm
Step5	Obtain the BN-AlexNet-ELM-CBA generalization ability on testing set
Step6	Run from Step1 to Step5 for five times
Output	The optimal number of layers to be replaced

4.1 Dataset

We obtained 359 samples in total and used 331 samples for training and 28 for testing. In training set, there are 163 abnormal samples and 168 normal controls. In testing set, there are 14 images for both classes, respectively. The dataset information is listed in Table 4.

4.2 Hyper-parameter settings

We added 6 batch normalization layers in AlexNet and modified its last three layers to obtain BN-AlexNet. The two fully connected layers ‘fc8’ and ‘fc9’ contained 256 and 2 nodes, respectively. The BN-AlexNet was trained on our brain images by stochastic gradient descent with momentum (SGDM) algorithm. The mini-batch size was

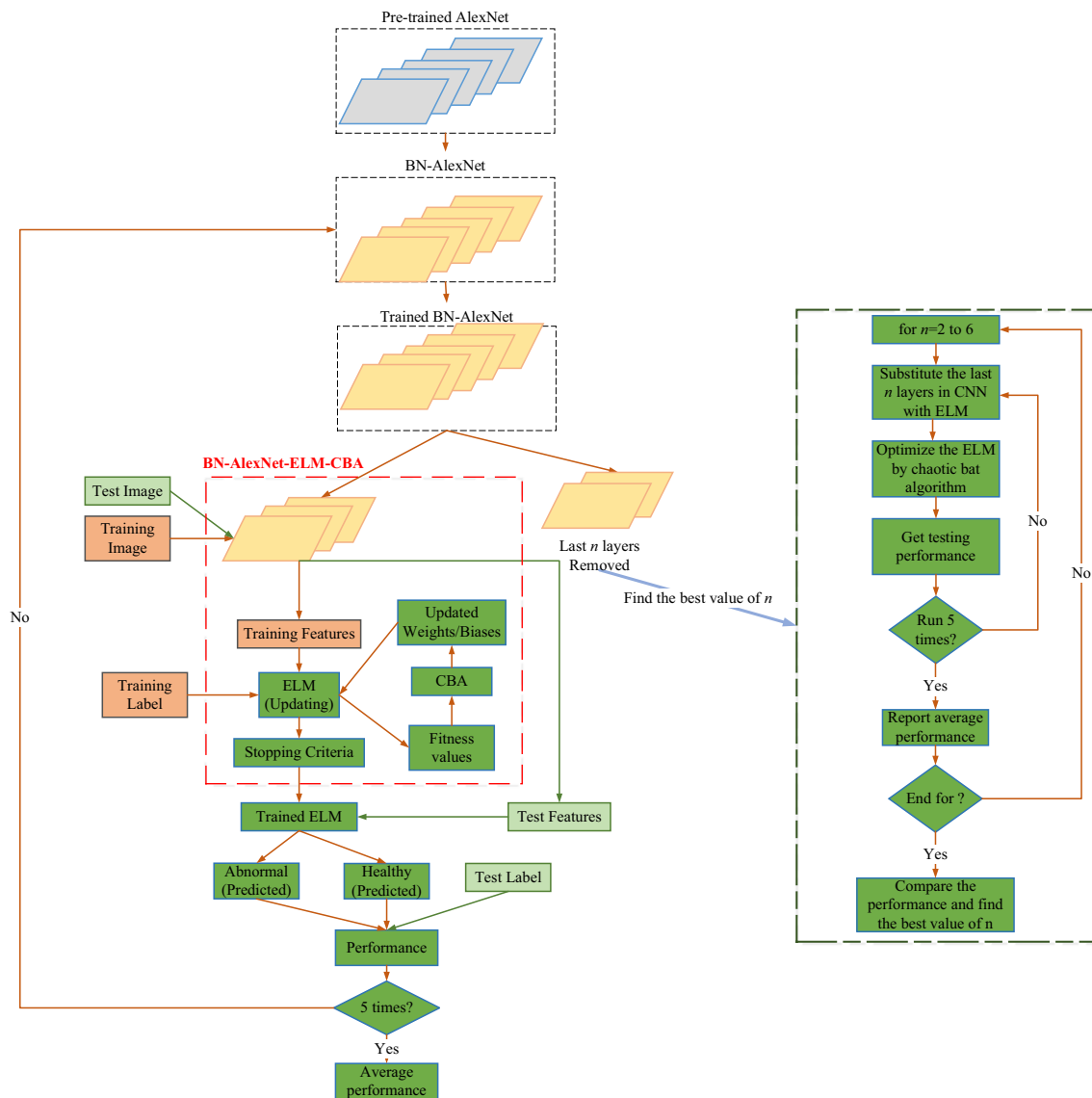


Fig. 7 Flowchart of our BN-AlexNet-ELM-CBA

Table 4 Dataset information

Total samples			
359			
Abnormal samples		Healthy samples	
177	182		
# of samples training set		# of samples testing set	
331		28	
Abnormal	Healthy	Abnormal	Healthy
163	168	14	14

40, maximum epoch was 3, and learning rate was 1e-4. The T-AlexNet was trained on the same settings as BN-AlexNet.

The ELM is a simple structure with only one hyper-parameter: number of hidden nodes. We set 500 hidden nodes in our model considering the input dimension. Finally, the hyper-parameters in CBA were determined. The population of bat particles was 20, and max iteration was 5. The values of all hyper-parameters are provided in Table 5.

4.3 Evaluation measurements

Six widely used measurements were employed to evaluate the classification performance of our method and compare with state-of-the-art approaches: sensitivity, specificity, accuracy, precision, F1 score and Matthew’s correlation coefficient (MCC). They can be computed by following equations:

Table 5 Hyper-parameter settings

Hyper-parameter	Value
Mini-batch	40
Maximum epoch	3
Learning rate	1e-4
# of hidden nodes in ELM	500
# of population of bats	20
Maximum iteration of bats	5

$$\text{Sensitivity} = \frac{TP}{TP + FN} \tag{17}$$

$$\text{Specificity} = \frac{TN}{TN + FP} \tag{18}$$

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{19}$$

$$\text{Precision} = \frac{TP}{TP + FP} \tag{20}$$

$$\text{F1score} = 2 \times \frac{\text{Precision} \times \text{Sensitivity}}{\text{Precision} + \text{Sensitivity}} \tag{21}$$

$$\text{MCC} = \frac{TP \times TN - TP \times FN}{\sqrt{(TP + FP) \times (TP + FN) \times (TN + FP) \times (TN + FN)}} \tag{22}$$

Where *TP* and *FP* denote the numbers of abnormal samples correctly classified and miss classified, respectively, and *TN* and *FN* represent the numbers of healthy samples correctly classified and miss classified, respectively.

5 Results and discussion

5.1 Performance of the proposed method

The classification performance of T-AlexNet and BN-AlexNet is presented in Tables 6 and 7, respectively. The classification results of BN-AlexNet-ELM and BN-AlexNet-ELM-CBA are provided in Tables 8 and 9,

respectively. The training time of T-AlexNet was 18 s for one time running. The comparison of the four methods is provided in Table 10 and Fig. 8. The BN-AlexNet achieved sensitivity of 78.57%, specificity of 97.14% and overall accuracy of 87.86%, which was better than T-AlexNet in terms of specificity and accuracy. So, the introduction of batch normalization did improved the classification performance of AlexNet. The performance of BN-AlexNet-ELM was better than BN-AlexNet with accuracy of 92.86%, and the BN-AlexNet-ELM-CBA outperformed BN-AlexNet-ELM with accuracy of 96.43%. The CBA optimization contributed to the high accuracy of BN-AlexNet-ELM-CBA. Though ELM training is extremely fast which can finish within 0.03 s, the ELM-CBA training can converge in 3 s, which is also acceptable for real-world application. The number of replaced layers was 3, and the chaotic map was Gaussian map in our BN-AlexNet-ELM-CBA. Detailed analysis was provided in following sections.

5.2 Optimal numbers of layers to be replaced

The numbers of layers to be replaced can make a great difference in our system, because the input dimension of the ELM is different with different layers to be replaced. As a result, the training and testing results of ELM vary. Therefore, we proposed to search the optimal replaced layers by evaluating the classification performance of our system with different layers to be replaced. The statistics were the average results of 5 × hold-out validation, shown in Table 11. It is clear that our system achieved over 90% accuracy except with 2 replaced layers. Because the feature dimension was only two, most information was lost. BN-AlexNet-ELM-CBA performed best with 3 layers to be replaced in terms of F1 score and MCC. Though the specificity was the highest when our system was with 5 replaced layers, the sensitivity was only 90.00%. Additionally, 4096 features were much more than 256 features, which inevitably increased the computational complexity. Therefore, the optimal number of layers to be replaced is 3 in this study.

Table 6 Classification performance of T-AlexNet

Run	Sensitivity (%)	Specificity (%)	Accuracy (%)	Precision (%)	F1 score (%)	MCC (%)
1	92.86	92.86	92.86	92.86	92.86	79.59
2	78.57	85.71	82.14	84.62	81.48	50.64
3	71.43	92.86	82.14	90.91	80.00	47.01
4	64.29	100.00	82.14	100.00	78.26	44.25
5	85.71	92.86	89.29	92.31	88.89	67.52
Average	78.57	92.86	85.71	92.14	84.30	57.80

Table 7 Classification performance of BN-AlexNet

Run	Sensitivity (%)	Specificity (%)	Accuracy (%)	Precision	F1 score	MCC
1	85.71	92.86	89.29	92.31	88.89	67.52
2	71.43	100.00	85.71	100.00	83.33	53.24
3	78.57	100.00	89.29	100.00	88.00	63.20
4	85.71	100.00	92.86	100.00	92.31	73.23
5	71.43	92.86	82.14	90.91	80.00	47.01
Average	78.57	97.14	87.86	96.64	86.51	60.84

Table 8 Classification performance of BN-AlexNet-ELM

Run	Sensitivity	Specificity	Accuracy	Precision	F1 score	MCC
1	92.86%	85.71%	89.29%	86.67%	89.66%	73.15%
2	85.71%	92.86%	89.29%	92.31%	88.89%	67.52%
3	92.86%	92.86%	92.86%	92.86%	92.86%	79.59%
4	92.86%	100.00%	96.43%	100.00%	96.30%	86.45%
5	100.00%	92.86%	96.43%	93.33%	96.55%	93.09%
Average	92.86%	92.86%	92.86%	93.03%	92.85%	79.96%

Table 9 Classification performance of BN-AlexNet-ELM-CBA

Run	Sensitivity (%)	Specificity (%)	Accuracy (%)	Precision (%)	F1 score (%)	MCC (%)
1	92.86	100.00	96.43	100.00	96.30	86.45
2	100.00	85.71	92.86	87.50	93.33	86.60
3	100.00	92.86	96.43	93.33	96.55	93.09
4	92.86	100.00	96.43	100.00	96.30	86.45
5	100.00	100.00	100.00	100.00	100.00	100.00
Average	97.14	95.71	96.43	96.17	96.50	90.52

Table 10 Classification performance comparison of our proposed four methods

Method	Sensitivity (%)	Specificity (%)	Accuracy (%)	Precision (%)	F1 score (%)	MCC (%)
T-AlexNet	78.57	92.86	85.71	92.14	84.30	57.80
BN-AlexNet	78.57	97.14	87.86	96.64	86.51	60.84
BN-AlexNet-ELM	92.86	92.86	92.86	93.03	92.85	79.96
BN-AlexNet-ELM-CBA	97.14	95.71	96.43	96.17	96.50	90.25

5.3 Optimal chaotic map

In this study, we tested the performance of our BN-AlexNet-ELM-CBA with four different chaotic maps: sine map, cosine map, Gaussian map and logistic map based on $5 \times$ hold-out validation. The results are given in Table 12. The ‘No map’ means the ELM was trained by bat algorithm with any chaotic maps. We can find that the introduction of chaotic maps generally improves the classification performance, except cosine map. The improvement was obvious in terms of specificity, from 87.14% to the best 95.71% obtained by Gaussian and cosine maps. The accuracy of BN-AlexNet-ELM-CBA with Gaussian map was marginally better than that of

logistic map, and it also achieved the best F1 score and MCC. So, Gaussian map was selected as the optimal chaotic map in our method. Gaussian map provides the better chaotic mechanism for optimization of the CBA. As a result, the ELM can achieve higher generalization ability.

5.4 Comparison of three classifiers

We compared the performance of ELM, SNN and RVFL with the same image features from BN-AlexNet and CBA optimization, i.e., the BN-AlexNet-ELM-CBA, BN-AlexNet-SNN-CBA and BN-AlexNet-RVFL-CBA. The statistics is presented in Table 13, which was obtained by $5 \times$ hold-out validation. It can be seen that all the three

Fig. 8 Comparison of the four proposed methods

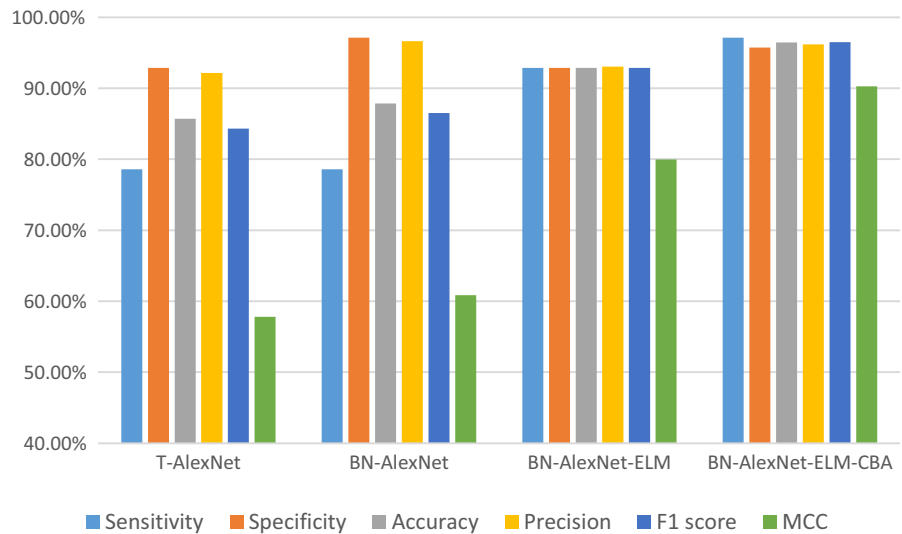


Table 11 Performance of BN-AlexNet-ELM-CBA with different number of replaced layers

# of replaced layers	Feature dimension	Sensitivity	Specificity	Accuracy	Precision	F1 score	MCC
2	2	98.57%	57.14%	77.86%	69.70%	81.66%	60.34%
3	256	97.14%	95.71%	96.43%	96.17%	96.50%	90.25%
4	4096	92.86%	91.43%	92.14%	91.55%	92.20%	78.27%
5	4096	90.00%	100.00%	95.00%	100.00%	94.74%	80.83%
6	4096	97.14%	92.86%	94.29%	93.15%	95.10%	87.51%

Table 12 Performance of BN-AlexNet-ELM-CBA with different chaotic maps

Chaotic map	Sensitivity (%)	Specificity (%)	Accuracy (%)	Precision (%)	F1 score (%)	MCC (%)
No map	97.14	87.14	92.14	87.18	91.89	81.02
Sine map	97.14	90.00	93.57	93.15	95.10	87.51
Cosine map	87.14	95.71	91.43	95.24	91.01	70.15
Gaussian map	97.14	95.71	96.43	96.17	96.50	90.25
Logistic map	97.14	92.86	95.00	93.15	95.10	87.51

Table 13 Performance of different classifier structures

Method	Sensitivity	Specificity	Accuracy	Precision	F1 score	MCC
BN-AlexNet-ELM-CBA	97.14%	95.71%	96.43%	96.17%	96.50%	90.25%
BN-AlexNet-SNN-CBA	98.57%	91.43%	95.00%	92.00%	95.17%	88.94%
BN-AlexNet-RVFL-CBA	98.57%	88.57%	93.57%	89.61%	93.88%	86.33%

methods achieved over 90% accuracy. The sensitivity of BN-AlexNet-SNN-CBA and BN-AlexNet-RVFL-CBA is 98.57% which is marginally better than BN-AlexNet-ELM-CBA, but the specificity of BN-AlexNet-ELM-CBA was the best among the three. Moreover, BN-AlexNet-ELM-CBA yielded over 90% performance for all the six

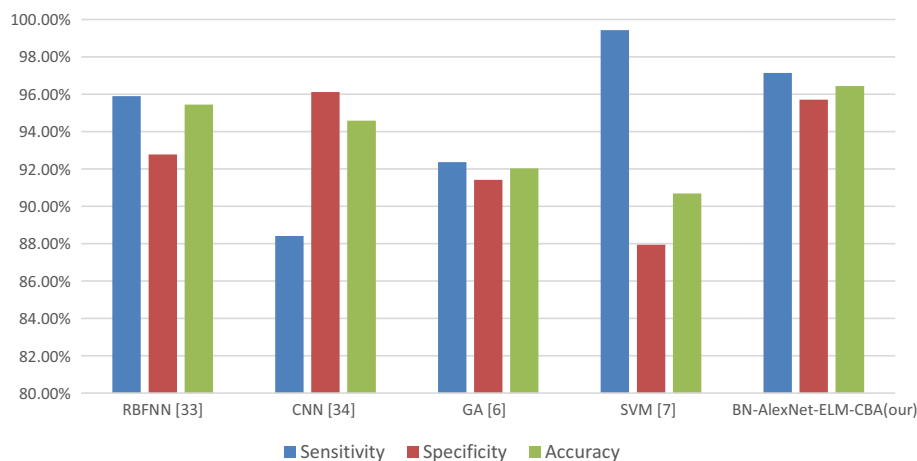
measurements, so we think it is marginally better than the other two methods.

5.5 Comparison with state-of-the-art methods

We offer a comparison between our BN-AlexNet-ELM-CBA and state-of-the-art methods in detecting abnormal

Table 14 Performance comparison

Methods	Sensitivity (%)	Specificity (%)	Accuracy (%)	Validation
RBFNN [33]	95.89	92.78	95.44	10 × 10-fold cross-validation
CNN [34]	88.41	96.12	94.58	Hold-out validation
GA [6]	92.36	91.42	92.03	No validation
SVM [7]	99.43	87.95	90.68	10 × 10-fold cross-validation
BN-AlexNet-ELM-CBA (ours)	97.14	95.71	96.43	5 × hold-out validation

Fig. 9 Comparison with state-of-the-art methods

brains in MRIs. The state-of-the-art methods include: RBFNN [33], CNN [34], GA [6] and SVM [7]. The detailed information is listed in Table 14. It is obvious that SVM yielded the best sensitivity and CNN achieved the best specificity. However, the difference between sensitivity and specificity of the two methods was relatively large, which resulted in the low accuracy. Our BN-AlexNet-ELM-CBA was marginally worse than SVM and CNN in terms of sensitivity and specificity, respectively, and achieved the best accuracy among the methods. Meanwhile, BN-AlexNet-ELM-CBA was also robust because of the small differences between the three measurements (Fig. 9).

6 Conclusion

In this study, we proposed four novel abnormal brain diagnosing method for brain MRI: T-AlexNet, BN-AlexNet, BN-AlexNet-ELM and BN-AlexNet-ELM-CBA. Experiment result revealed that BN-AlexNet-ELM-CBA was the best of the four with sensitivity of 97.14%, specificity of 95.71% and overall accuracy of 96.43%. Our method leveraged the feature learning ability of deep neural network and the particle intelligence for ELM optimization for classification on small dataset. The

introduction of batch normalization and chaotic bat algorithm improved the generalization ability of our system. Moreover, our method can provide a general framework to search the optimal feature layers in deep CNN models, which is applicable to other image classification tasks.

However, our system can only classify brain images as abnormal or healthy, multi-class classification is more useful in clinical diagnosis, which is one of our future research directions. We shall collect more samples and build bigger dataset to re-test our method. Because for deep models, generally, the bigger the training set is, the better they can perform. We shall adopt other swarm intelligent algorithms to optimize the ELM in the future.

Acknowledgement This paper was partially supported by Royal Society International Exchanges Cost Share Award, UK (RP202G0230); Medical Research Council Confidence in Concept (MRC CIC) Award, UK; Hope Foundation for Cancer Research, UK (RM60G0680); British Heart Foundation Accelerator Award, UK; Guangxi Key Laboratory of Trusted Software (kx201901); and Henan Key Research and Development Project (182102310629); Fundamental Research Funds for the Central Universities (CDLS-2020-03); Key Laboratory of Child Development and Learning Science (Southeast University), Ministry of Education.

Compliance with ethical standards

Conflict of interest The authors declare that there is no conflict of interest regarding the publication of this article.

References

- Chaplot S et al (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
- El-Dahshan E-SA et al (2010) Hybrid intelligent techniques for MRI brain images classification. *Digit Signal Process* 20(2):433–441
- Kalbkhani H et al (2013) Robust algorithm for brain magnetic resonance image (MRI) classification based on GARCH variances series. *Biomed Signal Process Control* 8(6):909–919
- Saritha M et al (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
- El-Dahshan E-SA et al (2014) Computer-aided diagnosis of human brain tumor through MRI: a survey and a new algorithm. *Expert Syst Appl* 41(11):5526–5545
- Bahadure NB et al (2018) Comparative approach of MRI-Based brain tumor segmentation and classification using genetic algorithm. *J Digit Imaging* 31(4):477–489
- Gudigar A et al (2019) Automated categorization of multi-class brain abnormalities using decomposition techniques with mri images: a comparative study. *IEEE Access* 7:28498–28509
- Acharya UR et al (2019) Automated detection of alzheimer's disease using brain mri images- a study with various feature extraction techniques. *J Med Syst* 43(9):302
- Pandey HM et al (2016) Grammar induction using bit masking oriented genetic algorithm and comparative analysis. *Appl Soft Comput* 38:453–468
- Pandey HM et al (2014) A comparative review of approaches to prevent premature convergence in GA. *Appl Soft Comput* 24:1047–1077
- Pandey HM et al (2016) Maintaining regularity and generalization in data using the minimum description length principle and genetic algorithm: case of grammatical inference. *Swarm Evolut Comput* 31:11–23
- Pandey HM et al (2019) A comprehensive classification of deep learning libraries. In: Yang XS et al (eds) *Third international congress on information and communication technology*. Springer, Cham, pp 427–435
- Pandey HM et al (2018) Bit mask-oriented genetic algorithm for grammatical inference and premature convergence. *Int J Bio-Inspir Comput* 12(1):54–69
- Nayak DR et al (2019) Deep extreme learning machine with leaky rectified linear unit for multiclass classification of pathological brain images. *Multimed Tools Appl*
- Deepak S et al (2019) Brain tumor classification using deep CNN features via transfer learning. *Comput Biol Med* 111:103345
- Han C et al (2019) Combining noise-to-image and image-to-image GANs: brain MR image augmentation for tumor detection. *IEEE Access* 7:156966–156977
- Lu S et al (2019) Pathological brain detection based on AlexNet and transfer learning. *J Comput Sci* 30:41–47
- Krizhevsky A et al (2012) ImageNet classification with deep convolutional neural networks. In: *International conference on neural information processing systems*, p. 1097–1105
- He K et al (2016) Deep residual learning for image recognition. In: *The IEEE conference on computer vision and pattern recognition (CVPR)*. p 770–778
- Szegedy C et al (2014) Going deeper with convolutions
- Simonyan K et al (2015) Very deep convolutional networks for large-scale image recognition. *International conference on learning representations*
- Guang-Bin H et al (2006) Extreme learning machine: theory and applications. *Neurocomputing* 70(1–3):489–501
- Huang G-B et al (2010) Optimization method based extreme learning machine for classification. *Neurocomputing* 74(1–3):155–163
- Tang J et al (2015) Extreme learning machine for multilayer perceptron. *IEEE Trans Neural Netw Learn Syst* 27(4):809–821
- Wang Y et al (2011) A study on effectiveness of extreme learning machine. *Neurocomputing* 74(16):2483–2490
- Zhang L et al (2017) Deep object recognition across domains based on adaptive extreme learning machine. *Neurocomputing* 239:194–203
- Yang X-C et al (2015) Pressure prediction of coal slurry transportation pipeline based on particle swarm optimization kernel function extreme learning machine. *Math Probl Eng* 2015:1–7
- Liu T et al (2018) Extreme learning machine for joint embedding and clustering. *Neurocomputing* 277:78–88
- Schmidt WF et al (1992) Feedforward neural networks with random weights, In: *Proceedings., 11th IAPR international conference on pattern recognition. Vol.II. conference b: pattern recognition methodology and systems*. p 1–4
- Pao YH et al (1994) Learning and generalization characteristics of random vector functional-link net. *Neurocomputing* 6:163–180
- Yang X-S (2010) A new metaheuristic Bat-inspired algorithm. In: *Nature inspired cooperative strategies for optimization (NICSO 2010)*, vol 284, pp 65–74
- Arasomwan AM et al (2014) An investigation into the performance of particle swarm optimization with various chaotic maps. *Math Prob Eng* 2014:1–17
- Lu Z et al (2016) A pathological brain detection system based on radial basis function neural network. *J Med Imaging Health Inf* 6(5):1218–1222
- Sajjad M et al (2019) Multi-grade brain tumor classification using deep CNN with extensive data augmentation. *J Comput Sci* 30:174–182

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