1181: MULTIMEDIA-BASED HEALTHCARE SYSTEMS USING COM-
PUTATIONAL INTELLIGENCE

A review on extreme learning machine

Jian Wang[1] · Siyuan Lu[1] · Shui-Hua Wang[2,3,4,5] · Yu-Dong Zhang[1,2,4] (iD)

Abstract

Extreme learning machine (ELM) is a training algorithm for single hidden layer feedforward neural network (SLFN), which converges much faster than traditional methods and yields promising performance. In this paper, we hope to present a comprehensive review on ELM. Firstly, we will focus on the theoretical analysis including universal approximation theory and generalization. Then, the various improvements are listed, which help ELM works better in terms of stability, efficiency, and accuracy. Because of its outstanding performance, ELM has been successfully applied in many real-time learning tasks for classification, clustering, and regression. Besides, we report the applications of ELM in medical imaging: MRI, CT, and mammogram. The controversies of ELM were also discussed in this paper. We aim to report these advances and find some future perspectives.

1 Introduction

Neural network, which usually refers to artificial neural network, was firstly proposed in last century for a history of more than 70 years [114]. Neural network was inspired by biological neural network and its basic structure mimics the nerve system in human brain. As a kind of

✉ Shui-Hua Wang
shuihuawang@ieee.org

✉ Yu-Dong Zhang
yudongzhang@ieee.org

Jian Wang
jw830@le.ac.uk

Siyuan Lu
sl672@le.ac.uk

Extended author information available on the last page of the article

machine learning method, a typical neural network consists of neurons, connections, and weights. In neural network, a neuron, also called perception [152], receives input signals from former neurons and sends out output signals to later neurons. Neurons within a neural network are often organized in different layers. The connection was defined as relation between neurons of different layers. A neuron may have lots of connections to neurons of its former layer and later layer. And each connection delivers an important parameter called weight. It is weights that decide how a neuron processes input signals to give out output signals.

To train the network is to determine the weights in its layers that makes it best approximates the data. Traditional training algorithms are often based on gradient, like back propagation (BP) [191]. BP was widely used in many domains because it is easy to understand and simple to implement. However, gradient based algorithms do not always come up with global best solution. Their solutions are dependent on the parameter initialization and the complexity of the feature space, so they are more likely to converge at local extrema. Extreme learning machine (ELM) was proposed by Guang-Bin and Qin-Yu [41], which was aim to train single-hidden layer feedforward networks (SLFNs). Different than gradient based methods, ELM assigns random values to the weights between input and hidden layer and the biases in the hidden layer, and these parameters are frozen during training. The nonlinear activation functions in hidden layer provide nonlinearity for the system. Then, it can be regarded as a linear system. The only parameter needs to learn is the weight between hidden layer and output layer. Hence, ELM converges much faster than traditional algorithms because it learns without iteration. Meanwhile, random hidden nodes promise the universal approximation ability. Theoretical analysis showed that ELM are more likely to reach global optimal solution with random parameters than traditional networks with all the parameters to be trained [81]. Compared with support vector machine (SVM) [23], ELM tends to yield better classification performance with less optimization constrains [55]. Due to its superior training speed and good generalization capability [82], ELM is widely applied in a variety of learning problems, such as classification, regression, clustering, and feature mapping. ELM evolved as many variants have been proposed to further improve its stability and generalization for specific applications [29, 61, 86, 223].

Over the past decade, extensive research on ELM has been carried out for three purpose: less manual intervention, higher classification accuracy and less training time. Theoretical analysis has also been done to investigate the generalization and global approximation ability of ELM [55].

In this study, we try to provide a comprehensive review on the development of ELM, including the theoretical analysis, its variants, recent advances, and real applications. The rest of this paper is organized as follows. In section 2, we will give the detailed theoretical analysis of basic ELM. Section 3 present recent advances on ELM for classification, regression, and clustering. Section 4 is a comparison between ELM and other machine learning algorithms. Section 5 gives a summary on machine learning tasks including recognition, prediction, representation, clustering, and surrogate modeling of ELM. Section 6 summarized the application fields where ELM was adopted such as medical analysis, chemistry, transportation, economy, robotics, management, geography, and food safety. Finally, Section 8 concludes the review.

2 Theoretical fundamentals

In this part, we are to show the theoretical analysis of basic ELM. ELM was invented to train SLFNs, which is the most widely used artificial neural network structure. A conventional

SLFN consists of three layers: input layer, hidden layer and output layer, shown in Fig. 1. The notations are given in Table 1. x and o denote the input and output vector. w and b represent the weight from input to hidden layer and the bias of hidden layer. β denotes the output weight. Training the network is to decide these parameters that reach the optimal solution.

2.1 SLFN training

In this section, we will briefly introduce the training problem for SLFN. Given a training set $S = \{(x_i, t_i) | x_i = (x_{i1}, x_{i2}, \ldots, x_{in})^T \in R^n, \ t_i = (t_{i1}, t_{i2}, \ldots, t_{im})^T \in R^m\}$, where x_i denotes the input value and t_i represents the target, the output o of an ELM with $\widehat{N}$ hidden neurons can be expressed as:

$$\sum_{i=1}^{\widehat{N}} \beta_i g\left(w_i x_j + b_i\right) = o_j, j = 1, \ldots, N \tag{1}$$

Where $g(x)$ means the activation function in the hidden layer. In ELM, activation functions are nonlinear ones to provide nonlinear mapping for the system. Table 2 lists several widely used activation functions.

The goal of training is to minimize the error between the target and the output of ELM. The most commonly used object function is mean squared error (MSE):

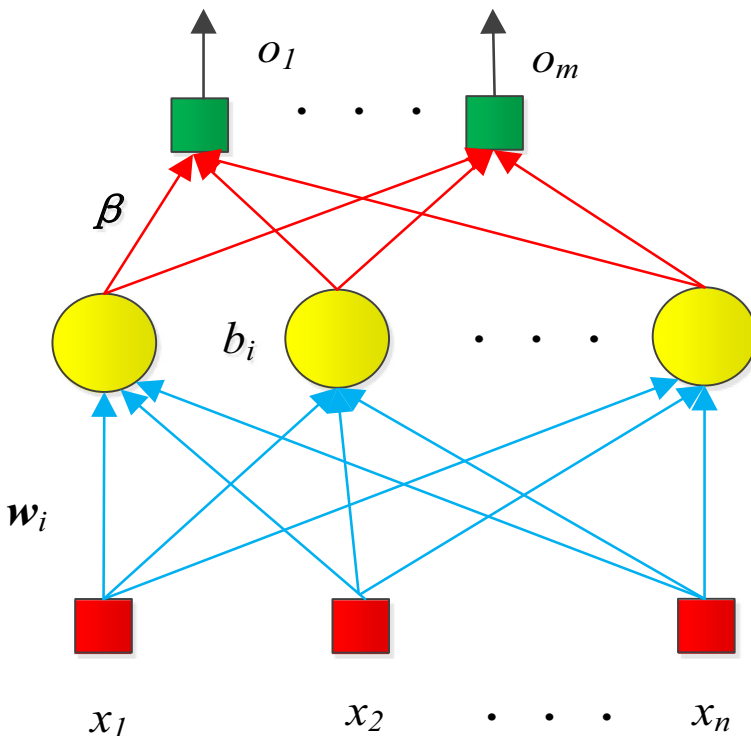$$\text{MSE} = \sum_{i=1}^{N} \left(t_{ij} - o_{ij}\right)^2, j = 1, \ldots, m \tag{2}$$

Fig. 1 Structure of SLFN

Table 1 Notations

Notation	Meaning
x_i	The input vector of the ith sample
t_i	The target vector of the ith sample
$\widehat{N}$	The number of hidden nodes in ELM
w_i	The weight vector from input layer to the ith hidden node
b_i	The bias of the ith hidden node
β_i	The weight vector from the ith hidden node to output layer

where N is the number of training samples, and i and j are the indexes for the training sample and output layer node. It can be proved that SLFN is able to approximate all the training samples when the number of hidden nodes N approaches to infinity,

$$\sum_{j=1}^{N} \left|\left| o_j - t_j \right|\right| = 0 \tag{3}$$

which is called the universal approximation capability, so there must be a set of w_i, b_i and β_i that suffice:

$$\sum_{i=1}^{\widehat{N}} \beta_i g\left(w_i x_j + b_i\right) = t_j, j = 1, \ldots, m \tag{4}$$

The formula above can be abbreviated as

$$H\beta = T \tag{5}$$

where

$$
H\left(w_1, \ldots, w_{\widehat{N}}, b_1, \ldots, b_{\widehat{N}}, x_1, \ldots, x_N\right) =
\begin{bmatrix}
g(w_1 x_1 + b_1) & \cdots & g\left(w_{\widehat{N}} x_1 + b_{\widehat{N}}\right) \\
\vdots & \ddots & \vdots \\
g(w_1 x_N + b_1) & \cdots & g\left(w_{\widehat{N}} x_N + b_{\widehat{N}}\right)
\end{bmatrix}
\tag{6}
$$

Table 2 Activation functions in ELM

Function	Formula
Sigmoid function	$G(a, b, x) = \frac{1}{1 + \exp(-(a \cdot x + b))}$
Hyperbolic tangent function	$G(a, b, x) = \frac{1 - \exp(-(a \cdot x + b))}{1 + \exp(-(a \cdot x + b))}$
Radial basis function	$G(a, b, x) = \exp(-b \| x - a \|)$
Multi-quadratic function	$G(a, b, x) = (\| x - a \| + b^2)^{\frac{1}{2}}$
Hard limit function	$G(a, b, x) = \{\, 1, a \cdot x + b \leq 0 \; 0, \text{otherwise}$
Cosine function	$G(a, b, x) = \cos(a \cdot x + b)$

Table 3 Training of ELM

Training set: $S = [(x_i, y_i) \| x_i \in R^n, y_i \in R^m, i = 1, \ldots, N]$
Initialization:
Assign random values to hidden weight w_i and bias b_i and calculate the output matrix of hidden layer H using training set.
Analytical solution:
Obtain β from $\mathbf{H}\beta = \mathbf{T}$ by Moore Penrose inverse. $\beta = \mathbf{H}^\dagger \mathbf{T}$, where $\mathbf{H}^\dagger$ is the Moore–Penrose generalized inverse of matrix $\mathbf{H}$

$$\beta = \begin{bmatrix} \beta_1^T \\ \vdots \\ \beta_{\widehat{N}}^T \end{bmatrix}_{\widehat{N} \times m}, T = \begin{bmatrix} t_1^T \\ \vdots \\ t_N^T \end{bmatrix}_{N \times m} \tag{7}$$

Therefore, training the SLFN is to find the best w_i, b_i and β_i.

2.2 Principles of ELM

The basic training of ELM can be regarded as two steps: random initialization and linear parameter solution. Firstly, ELM uses random parameters w_i and b_i in its hidden layer, and they are frozen during the whole training process. The input vector is mapped into a random feature space with random settings and nonlinear activation functions which is more efficient than those of trained parameters. With nonlinear piecewise continuous activation functions, ELM has the universal approximation capability [41, 58]. In the second step, β_i can be obtained by Moore-Penrose inverse as it is a linear problem $\mathbf{H}\beta = \mathbf{T}$. So, the training of ELM is summarized in Table 3:

ELM can yield better generalization ability without iteratively tuned hidden parameters. The theory of ELM was proved rigorously by Guang-Bin and Qin-Yu [41], and detailed information can be found in literature [59]. ELM can approximate arbitrary complex classification boundary with sufficient hidden nodes.

3 Variants of ELM

This section presents the recent advanced ELM variants. Various algorithms have been used to improve the performance of ELM for real applications.

3.1 Robustness improvement

Random hidden nodes enable ELM to converge much faster but may also result in fluctuation in classification performance. Calculating the output weight β_i another significant problem that related to ELM performance. Starting from these two points, researchers have been devoted to improving the robustness and stability of ELM.

Huang and Chen [54] proposed incremental ELM (I-ELM) in which the hidden nodes are added incrementally. With nonconstant piecewise continuous activation functions, I-ELM has universal approximation capability. Later, They improved the learning algorithm for I-ELM in

literature [53]. The convergence can be accelerated by convex optimization for recalculating the output weight of existing nodes. Xu and Yao [202] proposed an improved version of I-ELM called incremental regularized extreme learning machine (IR-ELM). The hidden nodes are added in one by one, and the output weight can be recursively updated in an efficient way. In incremental learning scheme, ELM can automatically adjust number of hidden nodes to obtain a better performance for regression and classification than those of random initialization.

Huang and Ding [55] compared ELM and support vector machine (SVM), and found out ELM was equivalent with SVM in terms of classification but ELM was more probably to reach better generalization performance. Liu and Wang [97] found that ELM may suffer from overfitting because it is trained by minimizing error on training set. So, they introduced ensemble learning and cross validation to ELM and proposed ensemble ELM (EN-ELM). In k fold cross validation, the training set was divided into k groups. In every iteration, k ELMs were trained by k-1 groups and validated by the other group. In testing phrase, the prediction was obtained by majority voting. Experimental results suggested that overfitting was alleviated and EN-ELM reached better generalization. Soria-Olivas and Gomez-Sanchis [167] combined Bayesian approach with ELM. By using prior knowledge, the confidence intervals can be calculated efficiently. The proposed Bayesian ELM outperformed classical ELM in six different regression tasks in their experiment. Wang and Cao [184] noticed that ELM sometimes suffers from low effectiveness because the hidden layer output matrix $\mathbf{H}$ is not full column rank. So, they proposed to make selections of the input weight and bias instead of using random parameters. In this way, the singular $\mathbf{H}$ can be avoided which accelerated the computation of generalized inverse. Experiments on benchmark data suggested that the proposed effective ELM (EELM) achieved better classification results with less training time. Cao and Lin [10] employed self-adaptive differential evolution algorithm to train hidden nodes parameters in ELM. With a strategy pool, the potential solution can evolve from previous experience into better ones. In comparison with several other evolutionary based ELM, the proposed self-adaptive differential evolution ELM was better in classification. Zhang and Qu [225] proposed a classification scheme based on evaluation ELM and weighted nearest-neighbor equality algorithm. The significance of features were measured by the weights of ELM input layer for feature selection and weighted nearest-neighbor equality algorithm served as the classifier. Cao and Lin [11] introduced voting mechanism to improve the robustness of ELM. They trained several individual ELMs with the same structure and the final classification result was obtained by voting on the results of these ELMs. The choice of kernel is an important part for kernel ELM. Traditionally, kernel choice is empirical, but it may not always suitable. Liu and Wang [103] proposed multiple kernel extreme learning machines (MK-ELM) for solving this problem. MK-ELM optimizes both the weights of kernel combination and network weights. They also introduced norm constrains and minimum enclosing ball to the learning framework. Three different optimization algorithms were employed in the experiment. The result suggested that the proposed learning framework was comparable to state-of-the-arts with less training time. Deng and Zheng [28] proposed reduced kernel extreme learning machine which aimed to greatly decrease time consumption while not affecting its performance compared to normal kernel extreme learning machine. Their proposed method replaced the original kernel matrix $K(X, X)$ with reduced kernel matrix $K\left(X, \widetilde{X}\right)$, choosing a subset randomly from given dataset [26]. After proving its efficiency, Deng and Zheng [27] also applied this method for cross-person activity recognition task and achieved success.

3.2 Parameter optimization

The pure ELM obtains the output weight by pseudo inverse and random initialization. Its performance and robustness can be further boosted. For regression problems, Chen and Lv [14] put forward a new method to alleviate side effect of outliers on ELM for regression. They proposed a robust regularized ELM using iteratively reweighted least squares (RELM-IRLS). They further compared the generalization performance of different loss functions and regularization terms. Zhang and Yang [226] proposed a hybrid learning algorithm named robust AdaBoost.RT based ensemble ELM (RAE-ELM). In training, RAE-ELM assigned a threshold for each weak learner, and the final output of the system was calculated by weighted ensemble. The proposed method outperformed several state-of-the-arts on real world datasets in their experiment.

Some researchers suggested to use swarm intelligence for optimization of the random initialized weights and biases in ELM. The optimization processes of these methods can be summarized in a flowchart in Fig. 2. Every particle in the population carries a potential solution to the optimization problem, which is updated with the iterations. For ELM optimization, the potential solution is set as the hidden weights and biases, and the fitness function is the training error, which can be expressed as

$$error = \sum_{i=1}^{N} \|\boldsymbol{t}_i - \boldsymbol{o}_i\|^2 \tag{8}$$

where $\boldsymbol{t}_i$ and $\boldsymbol{o}_i$ stand for the ground truth label and the ELM output of the ith training sample.

Zhang and Li [227] proposed a new method called firefly algorithm to optimize the parameters in hidden layer. Firefly algorithm employs a set of fireflies to search the solution space. Initially, every firefly contains a potential solution and they moves with a strategy that inspired by moving of firefly swarm. Zhang and Wu [233] proposed a Memetic Algorithm (MA)-based Extreme Learning Machine (M-ELM), which can reach the global optimal solution for the hidden layer parameters. The memetic algorithm employed a set of individuals to search the solution space. The solutions were set as genes of individuals. By mutation, crossover and selection, the genes in individuals were updated with the generation and finally stopped at the global best. Nayak and Dash [120] employed Improved Whale Optimization Algorithm (IWOA) which is a swarm optimization algorithm inspired from behavior of whales to optimize the parameters of hidden layer in ELM. The object function consisted of root mean squared error and output weight norm. Wu and Yao [194] utilized dolphin swarm optimization to train the parameters in ELM. Wang and Chen [181] trained the input weight and bias of kernel ELM (KELM) by grey wolf optimization (GWO) which belongs to a swarm optimization algorithm that mimics the hunting behavior of wolves. Rigorous comparison was made between KELMs trained by GWO, genetic algorithm, particle swarm optimization, and grid search.

Zhang and Wu [234] proposed instance clone ELM (IC-ELM) for overfitting preventing on small datasets. For each testing sample, IC-ELM finds k training samples that are nearest to the testing sample and clones these weighted ones into training set before training. The proposed IC-ELM outperformed traditional ELM and several other variants.

Zheng and Wang [239] proposed a multi-layer ELM for bearing fault detection. They developed a novel ant lion algorithm (NALO) to optimize the input weight and bias. The NALO can be regarded as an improved form of genetic algorithm with better searching ability. The classification performance of their approach was better than five algorithms.
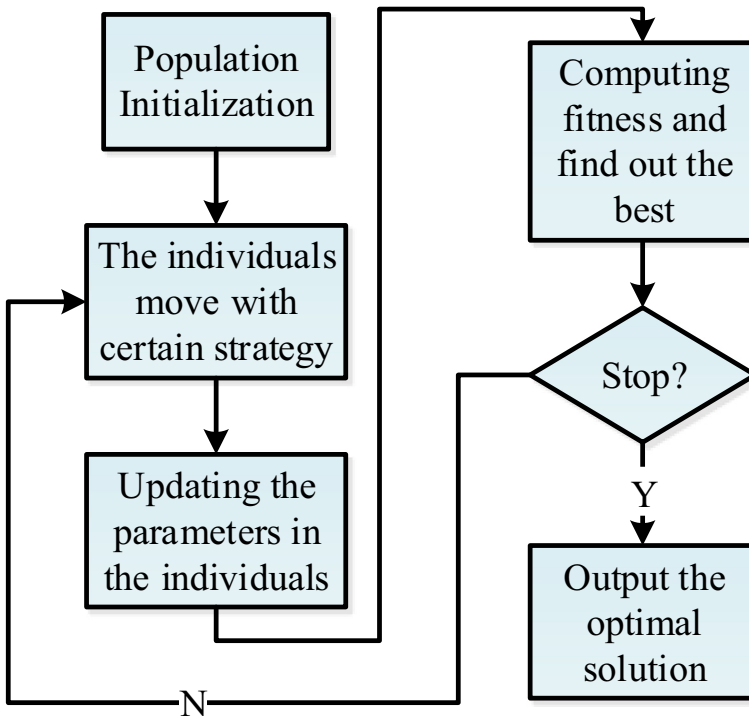
Fig. 2 ELM optimization using swarm intelligence methods

Missing data is another issue related to robustness, which is common in real world applications. Gao and Liu [36] put forward a sample-based ELM (S-ELM) for handling missing data learning. The traditional ELM trains the network in order to obtain the maximum margin and minimum norm of weight. However, it is hard to estimate the distance between the sample with missing data and the separating hype plane because calculating the distance in full feature space is unsuitable. So, they observed this problem and proposed to compute the distance in sample relevant feature space. The proposed S-ELM achieved good classification performance on three different datasets.

3.3 Structure modification

Researchers are also exploring to modify the structure of ELM to improve its performance or for specific tasks. Bai and Kasun [4] added local receptive fields to ELM and proposed ELM-LRF for object recognition. In its frame, the input was the raw image samples. The various local receptive fields were used for feature extraction. Then, a pooling layer with tuning-free nodes was employed to extract comprehensive representation. Finally, the representations were fed into an output layer for classification. The weight in output layer was calculated by least square solution. Experiment on NORB showed that ELM-LRF achieved better classification accuracy than deep belief network (DBN) and conventional neural network (CNN) with much less training time. Deng and Bai [29] combined singular value decomposition (SVD) with ELM for large scale data analysis. The SVD nodes were arranged in hidden layer to capture the innate characteristics of the input and reduce the dimension of data. A fast

approximation method was utilized in SVD to improve computational efficiency. Experimental results demonstrated that the proposed approach was superior to several state-of-the-art algorithms. Ouyang and Cheng [127] proposed a hybrid system based on fuzzy theory which included SVD and ELM for online learning. Instead of conventional Moore Penrose inverse to calculate output weight, they used a recursive SVD based optimization algorithm. Compared to those traditional application of fuzzy system theory [67], their combined approach showed better results. Geng and Dong [38] replaced the hidden layer with Central Neural System (CNS) and proposed Self-Organizing Extreme Learning Machine (SOELM). The training of SOELM consists the following three steps. Firstly, it tunes the input weight using Hebbian learning rule iteratively. Then, mutual information is employed to adjust the structure of the network in a self-organizing manner. Finally, the output weight is obtained by ELM algorithm. The proposed SOELM was applied in predicting complicated chemical processes and yielded better results than ELM. Wu and Qu [193] put forward a novel learning framework for multi-layer ELM. They implemented the hybrid hidden nodes optimization by ant clone algorithm and multiple GWO. Multiple hidden layer output matrixes were leveraged to generate the optimal structure.

3.4 ELM for online learning

In real world applications, there are various kinds of data, so ELM needs to be modified to better learn from these data. For instance, sometimes we may not be able to have access to the whole dataset because the dataset itself is growing. New samples are being added into the dataset from time to time. We need to re-train the ELM every time the set grows. However, the new samples are often accounted for only a small part, so it is inefficient to re-train the network using the whole dataset again and again. To handle this issue, Huang and Liang [56] proposed online sequential ELM (OS-ELM). The basic idea of OS-ELM is to avoid re-training over old samples by using sequential algorithm. After initialization, OS-ELM can adjust parameters over new samples sequentially. In this way, OS-ELM can be trained one by one or block by block. The detailed training process is given below in Table 4:

$$
\begin{aligned}
\beta_0 &= M_0 H_0^T T_0 \\
\text{where } M_0 &= \left(H_0^T H_0\right)^{-1}
\end{aligned}
\tag{9}
$$

$$
\begin{aligned}
\beta_{(t+1)} &= \beta_{(t)} + M_{t+1} h_{t+1}\left(t_i^T - h_{t+1}^T \beta_{(t)}\right) \\
\text{where } M_{t+1} &= M_t - \frac{M_t h_{t+1} h_{t+1}^T M_t}{1 + h_{t+1}^T M_t h_{t+1}}
\end{aligned}
\tag{10}
$$

Later, Zhao and Wang [237] introduced forgetting mechanism to OS-ELM to improve its performance on sequential data with time validity. Time is an important factor in many applications, such as stock market and preferences of people shopping. The outdated training samples should be down weighted and new samples emphasized. Also, Huynh and Won [64] proposed their approach to improve online sequential extreme learning machine. They applied the Tikhonov regularization to relieve the empirical error and tackle with ill-posed problems. The experiments presented its capacity of learning data one-by-one or chunk-by-chunk and

Table 4 Training of OS-ELM

Training set: $S = [(x_i, y_i) \| x_i \in R^n, y_i \in R^m, i = 1, \ldots, N]$ **Initialization:**   Set $t = 0$,
Calculate the output matrix of hidden layer H_0 using training set and random parameters.   Obtain the output weights using equation (9).   **Online sequential learning:**   For each new data block $[x_i, y_i]$, update the output of hidden layer $H_{(t+1)} = [h_t; h_{(t+1)}]$, where $h_{(t+1)}$ denotes the hidden output of the new data block, and tune the $\beta_{(t+1)}$ by equation (10).   Set $t = t + 1$.

superiority of better generalization ability. Moreover, Mirza and Lin [118] proposed their modified method called weighted online sequential extreme learning machine to mitigate the problem of imbalanced class. According to their experimental results, weighted online sequential extreme learning acquired better performance than other methods of class imbalance learning.

3.5 ELM for imbalanced data

Another problem in real application is the imbalanced distribution of samples in training set, which can hinder the performance of ELM severely. In ideal training sets, samples of different classes obey uniform distribution generally, but in real world datasets, the number of samples of some class may be times of that of other classes. Hence, the ELM cannot learn from minority classes effectively. Consequently, the trained network often recognize majority class samples more accurately than minorities. Zong and Huang [246] put forward weighted ELM (W-ELM) for classification of imbalanced data. They suggested to assign weights to each training samples according to class distribution. The formulation of W-ELM can be expressed as

$$\min_{\beta} \frac{1}{2} \|\beta\|^2 + \frac{1}{2} \sum_{i=1}^{N} C_i \|e_i\|^2, \quad s.t. \quad t_i^T - e_i^T = h(x_i)\beta, i = 1 \ldots N \tag{11}$$

where e_i denotes the error of sample x_i, and C_i is the penalty weight hyper-parameter. Practically, weights on majority classes are usually set smaller than those on minority classes. One of the most widely used weighting schemes is:

$$w_i = \frac{1}{n(c_i)} \tag{12}$$

where c_i denotes the class label, and $n(c_i)$ represents the total sample number. W-ELM can achieve good generalization both on imbalanced and balanced datasets and the scheme can be easily applied in cost sensor learning tasks. Its weighting algorithm is also simple and effective. Later, Tang and Chen [174] employed artificial bee colony optimization to train the W-ELM, which further improved its classification performance. Shen and Zhang [158] proposed their scheme for imbalanced binary classification. They used the minority samples as seeds to train an ELM autoencoder (ELM-AE). Then, more samples of minority class can be obtained by the ELM-AE. Those generated sample were added in the dataset so that it become balanced.

3.6 ELM for big data

ELM is now modified for big data. Liu and Huang [98] used ELM to reduce the dimension of high-dimensional data by spectral regression. Then, the output weight can be obtained. The proposed two-stage ELM can converge faster for high dimension data.

Xin and Wang [198] found that ELM on MapReduce was inefficient for updated big data learning and proposed elastic ELM for improvement. The most time-consuming part for big data learning of ELM is the matrix multiplication, so they improved the efficiency by calculating incrementally. Elastic ELM used intermediate matrix from the updated training subset multiplications and obtained the output weight by centralized calculating. Various experiments were conducted for evaluation, and result showed the proposed elastic ELM was rapid and effective for big data learning.

Yang and Xu [204] applied ELM for stream data classification and proposed an ensemble ELM. They embedded a concept drift detection module in ELM, which can recognize both gradual and abrupt concept drift by online sequential learning and classifier updating.

Chen and Song [18] developed a domain adaptation algorithm based on ELM. In many real-world applications, the training data and testing data are from different distributions but are related. So, there is domain transfer in classification. They employed an ELM based feature learning network to transfer the training and testing data into the same feature space and minimized the distance between the two domains. Afterwards, the transferred data can be used to train an adaptive ELM for classification.

3.7 ELM for transfer learning

It is known that datasets with annotation play an important role in many tasks of artificial intelligence. Despite of its importance, we need to realize that collecting labeled data is often a challenging work due to it usually consumes a lot of time and money. Transfer learning has its unique strength in solving this problem. By transferring knowledge from source domain to target domain, transfer learning could be implemented with only a small annotated dataset which makes it quite suitable for fields lack of annotation data such as computer-aided diagnosis, transportation, and emotion recognition. In recent years, extreme learning machine has been more and more applied in the field of transfer learning for its fast-learning speed and easy realization. Then, we will introduce several representative works of applying extreme learning machine in constructing transfer learning scheme.

In 2013, Chen and Wang [19] applied online sequential extreme learning machine in transfer learning scheme for recognizing various transportation modes. Their scheme consisted of three steps. First, they trained an original extreme learning machine on labeled source domain. Second, they calculated the mean and standard deviation for extracting trustable data from target domain. Third, they adopted online sequential extreme learning machine to retrain the original extreme learning machine classifier. According to their experimental results, this approach obtained better performance than other conventional methods.

In 2015, Zhang and Zhang [220] proposed a universal framework for improving the generalization capacity of traditional extreme learning machine. They applied both source domain adaptation method and target domain adaptation method to enhance the cross-domain ability of extreme learning machine and claimed to achieve state-of-the-art experimental results. Later they modified this generalized domain adaptation extreme learning machine

and tested it in the field of electronic nose sensor drift compensation [219] and visual knowledge domain adaptation [221], proving this framework's efficiency and potential.

In 2016, Li and Mao [87] presented an extreme learning machine based method to implement a transfer learning approach for data classification. An extreme learning machine classifier was trained with data from source domain first. During this process, they constructed two different penalty parameters. Second, they applied these two computed penalty parameters to obtain the remaining parameters of derived Lagrangian with data from target domain. In this way, the original extreme learning machine classifier was optimized, and knowledge was transferred from source domain to target domain. Their experimental results demonstrated this approach's efficiency. Further, Li and Mao [88] improved their approach via introducing a free sparse representation from unlabeled data and claimed to achieve promising results as well.

3.8 ELM for Semi-supervised and unsupervised learning

Apart from the supervised learning, ELM has also been applied in semi-supervised and unsupervised learning.

Liu and Xia [92] combined manifold regularization with ELM for semi-supervised learning. They gave a deep analysis on relation between manifold regularization and ELM. They also proposed manifold regularization ELM. While maintaining the features of traditional ELM, manifold regularization ELM can be applied to large scale datasets. Experiment results showed that manifold regularization ELM was superior to traditional semi-supervised learning methods in terms of scalability and generalization. Krishnasamy and Paramesran [76] found that Laplacian regularization for ELM failed in extrapolating ability and proposed a new algorithm for semi-supervised learning called Hessian semi-supervised ELM (HSS-ELM). HSS-ELM used Hessian regularization which preserved the manifold structure and improved the extrapolating ability while maintaining the main features of conventional ELM. The implementation of multiclass learning was also simple.

Liu and Liu [96] studied non-parametric kernel learning (NPKL) and pointed out traditional NPKL methods often require manifold assumption and lack scalability. So they employed ELM for NPKL which can obtain a low rank kernel matrix from the samples without clear manifold structure. Mehrizi and Yazdi [115] combined growing self-organizing map (GSOM) with ELM. The use of ELM solved the parameters calibration of GSOM which substantially boosted the speed and improved the classification performance. Zhang and Ding [224] introduced wavelet analysis to ELM for unsupervised learning and semi-supervised learning.

Ding and Zhang [30] employed ELM as auto encoder for feature learning and sparse representation. The embedded features were used for unsupervised ELM for clustering. Zhu and Miao [244] studied deep learning structures and proposed ELM autoencoder (ELM-AE) for feature representation from unlabeled images. ELM-AE was used to extract low hierarchical representation from local perceptive fields and the learnt features were fed into the last layer called trans-layer for high level feature learning. The block histogram was employed to achieve translation and rotation invariance for representation extraction. Experiments on MNIST and Caltech 101 suggested the effectiveness of their method. Cheng and Liu [20] improved ELM-AE and applied it for K-SVD. They put forward denoising deep ELM-AE (DDELM-AE) by orthonormal initialization of the input weight. Compared with complete random initialization, orthogonalization made better projection in ELM feature space and boosted the performance.

Sun and Zhang [169] added manifold regularization term in the object function of ELM-AE and proposed generalize extreme learning machine autoencoder (GELM-AE). They also developed a new deep network based on GELM-AE which converges faster than conventional deep networks. Nayak and Das [119]used ELM-AE to form a multilayer ELM and employed leaky rectified linear unit (LReLU) as activation function, which can be written as

$$LReLU(x) = \begin{cases} x, & if \ x > 0 \\ ax, & otherwise \end{cases} \tag{13}$$

Where $a \in (0, 1)$ is a hyper-parameter. Yousefi-Azar and McDonnell [209] proposed a hybrid approach for semi-supervised learning. They employed convolution and pooling to extract feature from input images and used ELM as the classifier. She and Hu [157] proposed to use hierarchical ELM to learn features from EEG and employ semi-supervised ELM for classification. Xia and Wang [196] studied OS-ELM and developed density-based semi-supervised OS-ELM which can learn patterns from unlabeled samples. A summary of ELM variants is presented in Table 5.

4 ELM vs other algorithms

Researchers have been devoted to analysis of ELM over the last decade, especially the comparison between ELM and other state-of-the art algorithms including Support Vector Machine (SVM), Random Neural Network (RNN), Radial Basis Function Neural Network (RBFNN), Hopfield Neural Network (HNN), Boltzmann Machine (BM), Restricted Boltzmann Machine (RBM), Deep Belief Network (DBN), and other deep learning methods. As representative machine learning algorithms, these classic methods stand for state-of-the-art technologies of machine learning in different age, not only playing an important role in the history of development of machine learning but also being active in research field nowadays. Via comparison between ELM and these machine learning algorithms, we could more easily to investigate the meaning of ELM in machine learning and more clearly to observe ELM's strength in related research. Generally speaking, ELM is obviously faster than those methods, and its variants are becoming more and more robust. With simple implementation, the performances of ELM are good in terms of accuracy.

4.1 ELM vs support vector machine (SVM)

As one of most well-known supervised algorithms for classification since last century, Support Vector Machine (SVM) was proposed by Cortes and Vapnik [24] in 1995. Basically speaking, considering data points scattering in space, we could find plenty of planes to separate these data points into several classes. In SVM, these planes are called hyperplanes. And among these hyperplanes, there exists one which would maximize the distance of every data point to the plane. That means this hyperplane could distinguish data points best. The mission of SVM is to find this best hyperplane, and it constructs essential idea of SVM.

More specifically, given data sample D

$$D = \left\{ (X_1, y_1), (X_2, y_2), \ldots, (X_n, y_n) \right\} \tag{14}$$

Table 5 Summary of variants of ELM

Authors	Contribution	Results
Huang and Chen [54]	I-ELM was proposed, in which the hidden nodes can be added into the network incrementally. The weights for new node can be obtained by $$\beta_{new} = \frac{EH_{new}^{T}}{H_{new}H_{new}^{T}}$$ where H_{new} stands for hidden output matrix with the new node, and E denotes the residual error.	I-ELM requires less user intervene in learning.
Huang and Chen [53]	The learning speed of I-ELM was accelerated by convex optimization algorithm.	Faster convergence of I-ELM.
Xu and Yao [202]	IR-ELM was proposed which is a variant of I-ELM. IR-ELM can adjust the random nodes automatically.	IR-ELM performed better in classification and regression than I-ELM.
Huang and Ding [55]	The comparison of ELM and SVM was carried out.	ELM tends to outperform SVM in classification because of less optimization constrains.
Liu and Wang [97]	EN-ELM was invented to improve generalization.	The overfitting was alleviated in ensemble ELM.
Soria-Olivas and Gomez-Sanchis [167]	Bayesian method was introduced to ELM to provide prior knowledge.	Generalization was improved for Bayesian ELM.
Wang and Cao [184]	By adjusting the input weights, the output matrix of hidden layer can be made full column rank which improved the effectiveness of ELM.	Effective ELM achieved better classification performance than original ELM at faster speed.
Cao and Lin [10]	Self-adaptive differential evolution algorithm was used to train ELM with mutation and crossover.	Their method achieved better classification result than several other evolutionary based ELM.
Zhang and Qu [225]	Weighted nearest-neighbor equality algorithm was applied to evaluate the significance of input features which is related to weights in evaluation ELM.	The approach yielded several state-of-the-arts in classification problems.
Cao and Lin [11]	Voting method was utilized to generate classification results based on a set of ELMs trained by subsets of the training set.	The robustness was improved in voting ELM.
Liu and Wang [103]	Multiple kernel learning framework was proposed for ELM.	The proposed learning scheme reached state-of-the-art performance.
Chen and Lv [14]	RELM-IRLS was put forward to deal with the outliers in regression problems. L1-norm, L2-norm, Huber, Bisquare, and Welsch were used for robust loss functions.	Comparison with models of different loss functions and regularization terms suggested the RELM-IRLS was better for regression.
Zhang and Yang [226]	Robust AdaBoost.RT was employed for ensemble ELM.	RAE-ELM outperformed several state-of-the-arts on real world datasets.
Zhang and Li [227]	Firefly algorithm was utilized to optimize the parameters in ELM. The hidden weights and biases were updated by the moving of fireflies.	ELM optimized by firefly algorithm performed better than similar methods with less hidden neurons.
Zhang and Wu [233]	Memetic Algorithm was used to train ELM to obtain global best solution of hidden weights and biases which worked by iterations of mutation, crossover, and selection.	Simulation on 46 UCI data sets revealed that the M-ELM performed obviously better than other state-of-the-art ELMs.

Table 5 (continued)

Authors	Contribution	Results
Nayak and Dash [120]	Improved Whale Optimization Algorithm was employed to train ELM.	IWOA-ELM outperformed other machine learning algorithms in pathological brain detection.
Wu and Yao [194]	Dolphin swarm optimization was combined with ELM.	DSO-ELM achieved better generalization performance than other algorithms in 13 benchmark datasets.
Wang and Chen [181]	Grey wolf optimization was used to train the input weights and biases in ELM.	GWO-ELM was better in classification compared with ELM trained by genetic algorithm, particle swarm optimization, and grid search.
Zhang and Wu [234]	IC-ELM was proposed to improve generalization of ELM on small datasets.	IC-ELM outperformed classical ELM and several variants in experiment.
Zheng and Wang [239]	NALO was applied in multi-layer ELM for optimizing input weight and bias.	Multi-layer ELM-NALO performed better than five other algorithms in bearing fault detection.
Gao and Liu [36]	S-ELM was invented to handle missing data.	The proposed S-ELM achieved good accuracy on three datasets.
Bai and Kasun [4]	Local receptive fields and pooling operation was added to ELM.	The proposed ELM-LRF yielded higher classification accuracy than DBN and CNN in much less training time.
Deng and Bai [29]	SVD was introduced to replace the hidden neurons in ELM.	Experimental results demonstrated that the proposed approach was superior to several state-of-the-art algorithms.
Ouyang and Cheng [127]	SVD, fuzzy theory and ELM was combined for online learning.	The proposed algorithm achieved better classification accuracy and regression result than several ELM variants.
Geng and Dong [38]	Central Neural System was employed to replace hidden layer in ELM.	The proposed SOELM was applied in predicting complicated chemical processes and yielded better results than ELM.
Wu and Qu [193]	A novel learning framework for multi-layer ELM was proposed based on optimization methods, subnetwork hidden nodes and cost sensitive learning.	The proposed method yielded better performance on regression and classification tasks than state-of-the-art ELMs.
Huang and Liang [56]	OS-ELM was put forward to solve online learning tasks.	OS-ELM produced better accuracy on real benchmark regression problems.
Zhao and Wang [237]	Forgetting mechanism was introduced to OS-ELM to improve its performance on sequential data with time validity.	The proposed ELM variant has better stability and performance for short-term prediction.
Zong and Huang [246]	Weighted ELM was proposed to handle class imbalanced data.	Experiments over benchmark datasets suggested W-ELM achieved better generalization for imbalanced data.
Tang and Chen [174]	Artificial bee colony optimization was employed to train W-ELM.	Results on 15 imbalanced datasets revealed the superiority of the optimized W-ELM.
Shen and Zhang [158]	ELM-AE was used to generate more samples of minority class to solve binary class imbalanced problem.	The proposed algorithm is simple and effective for binary classification.
Liu and Huang [98]		

Table 5 (continued)

Authors	Contribution	Results
	Spectral regression was employed to reduce dimension in ELM hidden layer.	The variant yielded better performance in terms of accuracy and training time than ELM for high dimension data.
Xin and Wang [198]	Elastic ELM was proposed for big data learning.	The proposed elastic ELM was rapid and effective
Yang, Xu, et al. (2018) [38]	Concept drift detection module was embedded in ELM for stream data analysis.	The ensemble ELM can detect concept drift accurately.
Chen and Song [18]	Domain adaptive ELM was invented for transfer learning.	The proposed model can transfer the training and testing samples to the same feature space, which improved the classification accuracy.
Chen and Wang [19]	Employ the mean and standard deviation to transfer knowledge from source domain to target domain	Better performance than conventional methods
Zhang and Zhang [220]	Construct a framework which elevate generalization ability by applying both source domain adaptation and target domain adaptation algorithms	State-of-the-art performance at that time
Zhang and Zhang [219]	Introduce transfer extreme learning machine to e-nose system	Efficient in e-nose sensor drift compensation
Zhang and Zhang [221]	Introduce transfer extreme learning machine to visual system	Efficient in visual knowledge domain adaptation
Li and Mao [87]	Apply penalty parameters to optimize the remaining parameters for retraining extreme learning machine classifier	Promising performance in data classification
Li and Mao [88]	Adopt free sparse representation for unlabeled data	Potential experimental results
Liu and Xia [92]	Manifold regularization ELM was proposed for semi-supervised learning.	Experiment results showed that manifold regularization ELM was superior to traditional semi-supervised learning methods in terms of scalability and generalization.
Krishnasamy and Paramesran [76]	HSS-ELM was proposed to preserve the manifold structure and improve extrapolating ability	Higher accuracy was achieved by HSS-ELM compared with SVM and least square methods.
Liu and Liu [96]	ELM was applied to NPKL to obtain a low rank kernel matrix from the samples without clear manifold structure.	The scalability of the algorithm was improved.
Mehrizi and Yazdi [115]	GSOM was combined with ELM.	The hybrid system converged faster with higher accuracy.
Zhang and Ding [224]	Wavelet analysis was introduced to ELM for unsupervised and semi-supervised learning.	ELM with wavelet kernel function achieved good performance for clustering and classification.
Ding and Zhang [30]	ELM served as an autoencoder for feature extraction and representation learning.	The proposed approach achieved comparable performance within less time.
Zhu and Miao [244]	Deep learning structure was employed to improve the feature learning of ELM.	Compared with conventional deep models, ELM-AE required less training time.
Cheng and Liu [20]	DDELM-AE was put forward for better feature projection.	DDELM-AE got better feature mapping than ELM-AE and the performance was boosted.
Sun and Zhang [169]		

Table 5 (continued)

Authors	Contribution	Results
	Manifold regularization term was introduced to the object function of ELM-AE	GELM-AE outperformed several state-of-the-arts in terms of accuracy and running time.
Nayak and Das [119]	Leaky rectified linear unit was used in multilayer ELM.	The robustness and generalization ability of the proposed system was improved.
Yousefi-Azar and McDonnell [209]	A hybrid system with convolution, clustering and ELM was proposed.	The proposed system yielded good classification accuracy with boosted speed.
She and Hu [157]	A semi-supervised algorithm was proposed for EEG classification based on ELM.	Their method achieved better performance than several state-of-the-art approaches.
Xia and Wang [196]	OS-ELM was modified to handle unlabeled data.	The proposed scheme was able to handle complex data.

where X_i is a vector consists of d elements which represents a coordinate in space of d dimension. And y_i is the value for classification and normally there exists $y \in \{1, -1\}$ for representation of positive class and negative one. In this space, we could always find a hyperplane which could be recorded as

$$X^T w + \mathrm{b} = 0 \tag{15}$$

It is easy to understand that we could always find a pair of hyperplanes which could separate D into positive and negative groups. And this pair of hyperplanes could be recorded as

$$\begin{cases} X^T w + b = +1 \\ X^T w + b = -1 \end{cases} \tag{16}$$

Thus, we could obtain the distance between this pair of hyperplanes and the distance in SVM is called margin, which could be written as

$$margin = \frac{2}{\|w\|} \tag{17}$$

Margin is demonstrated in Fig. 3.

Due to principle of SVM, support vector is defined as point on the pair of hyperplanes, which could be recorded as

$$y_i\left(X^T w + b\right) = 1 \tag{18}$$

And the purpose of SVM is to maximize the margin, which is equal to this equation as below

$$\max \frac{2}{\|w\|} \Leftrightarrow \min \frac{1}{2}\|w\|^2 \tag{19}$$

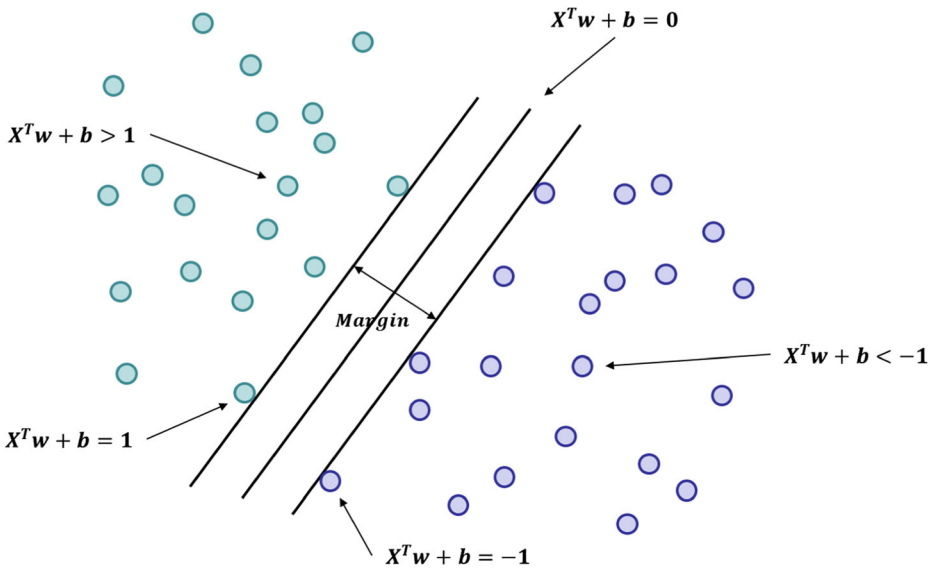So, concluding the restriction condition, the mathematical expression of SVM's goal is written as below

Fig. 3 Illustration of hyperplanes and margins

$$\min \frac{1}{2}\|w\|^2$$
$$s.t.y_i\left(X^T w + b\right) \geq 1, i = 1, 2, \ldots, n \tag{20}$$

However, we usually do not calculate the equation above directly since it is hard to figure out. Commonly, we call this equation a primal problem and we could apply Lagrange Function to solve its dual problem which is much easier to be calculated. In this situation, we introduced Lagrange Multiplier

$$\alpha_i \geq 0, i = 1, 2, \ldots, n \tag{21}$$

into primal equation and acquired its Lagrange Function and got the formula as below:

$$L(w, b, \alpha) = \frac{1}{2}\|w\|^2 - \sum_{i=1}^{n} \alpha_i \left[y_i\left(X^T w + b\right) - 1\right] \tag{22}$$

We could see it that only when all restricted conditions are required, $L(w, b, \alpha)$ could reach optimized value which could be shown as

$$L(w, b, \alpha) = \frac{1}{2}\|w\|^2 \tag{23}$$

So, the optimization of margin is equal to optimization of its Lagrange Function, when restricted conditions are all required. That is how we provided the basic principle of SVM method.

Huang and Ding [55] presented a detailed analysis and comparison of ELM and SVM. It is revealed that the maximal margin theory of SVM and the minimal weight norm of network are in conformity in ELM. In classification problems, ELM and SVM are equivalent, but ELM has less optimization constrains so it tends to yield better performance. Moreover, ELM is robust to user defined factors. Later, Huang and Zhou [57] compared ELM with least square SVM

(LS-SVM) and proximal SVM (PSVM) on wide types of datasets. They found that ELM can be regarded as a unified form of LS-SVM and PSVM. The performance of ELM for binary class problems was similar to those of LS-SVM and PSVM but for multiclass problem, ELM outperformed the two substantially. Furthermore, the convergence speed of ELM is thousands of times faster due to the random feature mapping. Random feature mapping not only provides nonlinearity but also enables the global approximation ability for ELM.

Liu and Loh [106] compared SVM with ELM in the field of text classification in 2005. Their experiments demonstrated that SVM showed better performance than ELM in most evaluation categories especially F1 value. In 2014, Chorowski and Wang [21] presented their research of implementing framework of convex optimization to assess different machine learning models including ELM and SVM. Zhong and Miao [240] compared several machine learning methods for corporate credit ratings. According to their experiments, SVM had the highest accuracy while single hidden layer feedforward networks showed best robustness and ELM proved its strength in efficiency, simplicity, and fast speed. In 2015, Qiu and Wang [140] investigated various methods for prediction of quality status of mandarin. They found that ELM and Random forest showed better prediction performance than SVM in mandarin quality prediction. In 2016, Zhang and Zhang [222] performed experiments to search best classifier for deep features extracted by convolutional neural network. They claimed that ELM showed obvious superiority to SVM in cross domain recognition tasks and kernel extreme learning machine achieved state-of-the-art performance. In 2017, Olatunji [125] compared ELM with SVM for email classification. According to their experimental results, SVM had better accuracy while ELM had faster speed.

4.2 ELM vs random neural network (RNN)

Random Neural Network (RNN) is a special kind of neural network which has its unique structure and learning strategy. It was proposed by Erol Gelenbe [37] in 1989, relatively late compared to birth year of other neural network model such as Support Vector Machine (SVM) or Boltzmann Machine (BM). Random Neural Network was inspired by simulation of the process that how biological neural network received and emitted stimulation signal. In actual fact, this process was determined by neuron's own electric potential. It was the same with the process occurring in RNN. RNN was an open neural network which consisted of several neurons. Each neuron's electric potential was recorded by Random Neural Network. And if the neuron received a positive signal, its electric potential would plus one. In turn, if the neuron received a negative signal, its electric potential would minus one. The process of emitting and receiving signal would occur continuously within RNN until the whole network network reached a balance condition.

Bhat and Merchant [5] proposed their approach on predicting melting points of organic compounds based on extreme learning machine. In their research, they also compared ELM based approach with other machine learning methods including Random Neural Network and found that ELM was superior according to their experiments. Feng and Huang [34] combined extreme learning machine with random structure of hidden neurons and improved the model performance in their research. Wang and Cao [184] addressed in their paper that according to their research results, extreme learning machine had several advantages in some conditions including higher accuracy and lower training cost against some other neural network structure, Random Neural Network included. In 2014, Zhu and Miao [245] claimed they established a novel neural network structure using constrained extreme learning machine. They adopted

random weights and connections derived from RNN and achieved outstanding performance in that year. Lima and Cannon [91] evaluated several representative neural network models in environmental sciences including extreme learning machine and RNN in 2015. According to their inspection, extreme learning machine were more widely applied in environmental sciences than RNN. Minemoto and Isokawa [117] performed research on optimization of feed forward neural network but in their research, they also acknowledged that extreme learning machine was a more powerful approach in most cases compared to Random Neural Network.

4.3 ELM vs radial basis function neural network (RBFNN)

RBFNN is a single hidden layer feedforward neural network based on function approximation proposed by Broomhead and Lowe [7] in the late 1980s. With the increasing maturity of the research, RBFNN has attracted great attention from researchers in various fields due to its simple structure, strong nonlinear approximation ability and good generalization ability, and has been widely used in many research fields such as pattern classification, function approximation and data mining [68].

The basic idea of RBFNN is to apply Radial Basis Function (RBF) as the "basis" of the hidden element to form a hidden layer space. The hidden layer transforms the input vector, transforms the input data of the low-dimensional mode into the high-dimensional space, and makes the linear inseparability problem in the low-dimensional space linearly separable in the high-dimensional space. One detail is to use the "basis" of the hidden element of RBF to form the hidden layer space, so that input vectors can be mapped directly (without weight connection) to the hidden space. When the center point of RBF is determined, the mapping relationship is also determined. The mapping from the hidden layer space to the output space is linear (note the distinction between linear mapping and nonlinear mapping here), that is, the network output is the linear weighted sum of the output of the element, and the weight here is the network adjustable parameter.

As a three-layer forward network with a single hidden layer, RBFNN's structure is shown in Fig. 4. The first layer is the input layer and consists of signal source nodes. The second layer to hidden layer, hidden layer node number depending on the described problems need, transformation function of neurons in hidden layer radial basis function is the center of radial symmetry and attenuation of nonnegative linear function, the function is local response function, local response of the concrete is embodied in its visible to hidden layer of transformation is different from other networks. The previous forward network transformation functions are all functions of global response. The third layer, the output layer, is the response to the input pattern. The input layer only plays the role of signal transmission, and the connection between the input layer and the hidden layer can be regarded as a connection with the weight value of 1. The tasks completed by the output layer and the hidden layer are different, so their learning strategies are also different. The output layer adjusts the linear weight and adopts the linear optimization strategy, so the learning speed is faster. However, the hidden layer adjusts the parameters of the activation function (Green's function, Gaussian function and generally the latter), and adopts the nonlinear optimization strategy, so the learning speed is slow. The understanding of this sentence can be found from the transformation between layers below.

The advantages of radial basis function neural network are as follows: the approximation ability, classification ability and learning speed are superior to BP neural network, the structure is simple, the training is simple, the learning convergence speed is fast, it can approximate any nonlinear function, and it can overcome the local minimum problem. The reason is that its
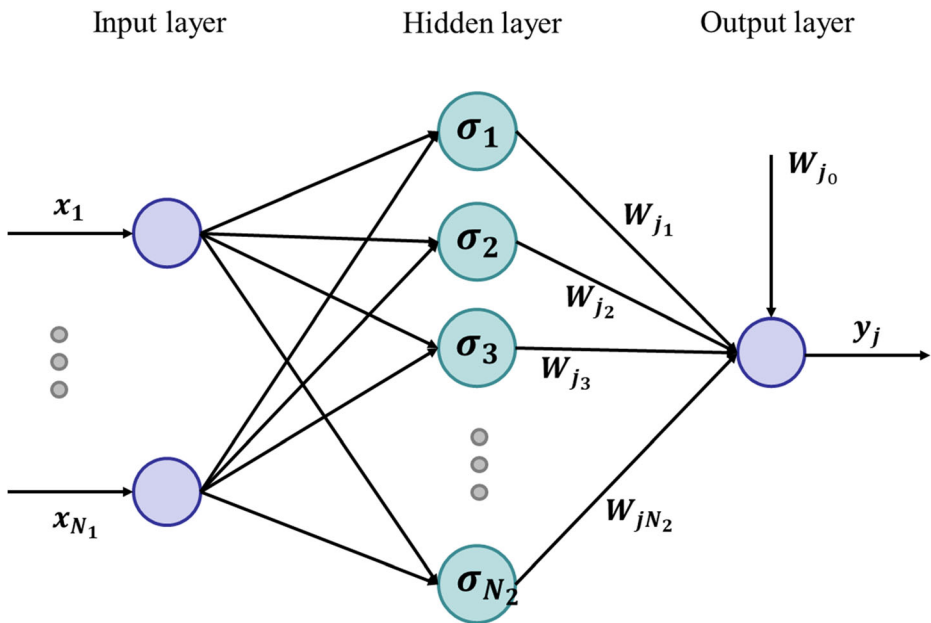
Fig. 4 Illustration of RBFNN

parameter initialization has certain method, not random initialization. For the RBF neural network training process, the key problem is the reasonable determination of the central parameters of the hidden layer neurons. The common method is to directly select the central parameter (or its initial value) from a given training sample set in a certain way, or to determine it by clustering method.

Researchers have been trying comparing Radial Basis Function Neural Network (RBFNN) with extreme learning machine (ELM) when solving some specific machine learning tasks for a long history. Li and Huang [84] inspected the performance comparison of ELM and RBFNN in the field of channel equalization in 2006. According to their study, ELM based methods obtained better results in symbol error rate and learning speed which meant ELM won the competition against RBFNN in the field of channel equalization application. Besides comparison of these two models, scientists also did experiments on improving performance by uniting advantages of both ELM and RBFNN. In 2011, Fernandez-Navarro and Hervas-Martinez [35] proposed a novel model which applied extreme learning machine as training method and used generalized radial basis function (GRBF) neural network as basis. Their merged new approach was called modified extreme learning machine-generalized radial basis (MELM-GRBF) neural network. They claimed that the proposed method was superior to either common extreme learning machine or radial basis function neural network. Ding and Ma [31] also considered applying extreme learning machine to train radial basis function neural network. They proposed a self-adaptive training strategy based on affinity propagation and demonstrated that their methodology showed obvious advantages beyond other models. Liu and Wan [102] analyzed the universal consistency of extreme learning machine when training radial basis function neural networks and which kernel function should be chosen in extreme learning machine according to different conditions. Lam and Wunsch [77] studied various unsupervised learning approaches for classification in the field of compute unified device architecture

(CUDA) including extreme learning machine (ELM), support vector machine (SVM) and radial basis function (RBF) neural network. Wen and Fan [190] proposed a classifier called hybrid structure-adaptive radial basis function-extreme learning machine (HSARBF-ELM) in 2017 and claimed their method combined advantages of both structure-adaptive radial basis function (SARBF) network and extreme learning machine (ELM). Yan [203] adopted both ELM and RBF to achieve higher accuracy of prediction in their experiments. And research of similar direction was concerned constantly in recent years [65, 201].

4.4 ELM vs hopfield neural network (HNN)

In 1982, US physicist Hopfield [50] proposed a novel neural network aimed at solving pattern recognition and optimization problem, named as Hopfield Neural Network (HNN). Since Hopfield Network came into people's view, scientists developed a series of creative structures of neural network by his inspiration. Hopfield Neural Network is a kind of recursive neural network with feedback connections from output to input. Each neuron is interconnected with all other neurons, also known as full interconnection network. Hopfield Neural Network is a neural network combining storage system and binary system. It guarantees convergence to a local minimum, but the condition that converging to the wrong local minimum instead of a global minimum can also happen. Hopfield Network also provides a model to simulate human memory. Hopfield Neural Network is a feedback neural network, and its output will be fed back to its input. Under the excitation of input, its output will produce constant state changes, and this feedback process will be repeated all the time. If Hopfield neural network is a convergent and stable network, the changes generated by this feedback and iterative calculation process will become smaller and smaller. Once a stable equilibrium state is reached, the Hopfield neural network will output a stable constant value. For Hopfield Neural Network, the key is to determine its weight coefficient under stable conditions.

Generally speaking, Hopfield Neural Network could be divided into two types, Discrete Hopfield Neural Network (DHNN) and Continuous Hopfield Neural Network (CHNN). The structure of DHNN is drawn as below in Fig. 5. There are few articles directly comparing Hopfield Neural Network (HNN) with extreme learning machine (ELM) due to scientists have developed more advanced neural network models based on HNN, such as Restricted Boltzmann Machine and Deep Belief Network (DBN) which are more widely applied since they faced the world. Although in this condition, we still reviewed some papers concerning performance comparison between HNN and ELM. Liang and Cheng [90] proposed their modified model based on ELM and claimed it was superior to some other popular neural networks including Hopfield neural network. Liu and Liu [95] constructed an improved multiple hidden layers ELM by employing random search enhancement. According to their experiments, their improved ELM based method had advantages in training speed and accuracy compared to other methods including HNN.

4.5 ELM vs boltzmann machine (BM)

As we discussed above, Hopfield Neural Network has the function of optimal calculation. However, the network can only evolve in strict accordance with the descending mode of energy function, and it is difficult to avoid the appearance of false state. Moreover, the weight is easy to fall into the local minimum value and cannot converge to the global optimal solution. If the iterative process of the feedback neural network is not so rigid, it can temporarily accept
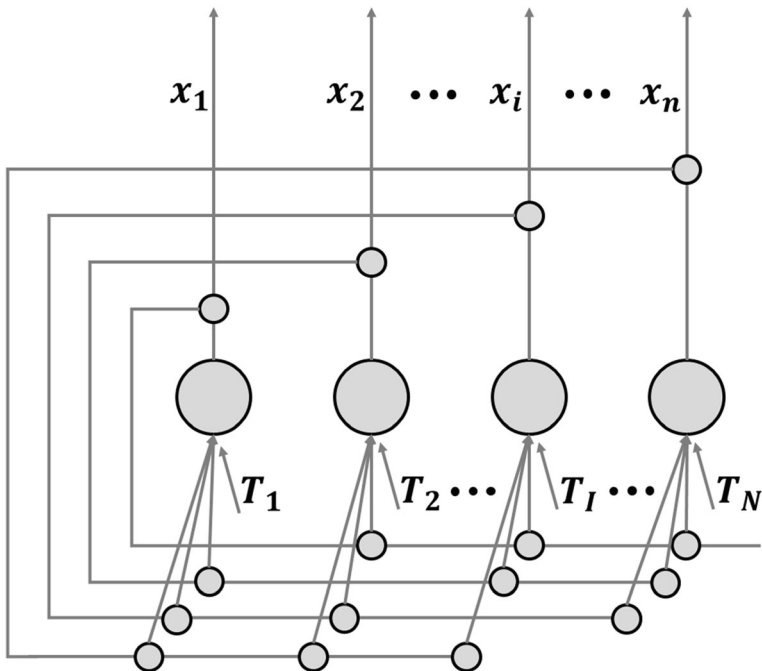
Fig. 5 Illustration of DHNN

the result of increasing the energy function to some extent, then it is possible to jump out of the local minimum. The core idea of stochastic neural network is to add probability factors into the network. The network does not evolve in the direction of the energy function reduction but evolves in the direction of a large probability to ensure the correct iteration direction. Meanwhile, the probability of the energy function increase also exists to prevent the network from falling into the local minimum.

Gradient descent is the most used method in machine learning and optimization combination problems. For example, BP neural network, the more units of the multilayer perceptron, the larger the corresponding weight matrix, and each weight can be regarded as one degree of freedom or variable. We know that the higher the degree of freedom, the more variables, the more complex the model, the more powerful the model. However, the stronger the model capability is, the easier it is for the model to overfit and be over-sensitive to noise. On the other hand, when using gradient descent to search for the optimal solution, the error surface of multiple variables is very similar to the undulating mountain peak. The more variables there are, the more mountains and valleys there are, which makes it very easy for the gradient descent method to fall into a small local valley and stop searching. This is the most common local optimal problem in solving multi-dimensional optimization problems by conventional gradient descent method. The reason is the search criterion of the gradient descent method, which is based on the negative direction of the gradient, blindly pursuing the reduction of network error or energy function, so that the search only has the ability to "go down the mountain", but not to "go up the mountain". The so-called "ability to climb" is that when the search falls into the local optimum, it can also have the ability to "climb over the mountains" to escape from the local optimum and continue to search for the global optimum. If you want to make a graphic metaphor for a system with multiple local minima. Imagine that there is a

concave and convex surface of multi-dimensional energy on the tray. If a ball is placed on the surface, it will roll into the nearest trough (local minimum point) under the action of gravity and cannot break away. But the trough is not necessarily the lowest trough on the surface (global minimum point). Therefore, the local minimum problem can only be solved by improving the algorithm. One way to do this is to give the algorithm the ability to "climb", as mentioned earlier, while ensuring that the search does not "climb" out of the globally optimal "valley".

Simulated annealing algorithm is an effective method to solve the problem of local minimum energy in stochastic network. Its basic idea is to simulate the process of metal annealing. The process of metal annealing is roughly as follows: firstly, the object is heated to a high temperature so that its atoms are in a high-speed motion state, and then the object has a high internal energy. Then, the temperature drops slowly. As the temperature drops, the atoms slow down and their internal energy drops. And then finally, the whole thing reaches its lowest internal energy state. As one of most significant derivatives of simulated annealing algorithm, Boltzmann machine was firstly proposed by Ackley and Hinton [1] in 1985. As it is shown in Fig. 6, a typical Boltzmann machine consists of hidden units, visible units, and edges of dependency. Boltzmann machine was the backbone of many later more well-known and applicative neural network models such as Restricted Boltzmann Machine (RBM) and Deep Boltzmann Machine (DBM). Due to these improved neural network models are more widely applied in research field, we will not provide separate paragraph on literature of comparing Boltzmann Machine with extreme learning machine but will review representative papers of comparing extreme learning machine with RBM or DBM in next section.
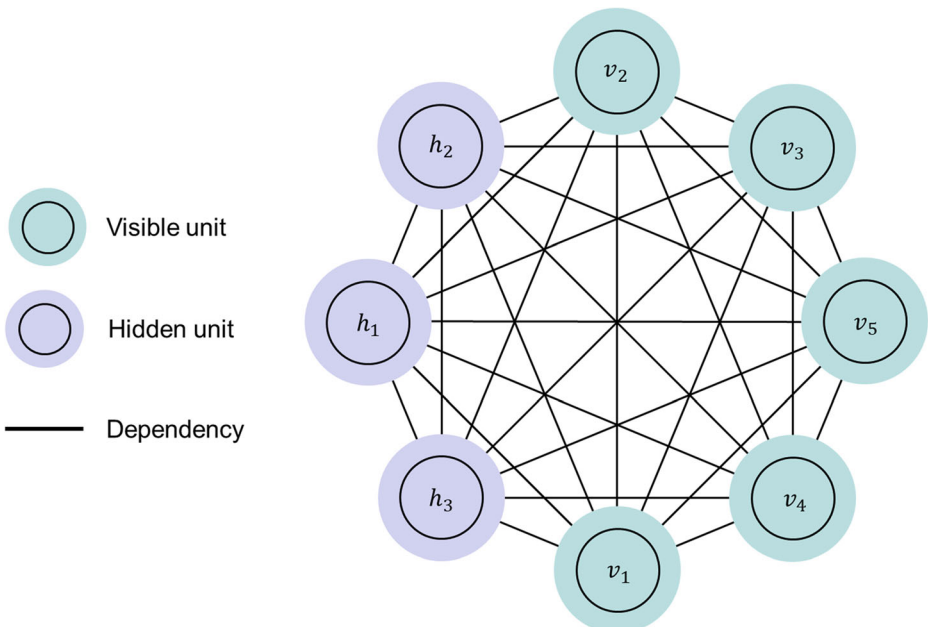

Fig. 6 Network structure of Boltzmann Machine

4.6 ELM vs restricted boltzmann machine (RBM)

Restricted Boltzmann Machine (RBM) is a randomly generated neural network that can learn probability distribution through input data sets. RBM was originally named Harmonium by its inventor, Paul Skitsky, in 1986, but the limited Boltzmann machine did not become well known until Hinton and Salakhutdinov [49] developed a fast-learning algorithm in the mid-2000s. Restricted Boltzmann machines have been used in dimension reduction, classification, collaborative filtering, feature learning and topic modeling. Depending on the task, the restricted Boltzmann machine can be trained using supervised learning or unsupervised learning.

Restricted Boltzmann Machine (which structure is shown is Fig. 7) is a special topological structure of Boltzmann machine (BM). The principle of BM originates from statistical physics, and it is a modeling method based on energy function, which can describe the higher-order interaction between variables. The learning algorithm of BM is more complex, but the established model and learning algorithm are based on relatively complete physical explanation and strict mathematical statistics theory. BM is a symmetrically coupled random feedback binary cell neural network, which is composed of visible layer and multiple hidden layers. Network nodes are divided into visible unit and hidden unit. Visible unit and hidden unit are used to express the learning model of random network and random environment, and the correlation between units is expressed through weights.

Once RBM learns the structure of the input data associated with the activation values of the first hidden layer, the data is passed down the network one layer. The first hidden layer
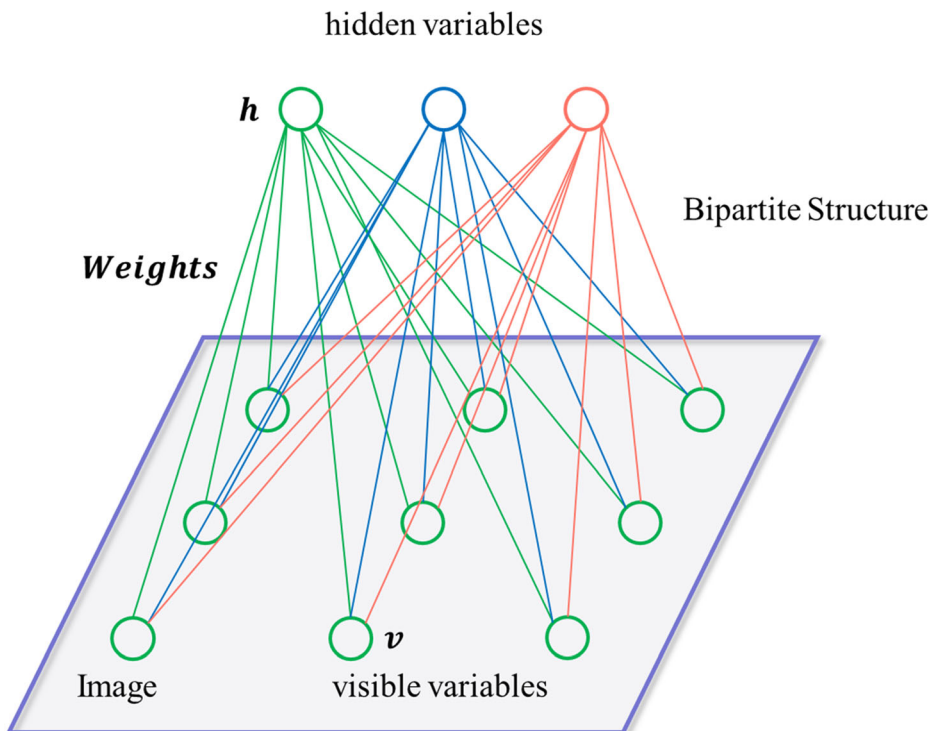

Fig. 7 Illustration of Restricted Boltzmann Machine (RBM)

becomes the new visible or input layer. The activation value of this layer is multiplied by the weight of the second hidden layer to produce another set of activations. This process of creating a sequence of activation values through feature grouping and grouping feature groups is the basis of feature hierarchy, through which neural networks learn more complex and abstract representations of data. For each new hidden layer, the weight is iteratively adjusted until the layer approximates the input from the previous layer. It is greedy, hierarchical, unsupervised pre-training. It does not require the use of tags to improve the weight of the network, which means that we can train on unlabeled data sets that have not been manually processed, which is the vast majority of data in reality. Often, algorithms with more data produce more accurate results, which is one reason for the rise of deep learning algorithms. Since these weights are already close to the characteristics of the data, the subsequent supervised learning stage can be learned more easily when deep belief networks are used for image classification. Although RBM has many application use, proper weight initialization to facilitate later classification is one of its main advantages. In a way, they perform something like backpropagation: they do a good job of adjusting the weights to better model the data. We could say that pre-training and back propagation are alternative ways to achieve the same goal.

As one of most famous and commonly used neural network models, Restricted Boltzmann Machine (RBM) has been hot in research field since it came out. Bu and Zhao [8] presented a methodology which applied RBM to enhance the potential capacity of extreme learning machine (ELM) in the field of spectral processing. Through their research, it was demonstrated that their joint architecture of RBM and ELM showed better performance than the algorithm of applying principal component analysis (PCA) with ELM learning method. Hassan [45] exploited various classification methods in the field of automated sleep apnea detection, including naïve bayes, k nearest neighbor (KNN), AdaBoost, random forest, RBM and ELM. According to their conclusion, ELM was superior to other methods (including RBM) in building automated sleep apnea detection system. In 2016, He and Wang [47] proposed a novel approach which used Gaussian Restricted Boltzmann Machine (Gaussian RBM) as basis and claimed it beat other popular methods including ELM, support vector machine (SVM) and deep belief network (DBN) in diagnosis of bearing fault due to it was more robust, accurate and effective. Chen and Yang [15] presented a method based on ELM and RBM to handle big data problem. Their method applied RBM by inserting hidden nodes into traditional ELM and was proved to achieve better performance than traditional methods. In 2017, Ramasamy and Rajaraman [147] proposed their modified approach called Meta-cognitive Restrict- ed Boltzmann Machine (McRBM) which was superior to other approaches such as support vector machine (SVM) and extreme learning machine (ELM) according to their experiment results. It is also demonstrated that RBM could be utilized to improve the performance of ELM by determining the input weights [130, 131]. Wang and Zhang [183] studied various methods for training multiple hidden layer feed-forward neural network, including back propagation (BP) and extreme learning machine (ELM). Their study showed that RBM had advantages such as fast training speed, better generalization, and high understandability compared to BP and ELM. In other fields such as aircraft auxiliary power unit (APU) sensing data prediction [104] and gas path fault diagnosis [108], scientists also revealed that Restricted Boltzmann Machine could be applied to promote performance of extreme learning machine.

4.7 ELM vs deep belief network (DBN)

Deep Belief Network (DBN) is a kind of neural network of machine learning, which can be used for both unsupervised learning and supervised learning. As a probabilistic generation

model, DBN was also proposed by Hinton and Osindero [48]. Compared with the neural network of traditional discriminant model, generation model is to establish a joint distribution between observation data and labels. By training the weights among the neurons, the whole neural network can generate training data according to the maximum probability. Not only can people use DBN to identify features, classify data, but people can also use it to generate data. DBN model is a very practical learning algorithm, which has a wide range of applications and strong expansibility. It can be used in the fields of handwriting recognition, speech recognition and image processing of machine learning.

DBN is composed of multiple layers of neurons (shown in Fig. 8), which are divided into dominant neurons and recessive neurons, which could also be called visible units and hidden units. Visible units are used to accept input, and hidden units are used to extract features. So hidden units also have another name, called feature detector (feature detectors). The connections between the top two levels are undirected, constituting associative memory. There are directed connections between the lower and upper layers. The lowest levels represent data vectors, each neuron representing one dimension of the data vector.

The component of DBN is Restricted Boltzmann Machines (RBM). The process of training DBN is carried out layer by layer. In each layer, the data vector is used to infer the hidden layer, and then this hidden layer is treated as the data vector of the next layer (one layer above). In fact, each RBM can be used separately as a cluster. RBM has only two layers of neurons, the first layer is called the visible layer, which is composed of visible units for the input of training data. Another layer is called the hidden layer, accordingly, consisting of hidden units, used as a feature detector.
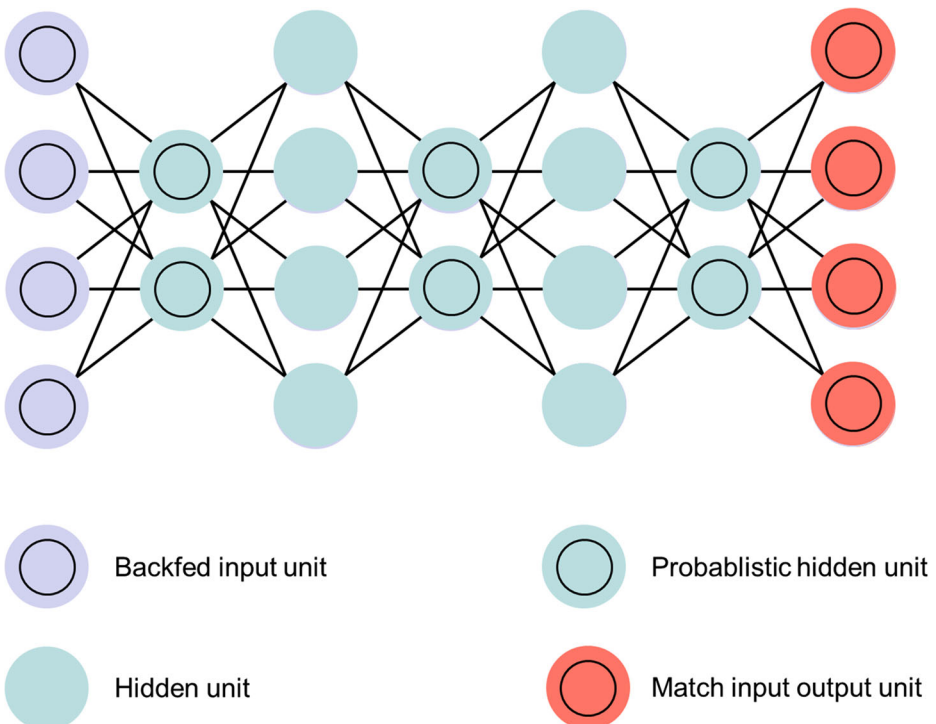

Fig. 8 Illustration of Deep Belief Network (DBN)

The training process of DBN has five steps. First, we need to fully train the first RBM. Second, we fix the first RBM's weights and bias and apply its hidden units as input vector of next RBM. The third step is to fully train the second RBM and heap up the second RBM above the first one. Next, the fourth step, this process should be looped for many times. In the end, if training data has label, not only visible units in uppermost RBM but also neurons with classification labels should be trained together.

Since Deep Belief Network (DBN) was proposed, scientists have not stopped enlarging its ability including embedding advantages of extreme learning machine (ELM) into it. Yu and Zhuang [211] presented a modified architecture of extreme learning machine by applying deep level representation with stacked generalization philosophy. They compared their scheme with other frequently-used methods including DBN and claimed their scheme could obtain better performance. Zhang and Ding [216] also adopted deep representation to improve extreme learning machine which turned out to be superior to DBN method according to their experiment results. Ronoud and Asadi [151] proposed a novel methodology to detect breast cancer by applying extreme learning machine to train a classifier with DBN. Their research proved that the methodology of DBN with ELM had advantage in performance compared to other combination methods such as DBN with genetic algorithm (GA). Zhang and Xiao [217] proposed their approach based on ELM called multilayer probability ELM (MP-ELM) for device-free localization. It was indicated via their research that their approach was better than some other popular methods including Deep Belief Network (DBN) and Deep Boltzmann Machine (DBM).

4.8 ELM vs other deep learning methods

Kasun and Zhou [75] combined autoencoder with ELM and proposed ELM-AE. They constructed a five-layer ELM for experiment on MNIST dataset which consists of 70,000 handwritten digit samples with 60,000 for training and the rest for testing. Results revealed that the multilayer ELM was comparable to deep belief network and deep Boltzmann machine but required less training time. They also discovered that the output weight β has a good capability of representation learning. The effectiveness of ELM-AE are further proven and improved by work in [20, 30, 169]. Zhao and Jiao [236] combined tensor operation with ELM and proposed sparse deep tensor extreme learning machine (SDT-ELM) which achieved good classification performance on three open datasets. Shen and Xiao [159] suggested to hybrid deep CNN and ELM for hyperspectral image classification. They designed a two-branch convolution module for spatial and spectral feature extraction respectively. The generated features were concatenated and sent into stack ELM for classification. Results from benchmark datasets revealed that their method achieved good classification performance with faster speed.

At last, in order to present brief comparison between ELM and these machine learning algorithms, we provide Table 6 to illustrate the strength and popularity of each method.

5 Tasks of ELM

As ELM has many advantages over other machine learning algorithms, it has been widely used to solve real world machine learning problems.

Table 6 Comparison between ELM and other machine learning algorithms

Algorithms	Accuracy	Time consumption	Cross domain adaptation	Popularity nowadays
Extreme learning machine	Comparatively high	Low	Comparatively high	Very popular
Support vector machine	High	Ordinary	Ordinary	Popular
Random neural network	Ordinary	Ordinary	Ordinary	Relatively low
Radial basis function network	Ordinary	Ordinary	Ordinary	Ordinary
Hopfield network	Ordinary	Ordinary	Ordinary	Relatively low
Boltzmann machine	Comparatively high	Slightly high	Ordinary	Relatively low
Restricted boltzmann machine	Comparatively high	Ordinary	Comparatively high	Comparatively popular
Deep belief Network	Comparatively high	Ordinary	Ordinary	Comparatively popular

5.1 Recognition

ELM is now an important and efficient tool for object recognition. Chacko and Vimal Krishnan [13] utilized wavelet energy as feature and ELM as classification algorithm for handwritten character recognition. Malayalam character set was used for evaluation in experiment. Rong and Jia [150] developed an aircraft recognition system based on ELM. Firstly, three different kinds of moments were extracted from aircraft images as the features. Then, the features were divided into subsets to train a set of ELMs. Finally, the output of the system was obtained by weighting on the outputs of each ELM. Generic object recognition remains a challenge for machine learning because of the Intra-class variabilities. Bai and Kasun [4] proposed local receptive fields based extreme learning machine (ELM-LRF) which used convolution and pooling for automatic feature learning and ELM for classification. Comparison was conducted on NORB, ETH-80 and COIL datasets and ELM-LRF achieved comparable classification results. Zhang and He [218] used a hybrid method for cross domain learning. They employed a deep convolution neural network to extract representation from input images and proposed adaptive ELM (AELM) for recognition. Domain related error terms and a regularization term were added in the object function for AELM. Experiment result suggested the effectiveness of their method. Zhou and Wang [242] used extreme learning machine for fabric defect recognition. They proposed to employ k-means singular value decomposition for dictionary learning and trained an adaptive differential evolution algorithm optimized regularization extreme learning machine (ADE-RELM) to recognize the defects such as broken warp, broken weft, etc. Zhang and Jiang [232] proposed to hybrid super pixel pattern and KELM for hyperspectral image recognition and classification which achieved 10% improvement of accuracy over other spectral methods. Tian and Zhang [177] used ELM for sensor-based activity recognition. They firstly trained a set of base ELMs independently and removed the similar ones. Then, glowworm swarm optimization was leveraged to select optimal sub-ensemble of the ELMs. Finally, the output of the system was generated by majority voting of the ensemble. Sharma and Giri [155] suggested to use ELM for intrusion detection. They firstly employed ExtraTrees classifier for feature selection of each intrusion type. Then, a set of ELMs were trained for every types of intrusions. The results from ELMs were refined by Softmax to generate final result. Rodriguez and Barba [149] combined wavelet

packet Fourier entropy and shallow kernel ELM for bearing fault detection. Qiu and Wu [141] proposed an insulator pollution detection method based on ELM for hyperspectral image. Lu and Zhang [109] integrated ELM with hypercomplex space for palm print recognition. Liu and Li [107] utilized 2D Gabor and local binary pattern for feature extraction and trained an ELM for human facial expression recognition.

5.2 Prediction

ELM was employed to solve prediction problems. Yang and Yan [205] used ELM for pressure prediction in the pipeline of coal slurry. Zhang and Liu [223] developed a ELM based system to predict the melt index in propylene polymerization process. Zeng and Zhang [213] trained ELM by switch delayed particle swarm optimization (SDPSO) and applied it to forecast short time load. In their experiment, ELM with SDPSO outperformed radial basis function neural network (RBFNN). Wang and Chen [181] employed grey wolf optimization to train KELM. The trained model was used for bankruptcy prediction. Zou and Yao [247] put forward a convex bidirectional ELM (CB-ELM) to verify and predict the temperature and humidity in greenhouse. In CB-ELM, when a new node is added into the hidden layer, the output weight can be obtained by convex optimization algorithm. Therefore, CB-ELM preserved the features of conventional ELM while improved the accuracy. Zhang and Wei [215] combined wavelet transform, principle component analysis and ELM for solar radiation forecasting. The parameters in ELM were optimized by bat algorithm. Sun and Duan [170] employed ELM optimized by PSO to predict carbon price. Prates [138] proposed spatial ELM and applied it on disease counting prediction. Naz and Javed [121] proposed an electric load and price prediction system based on ELM optimized by smart grid. Liu and Wang [99] leveraged wavelet transform and KELM for passenger flow prediction.

5.3 Representation/feature learning

ELM was used by researchers for representation and feature learning. Ding and Zhang [30] proposed to apply ELM-AE to unsupervised ELM and the features were the output of hidden layer. Zhang and Ding [216] put forward a hybrid scheme for feature learning and classification. Deep belief network was used for feature learning and ELM was used for feature mapping. Zhu and Miao [244] employed ELM-AE as the learning unit in local perceptive fields to construct a multilayer network. The low-level features were fed into the next layers to form high level representation. Then, block histogram was utilized to transform the obtained representations into translation and rotation invariant ones. Cheng and Liu [20] introduced ELM to K-SVD for improvement. Firstly, a denoising deep ELM-AE was used to extract features. Then, K-SVD was utilized to generate sparse representation from the features. The obtained representations can be used for classification and recognition. Experiment results showed the proposed method was comparable with other state-of-the-arts. Chen and Song [18] proposed an ELM based feature transfer method. First of all, like conventional ELM, random mapping was employed to generate features from input for both source and target domain samples. Then, domain transfer was performed with the aim of minimizing the domain distance and preserving the information in target domain as much as possible. Finally, the transferred features were used to train an adaptive ELM for classification. Song and Li [165] proposed an improved ELM for sparse feature learning named OAL1-ELM. They added l_1

regularization in loss function of ELM, which was solved by alternative direction method of multipliers. The OAL1-ELM can be trained on one-by-one samples or in batch mode. Pei and Wang [135] used ELM for label-specific feature learning.

5.4 Clustering

Researchers have paid attention to ELM for clustering, which is an unsupervised learning problem. He and Jin [46] found out that the ELM feature mapping outperformed Mercer kernel function based methods for clustering. Alshamiri and Singh [3] employed artificial bee colony to train ELM for feature projection. The proposed system can overcome the initialization randomness in cluster centers, which works more robustly. Huang and Liu [60] proposed a discriminative clustering based on ELM and Fisher's Linear Discriminant Analysis (LDA). With simple implementation, their approach achieved state-of-the-art performance. Peng and Zheng [137] put forward an improved unsupervised discriminative ELM (UDELM) which leveraged both local manifold and global discriminative learning. Huang and Yu [61] suggested to use Gaussian and sigmoid neurons in ELM hidden layer for feature learning. Alternative direction method was employed to obtain the parameters. Liu and Lekamalage [100] combined ELM and joint embedding clustering. Zhang and Wu [235] exploited the random mapping ability of ELM and proposed a multi-view clustering scheme.

5.5 Surrogate modeling

ELM has also been used as meta-model or surrogate model. Pavelski and Delgado [134] tried to leverage ELM to solve multi-objective optimization problems (MOPs). Unlike the conventional multi-objective evolutionary algorithms, they proposed an ELM surrogate-assisted method which is less expensive. Their method achieved state-of-the-art performance in simulation. Hao and Liu [44] proposed to convert the bottleneck stage scheduling problem to an expensive-to-evaluate optimization problem and used incremental ELM as the surrogate model. Then, a differential evolutionary was utilized to solve the problem. Kang and Li [73] employed an ELM based surrogate model for the analysis of soil slope reliability. Artificial bee colony optimization was used to obtain the best values for the hidden weights and biases in ELM. Pavelski and Delgado [133] proposed a multi-objective decomposition and differential evolutionary algorithm based on ELM. They tested the proposed method on three famous benchmark datasets. Ghiasi and Ghasemi [40] proposed two algorithms to detect structure damage and leveraged 10 surrogate models including ELM. Limited by space, we cannot cover all the subjects. A summary is presented in Table 7.

6 Application fields of ELM

Due to its superiority in training speed, accuracy and generalization, extreme learning machine (ELM) was employed in many application fields such as medicine, chemistry, transportation, economy, robotics and etc. In this section, we will review some representative works in these application fields.

Table 7 A summary of ELM tasks

Recognition

Chacko and Vimal Krishnan [13], Rong and Jia [150], Bai and Kasun [4], Zhang and He [218], Zhou and Wang [242], Zhang and Jiang [232], Tian and Zhang [177], Sharma and Giri [155], Rodriguez and Barba [149], Qiu and Wu [141], Lu and Zhang [109], Liu and Li [107]

Prediction

Feng and Qian [33], Bian and Yang [6], Yang and Yan [205], Zhang and Liu [223], Wang and Chen [181], Zeng and Zhang [213], Zou and Yao [247], Zhang and Wei [215], Sun and Duan [170], Prates [138], Naz and Javed [121], Liu and Wang [99], Kaloop and Kumar [72]

Representation/feature learning

Ding and Zhang [30], Zhang and Ding [216], Zhu and Miao [244], Cheng and Liu [20], Chen and Song [18], Song and Li [165], Pei and Wang [135]

Clustering

He and Jin [46], Alshamiri and Singh [3], Huang and Liu [60], Peng and Zheng [137], Huang and Yu [61], Liu and Lekamalage [100], Zhang and Wu [235]

Surrogate modeling

Pavelski and Delgado [134], Hao and Liu [44], Kang and Li [73], Pavelski and Delgado [133], Ghiasi and Ghasemi [40]

6.1 Medical application

Medical imaging generates a huge number of images every day. ELM is a popular tool in applications in medical imaging fields, such as magnetic resonance imaging (MRI), computerized tomography (CT), and mammogram.

6.1.1 MRI

Computer aided diagnosis is a hot research topic which employs machine learning approaches to detect diseases based on medical images. ELM was found in some computer aided diagnosis systems. Brain MRI analysis is one of fields which attract most scientists' interest. Although many approaches have been applied into Brain MRI analysis, ELM-based method still shows its unique strength. Peng and Lin [136] proposed an attention-deficit/hyperactivity disorder detection method based on ELM and structural MRI. Originally, 340 cortical features were obtained from brain MRIs and reduced by Sequential forward selection method. Both SVM and ELM were trained for classification. They found out that ELM was better in generalization and robustness than SVM. Termenon and Graña [175] proposed an ELM based cocaine dependency detection scheme for brain MRI. They firstly calculated volume of correlation values and segmented the images into regions by watershed algorithm. Then, they measured the region relevance by two methods. ELM based algorithm performed better than other classification algorithms in experiment. Qureshi and Min [142] proposed a novel recursive feature elimination SVM based method for feature selection and trained a hierarchical ELM to detect attention-deficit/hyperactivity disorder. Termenon and Graña [176] firstly used ELM to generate region candidates and employed a majority voting based classification strategy for detecting Alzheimer's disease. Lu and Lu [110] proposed a KELM based pathological brain detection system for magnetic resonance image (MRI), and improved the performance by ELM variants. They firstly performed 2 level 2D wavelet transform on brain images and computed entropies from the obtained sub-coefficients to form the feature vector. Then, the features with their labels were sent into ELM for training and classification. Lama and Gwak [78] proposed a PCA based method to extract features from brain MRIs and compared three

different classification algorithms in experiment: SVM, import vector machine (IVM) and regularized ELM. Simulation results suggested that regularized ELM performed best. Qureshi and Oh [143] put forward a schizophrenia diagnosis approach based on ELM and brain MRI. Independent component analysis was employed, and the functional connectivity-based features were obtained. The cortex features along with shape and textural features were also calculated. Then, a hybrid feature concatenation method was utilized for feature reduction. Finally, ELM was employed for classification. Zhang and Zhao [231] employed oversampling to deal with class imbalanced problem in brain MRI dataset and trained ELM by Jaya algorithm for automatic diagnosis. They utilized wavelet packet Tsallis entropy as the image feature. Nayak and Das [119] utilized multilayer ELM as the classifier and trained the network with multiclass pathological brain images. Their diagnosis system can detect four different categories of brain diseases with a competitive accuracy and at fast speed. Nguyen and Ryu [122] used ELM to identify Alzheimer's disease, mild cognitive impairment and cognitive normal state based on brain MRI. They totally combined ten 3D map features to form the feature vector. Leave one out validation and 10 fold cross validation were employed for evaluation. Gumaei and Hassan [42] employed a regularized ELM for brain tumor identification. They leveraged normalized GIST as feature descriptor and combined it with PCA.

6.1.2 CT

As we know, computed tomography (CT) images play a very important role in medical diagnosis. Huang and Tan [62] proposed a 3D liver computed tomography (CT) segmentation method based on ELM. They treated every voxel as the sample and carried out voxel-level classification by ELM. Then, morphological method was used to improve the segmentation results. Huang and Yang [63] used ELM autoencoder for image feature pre-processing. They trained a set of ELM and generated the final output by majority voting ensemble. Ramalho and Filho [146] applied ELM to detection lung disease in CT. Firstly, the CT images were segmented by an adaptive contour model. Then, structural information was calculated from the segmented images as features. Finally, an ELM was trained to identify two types of lung diseases and healthy control. Zhu and Huang [243] developed liver tumor detection and segmentation system for CT based on ELM. To alleviate the overfitting problem, they employed data ensemble and feature ensemble. A sequential kernel learning strategy was utilized to improve segmentation ability.

6.1.3 Ultrasound

As one of most commonly used medical imaging technologies, ultrasound has irreplaceable value in diagnosis and assessment of many diseases including thyroid nodules. Plenty of methods have been adopted to classify thyroid nodules into malignant one and benign ones [66]. According to some research, ELM-based methods outperformed other algorithms due to its strength in accuracy and fast speed. In 2017, Xia and Chen [195] claimed they firstly investigated the potential of ELM in thyroid nodules classification and achieved promising experimental results, which indicated that ELM could be successfully applied into thyroid nodules classification. In 2019, Cai and Gu [9] proposed a novel scheme for different machine learning tasks including thyroid nodules classification. They improved kernel extreme learning machine based on an enhanced grey wolf optimization strategy to boost the stochastic behavior with the construction of a new

hierarchical mechanism. Their experimental results demonstrated its advance and efficiency.

6.1.4 RNA classification

Researching in RNA assists human in exploring mystery of lives. RNA classification could enable scientists to investigate tumor formation and life mechanisms including growth, development, and immunity to a higher level [70]. Thus, scientists put efforts in developing methods for RNA classification and recognition with machine learning technologies, ELM included. In 2018, Chen and Zhang [16] presented an ELM-based approach for classification between cirRNAs and lncRNAs. They first applied feature selection to construct a feature representation framework extracted from RNA's graph, sequence, and conservation properties. Then they utilized hierarchical extreme learning machine as classifier and claimed that they obtained convincing results. In 2020, Niu and Zhang [123] proposed their method for recognition of cirRNAs based on improved extreme learning machine. They applied particle swarm optimization algorithm to elevate the performance of ELM classifier. According to the experimental results, their method showed potential in future research.

6.1.5 EEG

The electroencephalogram (EEG) is regarded as the most effective way to monitor the status of patients with epilepsy. Among many machine learning algorithms, ELM-based method was widely applied into seizure detection with EEG information for its efficiency and simplicity. In 2011, Yuan and Zhou [212] proposed an approach for epileptic EEG classification. They adopted extreme learning machine to train single hidden layer feedforward networks with extracted nonlinear features such as approximate entropy, Hurst exponent, and detrended fluctuation analysis exponent. They compared ELM-based approach with BP and SVM, concluding that ELM-based approach achieved best experimental results. Song and Crowcroft [166] presented an improved ELM-based approach for automated detection of seizure with optimized sample entropy algorithm. They claimed their approach not only obtained outstanding accuracy but also had the advantage of fast speed inherited from ELM which made it potential to build real-time detection system. In 2016, Song and Hu [164] put forward an automatic epileptic detection approach with EEG information. They fused several level features including both the Mahalanobis-similarity-based feature and the sample-entropy-based feature and employed ELM to construct seizure detection based on this fusion feature. Their experiments proved this approach's efficiency. Zhou and Tang [241] proposed a novel ELM-based method for epileptic EEG classification combined with cellular automata and achieved better results compared to BP and SVM.

6.1.6 Mammogram

Vani and Savitha [178] applied ELM for abnormality detection in Mammograms. They initially selected 15 different shape and textual feature descriptors for feature extraction and reduced the feature number gradually for experiment. They found that 9 features performed the best. Malar and Kandaswamy [111] combined grey level spatial dependence matrix with wavelet features to form the feature vector, and used ELM to detect breast cancer based on those features obtained from mammogram. Wang and Yu [187] used ELM to detect breast

cancer in mammography. For segmentation, wavelet transform, region growth and morphology were used. Then, textual and morphological features were obtained from the segmented images and sent into ELM for training and classification. Xie and Li [197] firstly segmented the masses in mammogram by a level set model. Then, multi-dimensional features were extracted and reduced by SVM and ELM based feature selection method. Finally, those features were sent into ELM classifier for training. The trained ELM can classify the masses into benign or malignant accurately. Wang and Qu [186] put forward a breast tumor detection method based on ELM and mammogram. They fused geometry and textural features from two views of mammogram and optimized the feature model with feature selection. ELM served as the classification algorithm. Hu and Yang [51] proposed a novel hidden Markov tree model of dual-tree complex wavelet transform for feature extraction from mammogram. They also optimized the features by genetic algorithm. ELM was trained to classify the samples as malignant or benign. Wang and Li [185] employed ELM for breast cancer detection. They firstly detect masses based on deep CNN features and unsupervised ELM features. Then, several different features were fused as the feature vector. Finally, an ELM was trained to classify the masses as benign or malignant. A summary is presented in Table 8.

6.2 Chemistry application

ELM has also been applied in other fields, like chemistry. Geng and Dong [38] proposed a self-organizing ELM to predict variables in chemical processes. They firstly tuned the input weight by Hebbian rule and optimized the structure using mutual information. Then, the output weight was obtained by pseudo inverse. Experiment results from UCI data revealed that their method was better in generalization compared to original ELM. Qin and Li [139] applied ELM to evaluate the green management in power generation enterprises in China. Firstly, an evaluation indicator system for low carbon sustainability was created by an improved dynamic hesitation degree. Then, an ELM with RBF kernel was trained to implement comprehensive evaluation. Their model achieved good results in experiment, and it can be used in similar projects. In 2013, Wang and Jin [179] introduced ELM into prediction of the nonlinear optical property. In 2017, You and Zhou [208] employed ELM in prediction of protein-protein interactions (PPIs) and claimed to obtain satisfying performance. In 2018, You and Zhou [208] designed and implemented experiments to compare different kinds of classifiers'

Table 8 Application summary in medical imaging

MRI
Peng and Lin [136], Termenon and Graña [175], Qureshi and Min [142], Termenon and Graña [176], Lu and Lu [110], Lama and Gwak [78], Qureshi and Oh [143], Zhang and Zhao [231], Nayak and Das [119], Nguyen and Ryu [122], Gumaei and Hassan [42]
CT
Huang and Tan [62], Huang and Yang [63], Ramalho and Filho [146], Zhu and Huang [243]
Ultrasound
Xia and Chen [195], Cai and Gu [9]
RNA classification
Chen and Zhang [16], Niu and Zhang [123]
EEG
Yuan and Zhou [212], Song and Crowcroft [166], Song and Hu [164], Zhou and Tang [241]
Mammogram
Vani and Savitha [178], Malar and Kandaswamy [111], Wang and Yu [187], Xie and Li [197], Wang and Qu [186], Hu and Yang [51], Wang and Li [185]

performance in predicting protein secondary structure. According their results, extreme learning machine (ELM) had the fastest training speed, but support vector machine (SVM) was the most accurate. Cao and Zhu [12] also compared various approaches to predict the toxicity of ionic liquids and they claimed ELM had the best performance, beating multiple linear regression (MLR) and support vector machine (SVM). Kang and Liu [74] brought the competition into predicting the Henry's law constant (HLC) of CO2. And it seemed that ELM won again among several competitors including MLR and SVM. In 2019, Lei and Wen [80] proposed their method of predicting protein-protein interactions based on regularized extreme learning machine and claimed to achieve promising performance. Further in 2020, Li and Shi [83] also presented their modified method to predict protein-protein interactions. They adopted weighted extreme learning machine as basis with combination of scale-invariant feature transform to improve the model's performance and gained the ideal results. Ouyang and Wang [129] applied ELM into measuring NOx in vehicle exhaust and achieved qualified experiment results. And it could be observed that more scientists were still putting efforts into applying ELM based methods to chemical analysis [162, 180].

6.3 Economy application

Extreme learning machine was also applied in solving economic problem. In 2012, Landa-Torres and Ortiz-Garcia [79] proposed their method based on ELM to evaluate the internationalization success of companies. In 2016, Sokolov-Mladenović and Milovančević [163] applied ELM with data of trade, import and export to predict economic growth and claimed to obtain convincing results. Marković and Petković [113] also adopted ELM for similar research but using data of science and technology development and investment in nations. Moreover, Milačić and Jović [116] employed ELM to predict economic growth with data of agriculture, manufacturing and industry. They utilized artificial neural network (ANN) with back propagation (BP) as contrast and concluded that ELM was superior to contrast approach. Further, Rakic and Milenkovic [145] presented their study of applying ELM to forecast economic growth based on information technology level of nations and also claimed to obtain promising results. Their research indicated that extreme learning machine could be applied successfully in forecasting economic growth based on statistical data.

The energy industry is one of most crucial departments in economy. The consumption of energy could reflect vitality of economy. Thus, scientists often study trend of economic development via analysis of energy industry and its consumption. For example, emission of carbon dioxide (CO2) presents a significant signal of energy consumption during economic activities. Many researchers have applied ELM into analyzing CO2 emission to predict economic development[112, 161, 171]. Except for data of carbon emission, scientists also utilized the data of demand and price from energy resources market to forecast economic growth via ELM [22, 153, 172] and proved that ELM was qualified in these situation of economic analysis field. Besides energy, in 2019, Shoumo and Dhruba [160] implemented a comparative research of various machine learning models in credit risk assessment including ELM, SVM, random forest and logistic regression. And ELM showed its advantage of quickness and simplicity among these various methods.

6.4 Transportation application

Transportation is a hot topic for applying machine learning into application. For example, scientists utilized machine learning technique to establish driver drowsiness monitoring system

to avoid dangerous driving and save people lives [69]. It has been a long time since extreme learning machine was applied in solving problems of transportation field. Sun and Ng [173] proposed a two-stage approach which combined linear programming (LP) and extreme learning machine (ELM) to optimize transportation system. Their experiments proved that their combined approach could improve lifetime of transportation system and increased its reliability. In 2016, Liu and Yang [101] studied semi-supervised methods for real-time driver distraction detection. They analyzed Laplacian support vector machine and semi-supervised extreme learning machine and concluded that these two semi-supervised methods could mitigate the cost of labeled training data in building driver distraction detection system. In 2017, Liu and Yan [93] applied radial basis function into extreme learning machine to predict road surface temperature and achieved hopeful experiment results. Oneto and Fumeo [126] applied ELM into predicting dynamic delay of large-scale railway network. Zeng and Xu [214] adopted ELM to recognize traffic sign with deep features. Zhang and He [229] employed ELM to analyze traffic accidents based on video data. In 2018, Wang and Chow [182] applied ELM with large-scale GPS data to help taxi drivers searching best routes. Similar to their research, more scientists have utilized ELM to predict traffic flow for drivers and governments [94, 199]. And it has become a popular trend in transportation research field.

6.5 Robotics application

It is because ELM had the advantage of fast training speed, it was widely employed in robotics application field. In 2015, Sun and Yu [168] proposed an ELM based method to perform object grasping detection. They applied ELM as the basis of their cascade classifier and adopted histograms of oriented gradients (HOG) as features. Their method was announced to achieve better performance than other benchmark methods. In 2016, Alcin and Ucar [2] also applied ELM in robotic arms control and claimed their approach was superior to Artificial Neural Network (ANN) model. In 2018, She and Hu [156] introduced ELM into the field of brain-computer interface. They utilized extreme learning machine to classifier electroenceph-alographic (EEG) signals and obtained promising experiment results. In 2019, Duan and Ou [32] applied ELM to build their dynamical system for robotic motion control. Via their experiments, it was claimed that ELM's characteristics of efficiency and accuracy brought their approach better performance compared to other approaches.

6.6 IoT Application

As Internet of Things attracting continuous attention from both academic and industrial field in recent years, more and more scientists developed quantity of approaches or applications in IoT with modern advanced information technologies [71]. There existed a lot of approaches which applied ELM in IoT application as well. Rathore and Park [148] proposed an ELM-based approach for detection of cyber-attack. They built an attack detection framework based on fog computing and adopted improved ELM as classifier to differentiate attack from normal visit. According to their experiments, the proposed approach worked. Yin and Dong [207] presented an algorithm to assign bug fixing jobs to engineers with ELM. They combined diversified feature selection with extreme learning machine to build classifier and claimed to achieve promising results. In 2019, Li and Wei [89] analyzed DDoS attack to propose joint entropy features. combining with extreme learning machine for detection of DDoS attack. The experimental results proved this method was efficient and rapid. In 2020, Zheng and Hong

[238] put forward a novel approach for intrusion detection which combined several useful scheme including extreme learning machine. They first applied improved linear discriminant analysis (LDA) for reduction of feature dimensions. Then the approach adopted single hidden layer feedforward neural network as classifier to process dimension-reduced data with ELM optimization. The experimental results demonstrated that their approach obtained the performance of outstanding real-time characteristics and generalization.

6.7 Geography application

ELM can also been applied in Geography. Huang and Yin [52] used ELM to generate the landslide susceptibility indexes. Firstly, a rough index map was obtained by self-organizing map network, and the very low susceptible areas were removed. Then, the final landslide susceptibility index map was generated using by ELM. Their SOM-ELM outperformed single ELM and SOM-SVM on real data. Wei and Liu [189] tried to evaluate the stability of rubble-mound breakwaters using ELM. They trained 5000 different ELMs to select the best one based on testing performance. Wei and Li [188] used ELM to predict sediment-carrying capacity. They used 30 sets of data for training and 4 sets of data for validation. The input vector of ELM included surface velocity, slope and settling velocity, and the output was sediment-carrying capacity. Detailed discussion of the parameters was presented. Niu and Feng [124] developed four smart systems for deriving the operation rule of hydropower reservoir based on multiple linear regression, artificial neural network, SVM and ELM. From simulation results, they found artificial neural network, SVM and ELM performed better than multiple linear regression. Li and Chen [85] presented an efficient method for hyperspectral imagery classification. They adopted local binary patterns which was to extract local features like edges or corners and then combined feature-level fusion and decision-level fusion. Using the classifier, which was constructed by extreme learning machine, they reported to achieve promising results better than other traditional methods.

6.8 Food Industry application

ELM is a popular tool in food safety. Geng and Zhao [39] proposed a food safety monitoring system based on ELM. Analytic hierarchy process (AHP) was utilized to generate process characteristic information (PCIs). Then, the PCIs were sent into ELM for prediction. The prediction results along with the PCIs can provide better information for food safety monitoring. Ouyang and Chen [128] raised an approach for detecting amino acid nitrogen content in soy sauce. They applied near infrared spectroscopy and then employed feature selection. Using ELM as classifier, they claimed their approach outperformed other popular methods in construction of amino acid nitrogen detection system for quality assessment of soy sauce. da Costa and Llobodanin [25] utilized similar approach of combining feature selection and ELM for classification of different kinds of wines. Liu and Li [105] demonstrated ELM was also adaptive for prediction during large-scale food sampling analysis. Zhang and Zhou [228] applied ELM in prediction of dairy food safety from big data and proved ELM's superiority to other methods such as BP neural network and support vector machine. Guo and Ma [43] employed impact acoustic signal to collect features of wheat kernel and then applied the feature with ELM classifier to differentiate kernel-damaged wheat from kernel-undamaged wheat. The experimental results proved that their ELM-based approach could successfully build a wheat kernel classification system.

6.9 Other interdisciplinary application fields

Control system: Zhang and Ma [230] put forward an error minimized regularized OS-ELM for estimating time-varying parameters in model-free adaptive control. The improved ELM algorithm can train the parameter and optimize the network structure simultaneously. Xu and Lei [200] proposed an adaptive neural control method for pure-feedback systems based on ELM. They firstly transformed the system into nonaffine system by equivalent transform. Then, system state variables were estimated by finite-time-convergence-differentiator. Finally, ELM was employed to estimate the unknown term of the nonaffine system. Wong and Huang [192] leveraged ELM for automotive engine idle speed control. A novel adaptive law was combined with ELM based on Lyapunov analysis. Simulation results suggested that their method outperformed other state-of-the-art approaches. Video analysis: Yi and Dai [206] proposed a moving cast shadow detection method based on ELM. Firstly, the pixel-level features and region-level features of different channels were combined. Then, ELM was trained to classify the pixels as shadow or not. Rajpal and Mishra [144] proposed their watermarking algorithm based on bi-directional ELM. For odd or even number of hidden nodes, the bi-directional ELM used two different methods to update the parameters in order to minimize the error to zero. A threshold based fuzzy algorithm was employed to select key frames and bi-directional ELM was trained to produce the LL3 sub-band coefficients of blue channel for watermarking. Their scheme was fast enough for real applications. Yu and Song [210] developed an automatic violent scene detection system based on KELM. 3D histogram of gradient orientation (HOG) was used as the feature extractor and visual words were generated by K means clustering. The combined features were fed into KELM for classification. Experiment results showed that the method was effective and efficient.

7 ELM controversies

Actually, there are some controversies about ELM as two anonymous reviewers suggested. The idea of randomizing some of weights to reduce the complexity of weight learning is not new. It was employed in reservoir computing in 2002, and in convolutional learning in 1989. The Schmidt neural network (SNN) proposed by Schmidt and Kraaijveld [154] and the random vector functional link net (RVFL) proposed by Pao and Park [132] were proposed over ten years before ELM. Both SNN and RVFL used random hidden weights and biases, and ELM can be regarded as a special form of RVFL or radial basis function neural network. Moreover, ELM is almost the same as SNN, the only difference of ELM and SNN is that SNN has trainable output biases while the output biases of ELM are always zeros.

8 Conclusion

In this review, we presented the development and application of ELM over the last decade. The principle and theoretical analysis of ELM are given and a wide range of methods to improve the performance of ELMs for specific learning tasks are reported as well. Due to its fast speed, good generalization and easy implementation, ELM is often preferred than other state-of-the-art approaches for artificial intelligence related problems in a variety of fields. Through the survey, we found several interesting research directions and problems for future research.

ELM has potential of playing a more important role in big data. With the invention of deep learning structures and effective training algorithms, we are able to solve big data machine learning problems. Deep learning has achieved astonishing performance, but the computational complexity is high, and the training process often requires long time. ELM shows the potential of random mapping which can achieve the same classification performance as the trained parameters do. Therefore, it will be interesting to combine deep learning and ELM to learn from big data. The hybrid network may be able to yield the same results as deep learning does with less training time.

Theoretical justification of ELM never stop evolution. Random mapping in hidden layer is the most significant idea for ELM, which enables the universal approximation and contributes to the extremely fast training speed. But it remains a problem for justifying the randomness. It may be helpful to compare ELM with other random mechanisms, such as random forest.

Another promising direction of ELM development may be connected to biological learning. Researchers have found that the neurons in our brain also use random mechanism for learning [17]. As artificial neural network is originally inspired by our brain, it will be interesting to investigate the underlying connection of ELM and biological learning.

References

1. Ackley DH, Hinton GE, Sejnowski TJ (1985) A learning algorithm for boltzmann machines. Cogn Sci 9(1):147–169
2. Alcin OF, Ucar F, Korkmaz D (2016) Extreme learning machine based robotic arm modeling. In: 2016 21st International Conference on Methods and Models in Automation and Robotics (MMAR). Miedzyzdroje, Poland, pp 1160–1163
3. Alshamiri AK, Singh A, Surampudi BR (2015) Artificial bee colony algorithm for clustering: an extreme learning approach. Soft Comput 20(8):3163–3176
4. Bai Z, Kasun LLC, Huang G-B (2015) Generic Object Recognition with Local Receptive Fields Based Extreme Learning Machine. Procedia Comput Sci 53:391–399
5. Bhat AU, Merchant SS, Bhagwat SS (2008) Prediction of melting points of organic compounds using extreme learning machines. Ind Eng Chem Res 47(3):920–925
6. Bian Y, Yang J, Li M, Lan R (2013) Automated Flare Prediction Using Extreme Learning Machine. Mathematic Prob Eng 2013:1–7
7. Broomhead D, Lowe D (1988) Radial basis functions, multi-variable functional interpolation and adaptive networks. royal signals and radar establishment malvern (United Kingdom). Vol. RSRE-MEMO-4148

8. Bu YD, Zhao G, Luo AL, Pan JC, Chen YQ (2015) Restricted Boltzmann machine: a non-linear substitute for PCA in spectral processing. Astron Astrophys 576:8 Article ID: A96

9. Cai Z, Gu J, Luo J, Zhang Q, Chen H, Pan Z, Li Y, Li C (2019) Evolving an optimal kernel extreme learning machine by using an enhanced grey wolf optimization strategy. Expert Syst Appl 138:112814

10. Cao J, Lin Z, Huang G-B (2012) Self-Adaptive Evolutionary Extreme Learning Machine. Neural Process Lett 36(3):285–305

11. Cao J, Lin Z, Huang G-B, Liu N (2012) Voting based extreme learning machine. Inf Sci 185(1):66–77

12. Cao LD, Zhu P, Zhao YS, Zhao JH (2018) Using machine learning and quantum chemistry descriptors to predict the toxicity of ionic liquids. J Hazard Mater 352:17–26

13. Chacko BP, Vimal Krishnan VR, Raju G, Babu Anto P (2011) Handwritten character recognition using wavelet energy and extreme learning machine. Int J Mach Learn Cybern 3(2):149–161

14. Chen K, Lv Q, Lu Y, Dou Y (2017) Robust regularized extreme learning machine for regression using iteratively reweighted least squares. Neurocomputing 230:345–358

15. Chen L, Yang L, Sun C, Xi H (2017) A fast RBM-hidden-nodes based extreme learning machine. In: 29th Chinese Control And Decision Conference (CCDC). IEEE, Chongqing, pp 2121–2126

16. Chen L, Zhang Y-H, Huang G, Pan X, Wang S, Huang T, Cai Y-D (2018) Discriminating cirRNAs from other lncRNAs using a hierarchical extreme learning machine (H-ELM) algorithm with feature selection. Mol Gen Genomics 293(1):137–149

17. Chen ST, Tan DP (2018) A SA-ANN-Based Modeling Method for Human Cognition Mechanism and the PSACO Cognition Algorithm. Complexity: 21: Article ID: 6264124

18. Chen Y, Song S, Li S, Yang L, Wu C (2019) Domain Space Transfer Extreme Learning Machine for Domain Adaptation. IEEE Trans Cybernet 49(5):1909–1922

19. Chen Z, Wang S, Shen Z, Chen Y, Zhao Z (2013) Online sequential ELM based transfer learning for transportation mode recognition. In: 2013 IEEE Conference on Cybernetics and Intelligent Systems (CIS). p. 78-83

20. Cheng X, Liu H, Xu X, Sun F (2016) Denoising deep extreme learning machine for sparse representation. Memetic Computing 9(3):199–212

21. Chorowski J, Wang J, Zurada JM (2014) Review and performance comparison of SVM- and ELM-based classifiers. Neurocomputing 128:507–516

22. Cogoljević D, Alizamir M, Piljan I, Piljan T, Prljić K, Zimonjić S (2018) A machine learning approach for predicting the relationship between energy resources and economic development. Physica A: Stat Mech Appl 495:211–214

23. CORTES C, VAPNIK V (1995) Support-Vector Networks. Mach Learn 20(20):273–297

24. Cortes C, Vapnik V (1995) Support-Vector Networks. Mach Learn 20(3):273–297

25. da Costa NL, Llobodanin LAG, de Lima MD, Castro IA, Barbosa R (2018) Geographical recognition of Syrah wines by combining feature selection with Extreme Learning Machine. Measurement 120:92–99

26. Deng W-Y, Ong Y-S, Zheng Q-H (2016) A Fast Reduced Kernel Extreme Learning Machine. Neural Netw 76:29–38

27. Deng W-Y, Zheng Q-H, Wang Z-M (2014) Cross-person activity recognition using reduced kernel extreme learning machine. Neural Netw 53:1–7

28. Deng W, Zheng Q, Zhang K (2013) Reduced Kernel Extreme Learning Machine. Springer International Publishing, Heidelberg, pp 63–69

29. Deng WY, Bai Z, Huang GB, Zheng QH (2016) A Fast SVD-Hidden-nodes based Extreme Learning Machine for Large-Scale Data Analytics. Neural Netw 77:14–28

30. Ding S, Zhang N, Zhang J, Xu X, Shi Z (2015) Unsupervised extreme learning machine with representational features. Int J Mach Learn Cybern 8(2):587–591

31. Ding SF, Ma G, Shi ZZ (2014) A novel self-adaptive extreme learning machine based on affinity propagation for radial basis function neural network. Neural Comput Applic 24(7-8):1487–1495

32. Duan J, Ou Y, Hu J, Wang Z, Jin S, Xu C (2019) Fast and Stable Learning of Dynamical Systems Based on Extreme Learning Machine. IEEE Trans Syst Man Cybernet: Syst 49(6):1175–1185

33. Feng G, Qian Z, Dai N (2012) Reversible watermarking via extreme learning machine prediction. Neurocomputing 82:62–68

34. Feng GR, Huang GB, Lin QP, Gay R (2009) Error Minimized Extreme Learning Machine With Growth of Hidden Nodes and Incremental Learning. IEEE Trans Neural Netw 20(8):1352–1357

35. Fernandez-Navarro F, Hervas-Martinez C, Sanchez-Monedero J, Gutierrez PA (2011) MELM-GRBF: A modified version of the extreme learning machine for generalized radial basis function neural networks. Neurocomputing 74(16):2502–2510

36. Gao H, Liu X-W, Peng Y-X, Jian S-L (2015) Sample-Based Extreme Learning Machine with Missing Data. Math Probl Eng 2015:1–11

37. Gelenbe E (1989) Random Neural Networks with Negative and Positive Signals and Product Form Solution 1(4): 502-510

38. Geng Z, Dong J, Chen J, Han Y (2017) A new Self-Organizing Extreme Learning Machine soft sensor model and its applications in complicated chemical processes. Eng Appl Artif Intell 62:38–50

39. Geng Z, Zhao S, Tao G, Han Y (2017) Early warning modeling and analysis based on analytic hierarchy process integrated extreme learning machine (AHP-ELM): Application to food safety. Food Control 78: 33–42

40. Ghiasi R, Ghasemi MR, Noori M (2018) Comparative studies of metamoneling and AI-Based techniques in damage detection of structures. Adv Eng Softw 125:101–112

41. Guang-Bin H, Qin-Yu Z, Chee-Kheong S (2006) Extreme learning machine: Theory and applications. Neurocomputing 70(1-3):489–501

42. Gumaei A, Hassan MM, Hassan MR, Alelaiwi A, Fortino G (2019) A Hybrid Feature Extraction Method With Regularized Extreme Learning Machine for Brain Tumor Classification. IEEE Access 7:36266–36273

43. Guo M, Ma Y, Yang X, Mankin RW (2019) Detection of damaged wheat kernels using an impact acoustic signal processing technique based on Gaussian modelling and an improved extreme learning machine algorithm. Biosyst Eng 184:37–44

44. J-h H, Liu M, Lin J-h WC (2016) A hybrid differential evolution approach based on surrogate modelling for scheduling bottleneck stages. Comput Oper Res 66:215–224

45. Hassan AR (2015) A comparative study of various classifiers for automated sleep apnea screening based on single-lead electrocardiogram. In: International Conference on Electrical & Electronic Engineering (ICEEE). IEEE, Rajshahi, pp 45–48

46. He Q, Jin X, Du C, Zhuang F, Shi Z (2014) Clustering in extreme learning machine feature space. Neurocomputing 128:88–95

47. He Xiao-Hui, Wang Dong, Li Yan-Feng, Zhou Chun-Hua (2016) A Novel Bearing Fault Diagnosis Method Based on Gaussian Restricted Boltzmann Machine. Mathematical Problems in Engineering. 2016, Vol. 2016. https://doi.org/10.1155/2016/2957083

48. Hinton GE, Osindero S, Teh Y-W (2006) A Fast Learning Algorithm for Deep Belief Nets. Neural Comput 18(7):1527–1554

49. Hinton GE, Salakhutdinov RR (2006) Reducing the Dimensionality of Data with Neural Networks. Science 313(5786):504

50. Hopfield JJ (1982) Neural networks and physical systems with emergent collective computational abilities. Proc Natl Acad Sci USA 79(8):2554–2558

51. Hu K, Yang W, Gao X (2017) Microcalcification diagnosis in digital mammography using extreme learning machine based on hidden Markov tree model of dual-tree complex wavelet transform. Expert Syst Appl 86:135–144

52. Huang F, Yin K, Huang J, Gui L, Wang P (2017) Landslide susceptibility mapping based on self-organizing-map network and extreme learning machine. Eng Geol 223:11–22

53. Huang G-B, Chen L (2007) Convex incremental extreme learning machine. Neurocomputing 70(16-18): 3056–3062

54. Huang G-B, Chen L, Siew C-K (2006) Universal approximaation using incremental constructive feedforward networks with random hidden nodes. IEEE Trans Neural Netw 17(4):879–892

55. Huang G-B, Ding X, Zhou H (2010) Optimization method based extreme learning machine for classification. Neurocomputing 74(1-3):155–163

56. Huang G-B, Liang N-Y, Rong H-J, Saratchandran P, Sundararajan N (2005) On-Line Sequential Extreme Learning Machine. In: IASTED International Conference on Computational Intelligence. IASTED. p. 123-130

57. Huang G-B, Zhou H, Ding X, Zhang R (2012) Extreme Learning Machine for Regression and Multiclass Classification. IEEE Trans Syst Man Cybernet—Part B: Cybernet 42(2):513–529

58. Huang G-B, Zhu Q-Y, Siew C-K (2006) Extreme Learning Machine: A New Learning Scheme of Feedforward Neural Networks. Proc Intl Joint Confer Neural Networks (IJCNN2004) 2:985–990

59. Huang G, Huang GB, Song S, You K (2015) Trends in extreme learning machines: a review. Neural Netw 61:32–48

60. Huang G, Liu T, Yang Y, Lin Z, Song S, Wu C (2015) Discriminative clustering via extreme learning machine. Neural Netw 70:1–8

61. Huang J, Yu ZL, Gu Z (2018) A clustering method based on extreme learning machine. Neurocomputing 277:108–119

62. Huang W, Tan ZM, Lin Z, Huang G-B, Zhou J, Chui CK, Su Y, Chang S (2012) A Semi-automatic Approach to the Segmentation of Liver Parenchyma from 3D CT Images with Extreme Learning Machine. In: 34th Annual International Conference of the IEEE EMBS. IEEE, San Diego, pp 28–33

63. Huang W, Yang Y, Lin Z, Huang G-B, Zhou J, Duan Y, Xiong W (2014) Random Feature Subspace Ensemble Based Extreme Learning Machine for Liver Tumor Detection and Segmentation. In: 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE, Chicago, pp 126–130
64. Huynh HT, Won Y (2011) Regularized online sequential learning algorithm for single-hidden layer feedforward neural networks. Pattern Recogn Lett 32(14):1930–1935
65. Jayaweera CD, Othman MR, Aziz N (2019) Improved predictive capability of coagulation process by extreme learning machine with radial basis function. J Water Process Eng 32:9 Article ID: 100977
66. Jiang Y, Deng Z, Chen J, Wu H, Choi K-S, Wang S (2017) Intelligent Diagnostic Methods for Thyroid Nodules. J Med Imaging Health Inform 7(8):1772–1779
67. Jiang Y, Deng Z, Choi K-S, Chung F-L, Wang S (2016) A novel multi-task TSK fuzzy classifier and its enhanced version for labeling-risk-aware multi-task classification. Inf Sci 357:39–60
68. Jiang Y, Xue J, Wang R, Xia K, Gu X, Zhu J, Liu L, Qian P (2019) Seizure Recognition Using a Novel Multitask Radial Basis Function Neural Network. J Med Imaging Health Inform 9(9):1865–1870
69. Jiang Y, Zhang Y, Lin C, Wu D, Lin C (2020) EEG-Based Driver Drowsiness Estimation Using an Online Multi-View and Transfer TSK Fuzzy System. IEEE Trans Intell Transport Syst: 1-13
70. Jiang Y, Zhu J, Gu X, Xue J, Zhao K, Ni T, Qian P (2020) Recognition of Large-Scale ncRNA Data Using a Novel Multitask Cross-Learning 0-Order TSK Fuzzy Classifier. J Med Imaging Health Inform 10(2): 502–507
71. Jiang YZ, Bi AQ, Xia KJ, Xue J, Qian PJ (2020) Exemplar-based data stream clustering toward Internet of Things. J Supercomput 76:2929–2957
72. Kaloop MR, Kumar D, Samui P, Gabr AR, Hu JW, Jin X, Roy B (2019) Particle Swarm Optimization Algorithm-Extreme Learning Machine (PSO-ELM) Model for Predicting Resilient Modulus of Stabilized Aggregate Bases. Appl Sci 9(16): Article ID: 3221
73. Kang F, Li J-S, Wang Y, Li J (2016) Extreme learning machine-based surrogate model for analyzing system reliability of soil slopes. Eur J Environ Civ Eng 21(11):1341–1362
74. Kang XJ, Liu CJ, Zeng SJ, Zhao ZJ, Qian JG, Zhao YS (2018) Prediction of Henry's law constant of CO_2 in ionic liquids based on S-EP and S sigma-profile molecular descriptors. J Mol Liq 262:139–147
75. Kasun LLC, Zhou H, Huang G (2013) Representational learning with ELMs for big data. IEEE Intell Syst 4:1–4
76. Krishnasamy G, Paramesran R (2016) Hessian semi-supervised extreme learning machine. Neurocomputing 207:560–567
77. Lam D, Wunsch D (2017) Unsupervised Feature Learning Classification With Radial Basis Function Extreme Learning Machine Using Graphic Processors. Ieee Trans Cybernet 47(1):224–231
78. Lama RK, Gwak J, Park JS, Lee SW (2017) Diagnosis of Alzheimer's Disease Based on Structural MRI Images Using a Regularized Extreme Learning Machine and PCA Features. J Healthc Eng 2017:5485080
79. Landa-Torres I, Ortiz-Garcia EG, Salcedo-Sanz S, Segovia-Vargas MJ, Gil-Lopez S, Miranda M, Leiva-Murillo JM, Ser JD (2012) Evaluating the Internationalization Success of Companies Through a Hybrid Grouping Harmony Search—Extreme Learning Machine Approach. IEEE J Select Topics Signal Process 6(4):388–398
80. Lei H, Wen Y, You Z, Elazab A, Tan E, Zhao Y, Lei B (2019) Protein–Protein Interactions Prediction via Multimodal Deep Polynomial Network and Regularized Extreme Learning Machine. IEEE J Biomed Health Inform 23(3):1290–1303
81. Leung HC, Leung CS, Wong EWM (2019) Fault and Noise Tolerance in the Incremental Extreme Learning Machine. IEEE Access 7:155171–155183
82. Li H-T, Chou C-Y, Chen Y-T, Wang S-H, Wu A-Y (2019) Robust and Lightweight Ensemble Extreme Learning Machine Engine Based on Eigenspace Domain for Compressed Learning. IEEE Trans Circuits Syst I: Regular Papers 66(12):4699–4712
83. Li J, Shi X, You Z, Yi H, Chen Z, Lin Q, Fang M (2020) Using Weighted Extreme Learning Machine Combined with Scale-invariant Feature Transform to Predict Protein-Protein Interactions from Protein Evolutionary Information. IEEE/ACM Trans Computational Biol Bioinformatics 17:1546–1554
84. Li MB, Huang GB, Saratchandran P, Sundararajan N, *Channel equalization using complex extreme learning machine with RBF kernels*, in *Advances in Neural Networks - Isnn 2006, Pt 3, Proceedings*, J. Wang, et al., Editors. 2006, Springer-Verlag Berlin: Berlin. p. 114-119.
85. Li W, Chen C, Su H, Du Q (2015) Local Binary Patterns and Extreme Learning Machine for Hyperspectral Imagery Classification. IEEE Trans Geosci Remote Sens 53(7):3681–3693
86. Li X, He S, Wei Z, Wu L (2019) Improved Online Sequential Extreme Learning Machine: A New Intelligent Evaluation Method for AZ-Style Algorithms. IEEE Access 7:124891–124901
87. Li X, Mao W, Jiang W (2016) Extreme learning machine based transfer learning for data classification. Neurocomputing 174:203–210

88. Li X, Mao W, Jiang W, Yao Y (2016) Extreme learning machine via free sparse transfer representation optimization. Memetic Computing 8(2):85–95
89. Li Z, Wei L, Li W, Wei L, Chen M, Lv M, Zhi X, Wang C, Gao N (2019) Research on DDoS Attack Detection Based on ELM in IoT Environment. In: 2019 IEEE 10th International Conference on Software Engineering and Service Science (ICSESS). p. 144-148
90. Liang H, Cheng H, Wei J, Zhang L, Yang L, Zhao Y, Guo H (2019) Memristive Neural Networks: A Neuromorphic Paradigm for Extreme Learning Machine. IEEE Trans Emerg Topics Computational Intell 3(1):15–23
91. Lima AR, Cannon AJ, Hsieh WW (2015) Nonlinear regression in environmental sciences using extreme learning machines: A comparative evaluation. Environ Model Softw 73:175–188
92. Liu B, Xia S-X, Meng F-R, Zhou Y (2015) Manifold regularized extreme learning machine. Neural Comput & Applic 27(2):255–269
93. Liu B, Yan S, You H, Dong Y, Li J, Li Y, Lang J, Gu R (2017) An Ensembled RBF Extreme Learning Machine to Forecast Road Surface Temperature. In: 2017 16th IEEE International Conference on Machine Learning and Applications (ICMLA). p. 977-980
94. Liu J, Li Q, Chen W, Yan Y, Wang X (2019) A Fast Fault Diagnosis Method of the PEMFC System Based on Extreme Learning Machine and Dempster–Shafer Evidence Theory. IEEE Trans Transport Electrification 5(1):271–284
95. Liu J, Liu X, Liu C, Le BT, Xiao D (2019) Random Search Enhancement of Incremental Regularized Multiple Hidden Layers ELM. IEEE Access 7:36866–36878
96. Liu M, Liu B, Zhang C, Wang W, Sun W (2016) Semi-supervised low rank kernel learning algorithm via extreme learning machine. Int J Mach Learn Cybern 8(3):1039–1052
97. Liu N, Wang H (2010) Ensemble Based Extreme Learning Machine. IEEE Signal Process Lett 17(8):754–757
98. Liu P, Huang Y, Meng L, Gong S, Zhang G (2014) Two-stage extreme learning machine for high-dimensional data. Int J Mach Learn Cybern 7(5):765–772
99. Liu R, Wang Y, Zhou H, Qian Z (2019) Short-Term Passenger Flow Prediction Based on Wavelet Transform and Kernel Extreme Learning Machine. IEEE Access 7:158025–158034
100. Liu T, Lekamalage LCK, Huang G-B, Lin Z (2018) Extreme Learning Machine for Joint Embedding and Clustering. Neurocomputing 277:78–88
101. Liu T, Yang Y, Huang G, Yeo YK, Lin Z (2016) Driver Distraction Detection Using Semi-Supervised Machine Learning. IEEE Trans Intell Transp Syst 17(4):1108–1120
102. Liu X, Wan AH (2015) Universal consistency of extreme learning machine for RBFNs case. Neurocomputing 168:1132–1137
103. Liu X, Wang L, Huang G-B, Zhang J, Yin J (2013) Multiple kernel extreme learning machine. Neurocomputing 149:253–264
104. Liu XL, Liu LS, Wang LL, Guo Q, Peng XY (2019) Performance Sensing Data Prediction for an Aircraft Auxiliary Power Unit Using the Optimized Extreme Learning Machine. Sensors 19(18):21 Article ID: 3935
105. Liu Y, Li X, Wang J, Chen F, Wang J, Shi Y, Zheng L (2018) Pattern Discovery from Big Data of Food Sampling Inspections Based on Extreme Learning Machine. Springer International Publishing, Cham, pp 132–142
106. Liu Y, Loh HT, Tor SB (2005) Comparison of Extreme Learning Machine with Support Vector Machine for Text Classification. Springer Berlin Heidelberg, Berlin, pp 390–399
107. Liu Z-T, Li S-H, Cao W-H, Li D-Y, Hao M, Zhang R (2019) Combining 2D Gabor and Local Binary Pattern for Facial Expression Recognition Using Extreme Learning Machine. J Adv Computational Intell Intell Inform 23(3):444–455
108. Lu F, Wu JD, Huang JQ, Qiu XJ (2020) Restricted-Boltzmann-Based Extreme Learning Machine for Gas Path Fault Diagnosis of Turbofan Engine. Ieee Trans Indust Inform 16(2):959–968
109. Lu L, Zhang X, Xu X (2019) Hypercomplex extreme learning machine with its application in multispectral palmprint recognition. PLoS One 14(4):e0209083
110. Lu S, Lu Z, Yang J, Yang M, Wang S (2018) A pathological brain detection system based on kernel based ELM. Multimed Tools Appl 77(3):3715–3728
111. Malar E, Kandaswamy A, Chakravarthy D, Giri Dharan A (2012) A novel approach for detection and classification of mammographic microcalcifications using wavelet analysis and extreme learning machine. Comput Biol Med 42(9):898–905
112. Marjanović V, Milovančević M, Mladenović I (2016) Prediction of GDP growth rate based on carbon dioxide (CO_2) emissions. J CO2 Utilization 16:212–217
113. Marković D, Petković D, Nikolić V, Milovančević M, Petković B (2017) Soft computing prediction of economic growth based in science and technology factors. Physica A: Stat Mech Appl 465:217–220

114. McCulloch WS, Pitts W (1943) A logical calculus of the ideas immanent in nervous activity. Bull Mathematical Biophys 5(4):115–133
115. Mehrizi A, Yazdi HS (2016) Semi-supervised GSOM integrated with extreme learning machine. Intelligent Data Analysis 20(5):1115–1132
116. Milačić L, Jović S, Vujović T, Miljković J (2017) Application of artificial neural network with extreme learning machine for economic growth estimation. Physica A: Stat Mech Appl 465:285–288
117. Minemoto T, Isokawa T, Nishimura H, Matsui N (2017) Feed forward neural network with random quaternionic neurons. Signal Process 136:59–68
118. Mirza B, Lin Z, Toh KA (2013) Weighted Online Sequential Extreme Learning Machine for Class Imbalance Learning. Neural Process Lett 38:465–486
119. Nayak DR, Das D, Dash R, Majhi S, Majhi B (2020) Deep extreme learning machine with leaky rectified linear unit for multiclass classification of pathological brain images. Multimed Tools Appl 79(21-22): 15381–15396
120. Nayak DR, Dash R, Majhi B (2017) Pathological Brain Detection using Extreme Learning Machine Trained with Improved Whale Optimization Algorithm. Multimed Tools Appl 77(17):22705–22733
121. Naz A, Javed M, Javaid N, Saba T, Alhussein M, Aurangzeb K (2019) Short-Term Electric Load and Price Forecasting Using Enhanced Extreme Learning Machine Optimization in Smart Grids. Energies. 12(5): Article ID: 866
122. Nguyen DT, Ryu S, Qureshi MNI, Choi M, Lee KH, Lee B (2019) Hybrid multivariate pattern analysis combined with extreme learning machine for Alzheimer's dementia diagnosis using multi-measure rs-fMRI spatial patterns. PLoS One 14(2):e0212582
123. Niu M, Zhang J, Li Y, Wang C, Liu Z, Ding H, Zou Q, Ma Q (2020) CirRNAPL: A web server for the identification of circRNA based on extreme learning machine. Computational and Struct Biotechnol J 18: 834–842
124. Niu W-J, Feng Z-K, Feng B-F, Min Y-W, Cheng C-T, Zhou J-Z (2019) Comparison of Multiple Linear Regression, Artificial Neural Network, Extreme Learning Machine, and Support Vector Machine in Deriving Operation Rule of Hydropower Reservoir. Water. 11(1): Article ID: 88
125. Olatunji SO (2017) Extreme Learning machines and Support Vector Machines models for email spam detection. In: 2017 IEEE 30th Canadian Conference on Electrical and Computer Engineering (CCECE). IEEE, Windsor, pp 1–6
126. Oneto L, Fumeo E, Clerico G, Canepa R, Papa F, Dambra C, Mazzino N, Anguita D (2017) Dynamic Delay Predictions for Large-Scale Railway Networks: Deep and Shallow Extreme Learning Machines Tuned via Thresholdout. IEEE Trans Syst Man Cybernet: Syst 47(10):2754–2767
127. Ouyang C-S, Cheng Y-Y, Kao T-C, Pan S-T, Wu C-H, Lee S-J (2017) Recursive SVD-based Fuzzy Extreme Learning Machine. In: IEEE International Conference on Information and Automation (ICIA). IEEE, Macau, pp 466–471
128. Ouyang Q, Chen Q, Zhao J, Lin H (2013) Determination of Amino Acid Nitrogen in Soy Sauce Using Near Infrared Spectroscopy Combined with Characteristic Variables Selection and Extreme Learning Machine. Food Bioprocess Technol 6(9):2486–2493
129. Ouyang T, Wang C, Yu Z, Stach R, Mizaikoff B, Huang G, Wang Q (2020) NOx Measurements in Vehicle Exhaust Using Advanced Deep ELM Networks. IEEE Trans Instrum Meas. https://doi.org/10.1109/TIM.2020.3013129.(EarlyAccess)
130. Pacheco AGC, Krohling RA (2018) An approach to improve online sequential extreme learning machines using restricted Boltzmann machines. In: 2018 International Joint Conference on Neural Networks (IJCNN). IEEE, Rio de Janeiro, pp 1–8
131. Pacheco AGC, Krohling RA, da Silva CAS (2018) Restricted Boltzmann machine to determine the input weights for extreme learning machines. Expert Syst Appl 96:77–85
132. Pao YH, Park GH, Sobajic DJ (1994) Learning and generalization characteristics of random vector functional-link net. Neurocomputing 6:163–180
133. Pavelski LM, Delgado MR, Almeida CP, Gonçalves RA, Venske SM (2016) Extreme Learning Surrogate Models in Multi-objective Optimization based on Decomposition. Neurocomputing 180:55–67
134. Pavelski LM, Delgado MR, Almeida CP, Goncalves RA, Venske SM (2014) ELMOEA/D-DE: Extreme Learning Surrogate Models in Multi-objective Optimization Based on Decomposition and Differential Evolution. In: Brazilian Conference on Intelligent Systems. IEEE, Sao Paulo, pp 318–323
135. Pei G, Wang Y, Cheng Y, Zhang L (2019) Joint Label-Density-Margin Space and Extreme Elastic Net for Label-Specific Features. IEEE Access 7:112304–112317
136. Peng X, Lin P, Zhang T, Wang J (2013) Extreme learning machine-based classification of ADHD using brain structural MRI data. PLoS One 8(11):e79476
137. Peng Y, Zheng W-L, Lu B-L (2016) An unsupervised discriminative extreme learning machine and its applications to data clustering. Neurocomputing 174:250–264

138. Prates MO (2019) Spatial extreme learning machines: An application on prediction of disease counts. Stat Methods Med Res 28(9):2583–2594

139. Qin Y, Li M, De G, Huang L, Yang S, Tan Q, Tan Z, Zhou F (2019) Research on Green Management Effect Evaluation of Power Generation Enterprises in China Based on Dynamic Hesitation and Improved Extreme Learning Machine. Processes. 7(7): Article ID: 474

140. Qiu S, Wang J, Tang C, Du D (2015) Comparison of ELM, RF, and SVM on E-nose and E-tongue to trace the quality status of mandarin (Citrus unshiu Marc.). J Food Eng 166:193–203

141. Qiu Y, Wu G, Xiao Z, Guo Y, Zhang X, Liu K (2019) An Extreme-Learning-Machine-Based Hyperspectral Detection Method of Insulator Pollution Degree. IEEE Access 7:121156–121164

142. Qureshi MN, Min B, Jo HJ, Lee B (2016) Multiclass Classification for the Differential Diagnosis on the ADHD Subtypes Using Recursive Feature Elimination and Hierarchical Extreme Learning Machine: Structural MRI Study. PLoS One 11(8):e0160697

143. Qureshi MNI, Oh J, Cho D, Jo HJ, Lee B (2017) Multimodal Discrimination of Schizophrenia Using Hybrid Weighted Feature Concatenation of Brain Functional Connectivity and Anatomical Features with an Extreme Learning Machine. Front Neuroinform 11:59

144. Rajpal A, Mishra A, Bala R (2019) A Novel fuzzy frame selection based watermarking scheme for MPEG-4 videos using Bi-directional extreme learning machine. Appl Soft Comput 74:603–620

145. Rakic G, Milenkovic D, Vujovic S, Vujovic T, Jović S (2019) Information system for e-GDP based on computational intelligence approach. Physica A: Stat Mech Appl 513:418–423

146. Ramalho GLB, Filho RPP, Medeiros FS, Cortez PC (2014) Lung disease detection using feature extraction and extreme learning machine. Revista Brasileira de Engenharia Biomédica 30(3):207–214

147. Ramasamy S, Rajaraman K (2017) A Hybrid Meta-cognitive Restricted Boltzmann Machine Classifier for Credit Scoring. In: Tencon 2017 - 2017 Ieee Region 10 Conference. IEEE, New York, pp 2313–2318

148. Rathore S, Park JH (2018) Semi-supervised learning based distributed attack detection framework for IoT. Appl Soft Comput 72:79–89

149. Rodriguez N, Barba L, Alvarez P, Cabrera-Guerrero G (2019) Stationary Wavelet-Fourier Entropy and Kernel Extreme Learning for Bearing Multi-Fault Diagnosis. Entropy. 21(6): Article ID: 540

150. Rong H-J, Jia Y-X, Zhao G-S (2014) Aircraft recognition using modular extreme learning machine. Neurocomputing 128:166–174

151. Ronoud S, Asadi S (2019) An evolutionary deep belief network extreme learning-based for breast cancer diagnosis. Soft Comput 23(24):13139–13159

152. Rosenblatt F (1958) The perceptron: A probabilistic model for information storage and organization in the brain. Psychol Rev 65(6):386–408

153. Sánchez-Oro J, Duarte A, Salcedo-Sanz S (2016) Robust total energy demand estimation with a hybrid Variable Neighborhood Search – Extreme Learning Machine algorithm. Energy Convers Manag 123:445–452

154. Schmidt WF, Kraaijveld MA, Duin RPW (1992) Feedforward neural networks with random weights. In: 11th IAPR International Conference on Pattern Recognition. IEEE, The Hague, pp 1–4

155. Sharma J, Giri C, Granmo O-C, Goodwin M (2019) Multi-layer intrusion detection system with ExtraTrees feature selection, extreme learning machine ensemble, and softmax aggregation. EURASIP Journal on Information Security. 2019(1): Article ID: 15

156. She Q, Hu B, Gan H, Fan Y, Nguyen T, Potter T, Zhang Y (2018) Safe Semi-Supervised Extreme Learning Machine for EEG Signal Classification. IEEE Access 6:49399–49407

157. She Q, Hu B, Luo Z, Nguyen T, Zhang Y (2018) A hierarchical semi-supervised extreme learning machine method for EEG recognition. Med Biol Eng Comput 57(1):147–157

158. Shen C, Zhang S-F, Zhai J-H, Luo D-S, Chen J-F (2018) Imbalanced Data Classification Based on Extreme Learning Machine Autoencoder. In: International Conference on Machine Learning and Cybernetics (ICMLC). IEEE, Chengdu, pp 399–404

159. Shen Y, Xiao L, Chen J, Pan D (2019) A Spectral-Spatial Domain-Specific Convolutional Deep Extreme Learning Machine for Supervised Hyperspectral Image Classification. IEEE Access 7:132240–132252

160. Shoumo SZH, Dhruba MIM, Hossain S, Ghani NH, Arif H, Islam S (2019) Application of Machine Learning in Credit Risk Assessment: A Prelude to Smart Banking. In: TENCON 2019 - 2019 IEEE Region 10 Conference (TENCON). p. 2023-2028

161. Shukla AK, Kumar S, Jagdev R, Muhuri PK, Lohani QMD (2018) Interval Type-2 Fuzzy weighted Extreme Learning Machine for GDP Prediction. In: 2018 International Joint Conference on Neural Networks (IJCNN). p. 1-8

162. Singh S, Pareek M, Changotra A, Banerjee S, Bhaskararao B, Balamurugan P, Sunoj RB (2020) A unified machine-learning protocol for asymmetric catalysis as a proof of concept demonstration using asymmetric hydrogenation. Proc Natl Acad Sci USA 117(3):1339–1345

163. Sokolov-Mladenović S, Milovančević M, Mladenović I, Alizamir M (2016) Economic growth forecasting by artificial neural network with extreme learning machine based on trade, import and export parameters. Comput Hum Behav 65:43–45
164. Song J-L, Hu W, Zhang R (2016) Automated detection of epileptic EEGs using a novel fusion feature and extreme learning machine. Neurocomputing 175:383–391
165. Song T, Li D, Liu Z, Yang W (2019) Online ADMM-Based Extreme Learning Machine for Sparse Supervised Learning. IEEE Access 7:64533–64544
166. Song Y, Crowcroft J, Zhang J (2012) Automatic epileptic seizure detection in EEGs based on optimized sample entropy and extreme learning machine. J Neurosci Methods 210(2):132–146
167. Soria-Olivas E, Gomez-Sanchis J, Martin JD, Vila-Frances J, Martinez M, Magdalena JR, Serrano AJ (2011) BELM: Bayesian Extreme Learning Machine. IEEE Trans Neural Netw 22(3):505–509
168. Sun C, Yu Y, Liu H, Gu J (2015) Robotic grasp detection using extreme learning machine. In: International Conference on Robotics and Biomimetics (ROBIO). IEEE, Zhuhai, pp 1115–1120
169. Sun K, Zhang J, Zhang C, Hu J (2017) Generalized extreme learning machine autoencoder and a new deep neural network. Neurocomputing 230:374–381
170. Sun W, Duan M (2019) Analysis and Forecasting of the Carbon Price in China's Regional Carbon Markets Based on Fast Ensemble Empirical Mode Decomposition, Phase Space Reconstruction, and an Improved Extreme Learning Machine. Energies. 12(2): Article ID: 277
171. Sun W, Wang C, Zhang C (2017) Factor analysis and forecasting of CO_2 emissions in Hebei, using extreme learning machine based on particle swarm optimization. J Clean Prod 162:1095–1101
172. Sun W, Zhang C (2018) Analysis and forecasting of the carbon price using multi—resolution singular value decomposition and extreme learning machine optimized by adaptive whale optimization algorithm. Appl Energy 231:1354–1371
173. Sun Z, Ng KM, Soszynska-Budny J, Habibullah MS (2011) Application of the LP-ELM Model on Transportation System Lifetime Optimization. IEEE Trans Intell Transp Syst 12(4):1484–1494
174. Tang X, Chen L (2019) Artificial bee colony optimization-based weighted extreme learning machine for imbalanced data learning. Clust Comput 22:6937–6952
175. Termenon M, Graña M, Barrós-Loscertales A, Ávila C (2013) Extreme Learning Machines for Feature Selection and Classification of Cocaine Dependent Patients on Structural MRI Data. Neural Process Lett 38(3):375–387
176. Termenon M, Graña M, Savio A, Akusok A, Miche Y, Björk K-M, Lendasse A (2016) Brain MRI morphological patterns extraction tool based on Extreme Learning Machine and majority vote classification. Neurocomputing 174:344–351
177. Tian Y, Zhang J, Chen L, Geng Y, Wang X (2019) Selective Ensemble Based on Extreme Learning Machine for Sensor-Based Human Activity Recognition. Sensors (Basel). 19(16): Article ID: 3468
178. Vani G, Savitha R, Sundararajan N (2010) Classification of Abnormalities in Digitized Mammograms using Extreme Learning Machine. In: 11th Int. Conf. Control, Automation, Robotics and Vision. IEEE, Singapore, pp 2114–2117
179. Wang JN, Jin JL, Geng Y, Sun SL, Xu HL, Lu YH, Su ZM (2013) An accurate and efficient method to predict the electronic excitation energies of BODIPY fluorescent dyes. J Comput Chem 34(7):566–575
180. Wang L, You Z, Huang D, Zhou F (2020) Combining High Speed ELM Learning with a Deep Convolutional Neural Network Feature Encoding for Predicting Protein-RNA Interactions. IEEE/ACM Trans Computational Biol Bioinform 17(3):972–980
181. Wang M, Chen H, Li H, Cai Z, Zhao X, Tong C, Li J, Xu X (2017) Grey wolf optimization evolving kernel extreme learning machine: Application to bankruptcy prediction. Eng Appl Artif Intell 63:54–68
182. Wang R, Chow C, Lyu Y, Lee VCS, Kwong S, Li Y, Zeng J (2018) TaxiRec: Recommending Road Clusters to Taxi Drivers Using Ranking-Based Extreme Learning Machines. IEEE Trans Knowl Data Eng 30(3):585–598
183. Wang X, Zhang T, Wang R (2019) Noniterative Deep Learning: Incorporating Restricted Boltzmann Machine Into Multilayer Random Weight Neural Networks. IEEE Trans Syst Man Cybernet: Syst 49(7): 1299–1308
184. Wang Y, Cao F, Yuan Y (2011) A study on effectiveness of extreme learning machine. Neurocomputing 74(16):2483–2490
185. Wang Z, Li M, Wang H, Jiang H, Yao Y, Zhang H, Xin J (2019) Breast Cancer Detection Using Extreme Learning Machine Based on Feature Fusion With CNN Deep Features. IEEE Access 7:105146–105158
186. Wang Z, Qu Q, Yu G, Kang Y (2014) Breast tumor detection in double views mammography based on extreme learning machine. Neural Comput & Applic 27(1):227–240
187. Wang Z, Yu G, Kang Y, Zhao Y, Qu Q (2014) Breast tumor detection in digital mammography based on extreme learning machine. Neurocomputing 128:175–184

188. Wei X, Li C, Huang W, Gu J, Liu J, Zhu M (2019) Sediment-carrying capacity prediction using extreme learning machine. IOP Conference Series: Earth and Environ Sci 304(4): Article ID: 042006

189. Wei X, Liu H, She X, Lu Y, Liu X, Mo S (2019) Stability Assessment of Rubble Mound Breakwaters Using Extreme Learning Machine Models. Journal of Marine Science and Engineering. 7(9): Article ID: 312

190. Wen H, Fan HG, Xie WX, Pei JH (2017) Hybrid Structure-Adaptive RBF-ELM Network Classifier. IEEE Access 5:16539–16554

191. Werbos P, John P (1974) Beyond regression : new tools for prediction and analysis in the behavioral sciences. Ph. D. dissertation, Harvard University

192. Wong PK, Huang W, Vong CM, Yang Z (2020) Adaptive neural tracking control for automotive engine idle speed regulation using extreme learning machine. Neural Comput & Applic 32:14399–14409

193. Wu D, Qu Z, Guo F, Wan Q (2019) Multilayer Incremental Hybrid Cost-Sensitive Extreme Learning Machine With Multiple Hidden Output Matrix and Subnetwork Hidden Nodes. IEEE Access 7:118422–118434

194. Wu T, Yao M, Yang J (2017) Dolphin Swarm Extreme Learning Machine. Cogn Comput 9(2):275–284

195. Xia J, Chen H, Li Q, Zhou M, Chen L, Cai Z, Fang Y, Zhou H (2017) Ultrasound-based differentiation of malignant and benign thyroid Nodules: An extreme learning machine approach. Comput Methods Prog Biomed 147:37–49

196. Xia M, Wang J, Liu J, Weng L, Xu Y (2020) Density-based semi-supervised online sequential extreme learning machine. Neural Comput & Applic 32:7747–7758

197. Xie W, Li Y, Ma Y (2016) Breast mass classification in digital mammography based on extreme learning machine. Neurocomputing 173:930–941

198. Xin J, Wang Z, Qu L, Wang G (2015) Elastic extreme learning machine for big data classification. Neurocomputing 149:464–471

199. Xing Y, Ban X, Liu R (2018) A Short-Term Traffic Flow Prediction Method Based on Kernel Extreme Learning Machine. In: IEEE International Conference on Big Data and Smart Computing (BigComp). IEEE, Shanghai, pp 533–536

200. Xu C, Lei H, Li J, Ye J, Zhang D (2019) Adaptive Neural Control for Nonaffine Pure-Feedback System Based on Extreme Learning Machine. Math Probl Eng 2019:1–13

201. Xu XZ, Shan D, Li S, Sun TF, Xiao PC, Fan JP (2019) Multi-label learning method based on ML-RBF and laplacian ELM. Neurocomputing 331:213–219

202. Xu Z, Yao M, Wu Z, Dai W (2016) Incremental regularized extreme learning machine and it′s enhancement. Neurocomputing 174:134–142

203. Yan JY (2017) RBFL embedded network data algorithm based on machine learning and radial basis function. Agro Food Industry Hi-Tech 28(1):2321–2325

204. Yang R, Xu S, Feng L (2018) An Ensemble Extreme Learning Machine for Data Stream Classification. Algorithms 11(7):107

205. Yang X-c, X-r Y, C-f S (2015) Pressure Prediction of Coal Slurry Transportation Pipeline Based on Particle Swarm Optimization Kernel Function Extreme Learning Machine. Math Probl Eng 2015:1–7

206. Yi Y, Dai J, Wang C, Hou J, Zhang H, Liu Y, Gao J (2019) An Effective Framework Using Spatial Correlation and Extreme Learning Machine for Moving Cast Shadow Detection. Appl Sci 9(23): Article ID: 5042

207. Yin Y, Dong X, Xu T (2018) Rapid and Efficient Bug Assignment Using ELM for IOT Software. IEEE Access 6:52713–52724

208. You Z, Zhou M, Luo X, Li S (2017) Highly Efficient Framework for Predicting Interactions Between Proteins. IEEE Trans Cybernet 47(3):731–743

209. Yousefi-Azar M, McDonnell MD (2017) Semi-supervised Convolutional Extreme Learning Machine. International Joint Conference on Neural Networks (IJCNN): 1968-1974

210. Yu J, Song W, Zhou G, J-j H (2018) Violent scene detection algorithm based on kernel extreme learning machine and three-dimensional histograms of gradient orientation. Multimed Tools Appl 78(7):8497–8512

211. Yu W, Zhuang F, He Q, Shi Z (2015) Learning deep representations via extreme learning machines. Neurocomputing 149:308–315

212. Yuan Q, Zhou W, Li S, Cai D (2011) Epileptic EEG classification based on extreme learning machine and nonlinear features. Epilepsy Res 96(1):29–38

213. Zeng N, Zhang H, Liu W, Liang J, Alsaadi FE (2017) A switching delayed PSO optimized extreme learning machine for short-term load forecasting. Neurocomputing 240:175–182

214. Zeng Y, Xu X, Shen D, Fang Y, Xiao Z (2017) Traffic Sign Recognition Using Kernel Extreme Learning Machines With Deep Perceptual Features. IEEE Trans Intell Transp Syst 18(6):1647–1653

215. Zhang, Wei (2019) A Hybrid Model Based on Principal Component Analysis, Wavelet Transform, and Extreme Learning Machine Optimized by Bat Algorithm for Daily Solar Radiation Forecasting. Sustainability. 11(15): Article ID: 4138

216. Zhang J, Ding S, Zhang N, Shi Z (2015) Incremental extreme learning machine based on deep feature embedded. Int J Mach Learn Cybern 7(1):111–120

217. Zhang J, Xiao WD, Li YJ, Zhang S, Zhang ZQ (2020) Multilayer probability extreme learning machine for device-free localization. Neurocomputing 396:383–393

218. Zhang L, He Z, Liu Y (2017) Deep object recognition across domains based on adaptive extreme learning machine. Neurocomputing 239:194–203

219. Zhang L, Zhang D (2015) Domain Adaptation Extreme Learning Machines for Drift Compensation in E-Nose Systems. IEEE Trans Instrum Meas 64(7):1790–1801

220. Zhang L, Zhang D (2015) Domain Adaptation Transfer Extreme Learning Machines. Springer International Publishing, Cham, pp 103–119

221. Zhang L, Zhang D (2016) Robust Visual Knowledge Transfer via Extreme Learning Machine-Based Domain Adaptation. IEEE Trans Image Process 25(10):4959–4973

222. Zhang L, Zhang D, Tian F (2016) SVM and ELM: Who Wins? Object Recognition with Deep Convolutional Features from ImageNet. Springer International Publishing, Cham, pp 249–263

223. Zhang M, Liu X, Zhang Z (2016) A soft sensor for industrial melt index prediction based on evolutionary extreme learning machine. Chin J Chem Eng 24(8):1013–1019

224. Zhang N, Ding S (2016) Unsupervised and semi-supervised extreme learning machine with wavelet kernel for high dimensional data. Memetic Computing 9(2):129–139

225. Zhang N, Qu Y, Deng A (2015) Evolutionary Extreme Learning Machine Based Weighted Nearest-Neighbor Equality Classification. 7th International Conference on Intelligent Human-Machine Systems and Cybernetics: 274-279

226. Zhang P, Yang Z (2015) A Robust AdaBoost.RT Based Ensemble Extreme Learning Machine. Math Probl Eng 2015:1–12

227. Zhang Q, Li H, Liu C, Hu W (2013) A New Extreme Learning Machine Optimized by Firefly Algorithm. Sixth International Symposium on Computational Intelligence and Design: 133-136

228. Zhang R, Zhou L, Zuo M, Zhang Q, Bi M, Jin Q, Xu Z (2018) Prediction of Dairy Product Quality Risk Based on Extreme Learning Machine. In: 2018 2nd International Conference on Data Science and Business Analytics (ICDSBA). p. 448-456

229. Zhang X, He T, Lu L, Yue S, Cheng D, Xu X (2017) Video analysis of traffic accidents based on projection extreme learning machine. In: 2017 International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS). p. 149-154

230. Zhang X, Ma H (2019) Data-Driven Model-Free Adaptive Control Based on Error Minimized Regularized Online Sequential Extreme Learning Machine. Energies. 12(17): Article ID: 3241

231. Zhang Y-D, Zhao G, Sun J, Wu X, Wang Z-H, Liu H-M, Govindaraj VV, Zhan T, Li J (2017) Smart pathological brain detection by synthetic minority oversampling technique, extreme learning machine, and Jaya algorithm. Multimed Tools Appl 77(17):22629–22648

232. Zhang Y, Jiang X, Wang X, Cai Z (2019) Spectral-Spatial Hyperspectral Image Classification with Superpixel Pattern and Extreme Learning Machine. Remote Sens 11(17): Article ID: 1983

233. Zhang Y, Wu J, Cai Z, Zhang P, Chen L (2016) Memetic Extreme Learning Machine. Pattern Recogn 58: 135–148

234. Zhang Y, Wu J, Zhou C, Cai Z (2017) Instance cloned extreme learning machine. Pattern Recogn 68:52–65

235. Zhang Y, Wu J, Zhou C, Cai Z, Yang J, Yu PS (2019) Multi-View Fusion with Extreme Learning Machine for Clustering. ACM Trans Intell Syst Technol 10(5):1–23

236. Zhao J, Jiao L (2019) Sparse Deep Tensor Extreme Learning Machine for Pattern Classification. IEEE Access 7:119181–119191

237. Zhao J, Wang Z, Park DS (2012) Online sequential extreme learning machine with forgetting mechanism. Neurocomputing 87:79–89

238. Zheng D, Hong Z, Wang N, Chen P (2020) An Improved LDA-Based ELM Classification for Intrusion Detection Algorithm in IoT Application. Sensors (Basel, Switzerland) 20(6):1706

239. Zheng L, Wang Z, Zhao Z, Wang J, Du W (2019) Research of Bearing Fault Diagnosis Method Based on Multi-Layer Extreme Learning Machine Optimized by Novel Ant Lion Algorithm. IEEE Access 7:89845–89856

240. Zhong H, Miao C, Shen Z, Feng Y (2014) Comparing the learning effectiveness of BP, ELM, I-ELM, and SVM for corporate credit ratings. Neurocomputing 128:285–295

241. Zhou T, Tang L, Hang H, Yang P (2018) Classification of Epileptic Electroencephalograms Signal Based on Improved Extreme Learning Machine. J Med Imaging Health Inform 8(1):33–37

242. Zhou Z, Wang C, Gao X, Zhu Z, Hu X, Zheng X, Jiang L (2019) Fabric Defect Detection and Classifier via Multi-Scale Dictionary Learning and an Adaptive Differential Evolution Optimized Regularization Extreme Learning Machine. Fibres Textiles Eastern Europe 27(1(133)):67–77

243. Zhu W, Huang W, Lin Z, Yang Y, Huang S, Zhou J (2015) Data and feature mixed ensemble based extreme learning machine for medical object detection and segmentation. Multimed Tools Appl 75(5): 2815–2837

244. Zhu W, Miao J, Qing L, Huang G-B (2015) Hierarchical Extreme Learning Machine for Unsupervised Representation Learning. In: International Joint Conference on Neural Networks (IJCNN). IEEE, Killarney, pp 12–17

245. Zhu WT, Miao J, Qing LY (2014) Constrained Extreme Learning Machine: a Novel Highly Discriminative Random Feedforward Neural Network, in International Joint Conference on Neural Networks. IEEE, New York, pp 800–807

246. Zong W, Huang G-B, Chen Y (2013) Weighted extreme learning machine for imbalance learning. Neurocomputing 101:229–242

247. Zou W, Yao F, Zhang B, He C, Guan Z (2017) Verification and predicting temperature and humidity in a solar greenhouse based on convex bidirectional extreme learning machine algorithm. Neurocomputing 249:72–85

Affiliations

Jian Wang [1] · Siyuan Lu [1] · Shui-Hua Wang [2,3,4,5] · Yu-Dong Zhang [1,2,4]

[1]   School of Informatics, University of Leicester, Leicester LE1 7RH, UK

[2]   School of Computer Science and Technology, Henan Polytechnic University, Jiaozuo, Henan 454000, People's Republic of China

[3]   Department of Cardiovascular Sciences, University of Leicester, Leicester LE1 7RH, UK

[4]   Department of Information Systems, Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah 21589, Saudi Arabia

[5]   School of Architecture Building and Civil engineering, Loughborough University, Loughborough LE11 3TU, UK