International Journal of Applied Engineering Research ISSN 0973-4562 Volume 12, Number 14 (2017) pp. 4610-4623

Extreme Learning Machine: A Review

Musatafa Abbas Abbood Albadr[a] and Sabrina Tiun[a]

[a] *CAIT, Faculty of Information Science and Technology, Universiti Kebangsaan Malaysia, Bangi, Selangor, Malaysia.*

Abstract

Feedforward neural networks (FFNN) have been utilised for various research in machine learning and they have gained a significantly wide acceptance. However, it was recently noted that the feedforward neural network has been functioning slower than needed. As a result, it has created critical bottlenecks among its applications. Extreme Learning Machines (ELM) were suggested as alternative learning algorithms instead of FFNN. The former is characterised by single-hidden layer feedforward neural networks (SLFN). It selects hidden nodes randomly and analytically determines their output weight. This review aims to, first, present a short mathematical explanation to explain the basic ELM. Second, because of its notable simplicity, efficiency, and remarkable generalisation performance, ELM has had wide uses in various domains, such as computer vision, biomedical engineering, control and robotics, system identification, etc. Thus, in this review, we will aim to present a complete view of these ELM advances for different applications. Finally, ELM's strengths and weakness will be presented, along with its future perspectives.

Keywords: Extreme Learning Machine, Single-Hidden Layer Feedforward Neural Networks.

INTRODUCTION

Ever since the popular backpropagation (BP) algorithm has been introduced, feedforward neural networks (FNN) have been studied well and used widely (Rumelhart et al. 1988). Traditional BP algorithm is considered a first order gradient method that can be used to optimise parameters. However, it suffers from local minimum problem and slow convergence. Researchers have suggested different techniques to improve the optimality or efficiency in FNN training, such as subset selection methods (Chen et al. 1991; Li et al. 2005), second order optimisation methods (Hagan & Menhaj 1994; Wilamowski & Yu 2010), or global optimisation methods (Yao 1993; Branke 1995). Despite the fact that it exhibits better generalisation performance or faster training speed compared to the BP algorithm, majority of these methods are still not capable of guaranteeing a global optimal solution.

It has been recently proposed that Extreme Learning Machines (ELM) can be used to train single hidden layer feedforward neural networks (SLFNs). In ELM, initiation of the hidden nodes is done randomly and before it is fixed without iterative tuning. Furthermore, ELM's hidden nodes do not even need to be neuron alike. The free parameter that it has to learn is the connections (or weights) between the output layer and the hidden layer. As such, ELM is developed as a linear-in-the-parameter model that is ultimately concerned with solving a linear system. Unlike traditional FNN learning methods, ELM is significantly more efficient and it has a greater tendency to achieve a global optimum. It has been shown by theoretical studies that ELM is capable of maintaining the SLFNs' universal approximation capability even if it works with randomly generated hidden nodes (Huang et al. 2006; Huang & Chen 2007; Huang & Chen 2008). With frequently utilised activation functions, ELM is capable of achieving the traditional FNN's almost optimal generalisation bound, where it learns all the parameters (Liu et al. 2015). ELM's advantages over traditional FNN algorithms in generalisation and efficiency performance have been observed on a vast range of problems from various fields(Huang et al. 2006; Huang et al. 2012). ELM has been observed to generally have more efficiency compared to least square support vector machines (LS-SVMs) (Suykens & Vandewalle 1999), support vector machines (SVMs) (Cortes & Vapnik 1995), and other advanced algorithms. Empirical studies revealed that ELM's generalisation ability is comparable or even superior to that of SVMs and SVMs' variants (Huang et al. 2006; Huang et al. 2012; Fernández-Delgado et al. 2014; Huang et al. 2014). ELM and SVM were compared in detail in (Huang 2014) and (Huang et al. 2012). In the past decade, ELM applications and theories have been investigated extensively. From a learning efficiency standpoint, ELM's original design has three objectives: high learning accuracy, least human invention, and fast learning speed (as demonstrated in Fig. 1). The original ELM model has been equipped with various extensions to make it more suitable and efficient for specific applications. The authors of (Huang et al. 2015) wrote a review paper that did a wide-ranging study on ELM. Also, (Huang et al. 2011) gave a literature survey about ELM's applications and theories. Since then, there has been more active research on ELM. From a theoretical point of view, ELM's universal approximation capability was investigated further in (Huang et al. 2012). ELM's generalisation ability was studied using the framework from the initial localized generalisation error model (LGEM) (Xi-Zhao et al. 2013) and statistical learning theory (Liu et al. 2012; Lin et al. 2015; Liu et al. 2015). Many ELM variants have been suggested to meet specific application requirements. For example, the test time must be minimized in cost sensitive learning, which needs a compact network so that it can satisfy the test time budget. In this context, ELM was able to successfully adapt in order to attain high compactness in terms of network size (Deng et al. 2011; He et al. 2011; MartíNez-MartíNez et al. 2011; Yang et al. 2012; Du et al. 2013; Lahoz et al. 2013; Li et al. 2013; Yang et al. 2013; Bai et al. 2014; Wang et al. 2014). ELM extensions for noisy/missing data (Miche et al. 2010; Man et al. 2011; Horata et al. 2013; Yu et al. 2013), online sequential data (Liang et al. 2006; Lan et al. 2009; Rong et al. 2009; Zhao et al. 2012; Ye et al. 2013), imbalanced data (Horata et al. 2013; Zong et al. 2013; Huang et al. 2014), etc. have also been observed. Furthermore, apart from its uses in regression and traditional classification tasks, ELM's applications have recently been extended to feature selection, clustering, and representational learning (Benoít et al. 2013; Kasun et al. 2013; Huang et al. 2014). This review will first give a short mathematical explanation for the basic ELM. Next, it will present a roadmap for the newest optimisations on ELM and ELM's applications. Finally, ELM's strengths and weaknesses will be presented.

It should be noted that the randomised strategies of ELM learning frameworks for nonlinear feature construction have attracted a large amount of interest in the machine learning and computational intelligence community (Rahimi & Recht 2008; Rahimi & Recht 2008; Rahimi & Recht 2009; Saxe et al. 2011; Le et al. 2013; Widrow et al. 2013). These approaches have a close relationship to ELM and a number of them can be considered special cases since they have many common properties. For instance, (Rahimi & Recht 2009)introduced the Random Kitchen Sinks (RKS), which is a special kind of ELM that restricts the construction of its hidden layer to the Fourier basis. The No-Prop algorithm demonstrated in (Widrow et al. 2013) possesses a spirit similar to that of ELM. However, the former trains its output weights with the use of the Least Mean Square (LMS) method.

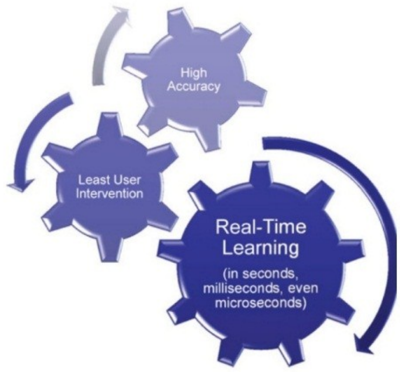

Figure 1: Learning targets of ELM framework (Huang et al. 2015)

The succeeding parts of this review will be organised into the following sections: In Section 2, the formulation of classical ELM will be introduced. Sections 3, 4, will offer an intensive review of ELM's extensions and improvements for its various applications. Section 5 will present ELM's strengths and weakness, along with its future perspectives. Section 6 presents the conclusion of the paper.

CLASSICAL EXTREME LEARNING MACHINES

This section will introduce the classical ELM model along with its basic variants for supervised regression and classification (Huang et al. 2004; Huang et al. 2006; Huang et al. 2011; Huang et al. 2012).The feature mappings, hidden nodes, and feature space of ELM were suggested for use in ‘‘generalised'' single-hidden layer feedforward networks. In these networks, the hidden layer does not have to be neuron alike (Huang & Chen 2007; Huang & Chen 2008; Huang et al. 2012). ELM's output function for generalised SLFNs is represented by the following equation

$$f_L(\mathrm{x}) = \sum_{i=1}^{L} \beta_i \, h_i(\mathrm{x}) = h(\mathrm{x})\beta \qquad (1)$$

Where $\beta=[\beta_1, \ldots, \beta_L]^T$is represents the output weight vector between the L nodes' hidden layer to the m ≥ 1 output nodes, and $h(\mathrm{x}) = [h1(\mathrm{x}), \ldots, hL(\mathrm{x})]$ represents the nonlinear feature mapping of ELM (Fig. 2), e.g., the hidden layer's output (row) vector in terms of the input x. $h_i(x)$ is the *ith* hidden node's output. The hidden nodes' output functions are not always unique. One can use different output functions in different hidden neurons. Specifically, in real applications, $h_i(x)$ can be presented as

$$h_i(\mathrm{x}) = G(\mathbf{a}_i, b_i, \mathrm{x}), \mathbf{a}_i \in \mathbf{R}^d, b_i \in R \qquad (2)$$

Where G (a, b, x) (having hidden node parameters (a, b)) represents a nonlinear piecewise continuous function that meets the capability theorems of ELM universal approximation (Huang et al. 2006; Huang & Chen 2007; Huang & Chen 2008).

International Journal of Applied Engineering Research ISSN 0973-4562 Volume 12, Number 14 (2017) pp. 4610-4623

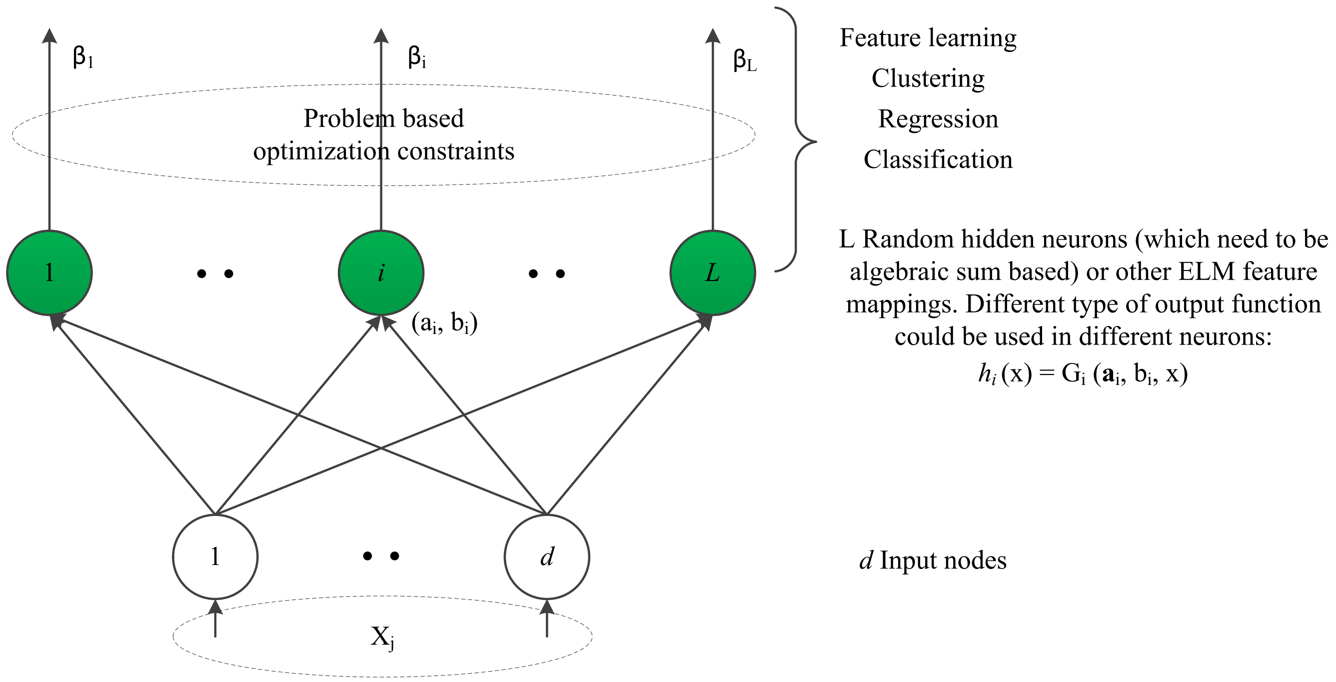

Figure 2: ELM architecture; the hidden nodes in ELM can be combinatorial nodes that are made up of different types of computational nodes (Huang & Chen 2007)

Table 1: Commonly used mapping functions in ELM

Sigmoid function	$\mathbf{G}(\mathbf{a},\mathbf{b},\mathbf{x}) = \dfrac{\mathbf{1}}{\mathbf{1}+\mathbf{exp}(-(\mathbf{a}\cdot\mathbf{x}+\mathbf{b}))}$
Hyperbolic tangent function	$G(a,b,x) = \dfrac{1-\exp(-(a\cdot x+b))}{1+\exp(-(a\cdot x+b))}$
Gaussian function	$G(a,b,x) = \exp(-b \parallel x - a \parallel)$
Multiquadric function	$G(a,b,x) = (\parallel x - a \parallel + b2)^{1/2}$
Hard limit function	$G(a,b,x) = \begin{cases} 1, & \text{if } \mathbf{a}\cdot x + b \leq 0 \\ 0, & \text{otherwise} \end{cases}$
Cosine function/Fourier basis	$G(a,b,x) = \cos(a \cdot x + b)$

Essentially, there are two main stages involved when ELM trains an SLFN: (1) randomised feature mapping and (2) solving of linear parameters. During the first stage, the hidden layer is randomly initialised by the ELM so that the input data can be mapped into a feature space (known as the ELM feature space) using some of the nonlinear mapping functions (see Fig. 3). This first stage differentiates ELM from numerous existing learning algorithms like SVM, which utilizes kernel functions to map features, or deep neural networks (Bengio 2009), which utilise Auto-Encoders/Auto-Decoders or Restricted Boltzmann Machines (RBM) for feature learning. In ELM, the nonlinear mapping functions can be any of the nonlinear piecewise continuous functions. Table 1 shows some of the most commonly used ones.

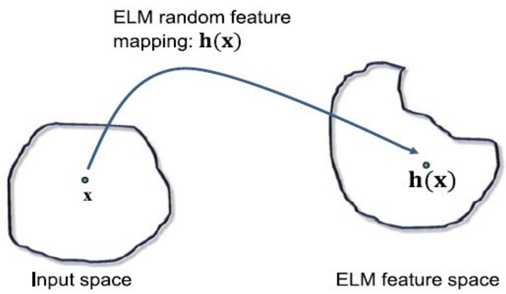

Figure 3: ELM feature mappings and feature space (Huang et al. 2015)

Instead of being explicitly trained, ELM randomly generates (does not depend on the training data) the hidden node parameters (a, b) based on any of the continuous probability distribution. This results into notable efficiency in comparison to traditional BP neural networks. Aside from the activation functions presented in Table 1, other special mapping functions are also utilized in ELM and its variants, like those utilized in wavelet ELM (Cao et al. 2010; Malathi et al. 2010; Malathi et al. 2011; Avci & Coteli 2012) and fuzzy ELM (Qu et al. 2011; Daliri 2012; Zhang & Ji 2013).

Basic Extreme Learning Machine

(Huang et al. 2006) proposed the original ELM algorithm that can be used to train SLFN. In ELM, the main idea involves the hidden layer weights. Furthermore, the biases are randomly generated and the calculation of the output weights is done using the least-squares solution. Furthermore, they have been defined by the outputs of the targets and the hidden layer. Figure 4 shows an overview of the training algorithm and the ELM structure. A brief description of ELM will be given in the next section.

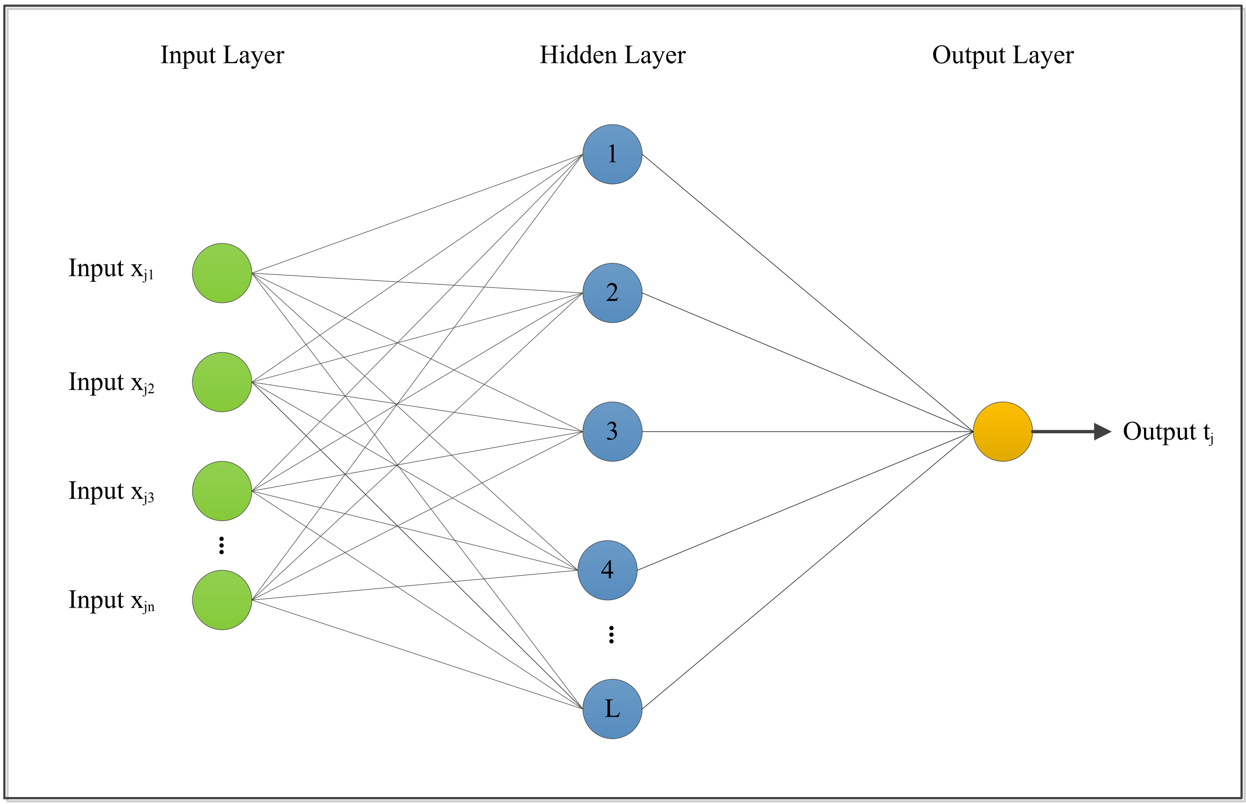

Figure 4: Diagram of Extreme Learning Machine (Huang et al. 2011)

Where:

N refers to a set of unique samples (X_i, t_i), where $X_i = [x_{i1}, x_{i2} \ldots x_{in}]^T \in R^n$ and $t_i = [t_{i1}, t_{i2} \ldots t_{im}]^T \in R^m$.

L represents the hidden layer nodes.

g(x) represents the activation function, which is also a mathematical model that is represented by the following equation:

$$\sum_{i=1}^{L} \beta_i \, g_i(X_j) = \sum_{i=1}^{L} \beta_i \, g_i(W_i . X_j + b_i) \qquad (3)$$

J = 1… N.

Where:

$W_i = [W_{i1}, W_{i2 \ldots} W_{in}]^T$ represents the weight vector that connects the hidden node and the *ith* input nodes.

$\beta i = [\beta i1, \beta i2, \ldots, \beta im]^T$ represents the weight vector that connects the hidden node and the *ith* output nodes.

b_i represents the *ith* hidden node's threshold.

$W_i . X_j$ represents the inner product of W_i and X_j. Selection of the output nodes is linearly done, however.

The standard of SLFNs and L hidden nodes in the activation function g(x) can be taken as samples of N without error. In other words, mean: $\sum_{j=1}^{L} || o_j - t_j || = 0$, i.e., and there exist $\beta i, W_i,$ and b_i in such a way that

$$\sum_{i=1}^{L} \beta_i \, g_i(W_i . X_j + b_i) = t_j, \quad j = 1, \ldots, N. \qquad (4)$$

From the equations given above for N, it can then be presented as follows:

$$H \beta = T \qquad (5)$$

Where:

H ($W_1 \ldots W_L, b_1 \ldots b_L, X_1 \ldots X_N$)

$$= \begin{bmatrix} g(W_1 . X_1 + b_1) & \cdots & g(W_L . X_1 + b_L) \\ \vdots & \cdots & \vdots \\ g(W_1 . X_N + b_1) & \cdots & g(W_L . X_N + b_L) \end{bmatrix}$$

$$\beta = \begin{bmatrix} \beta_1^T \\ \\ \beta_L^T \end{bmatrix}_{L*m} \text{ and } \mathrm{T} = \begin{bmatrix} t_1^T \\ . \\ t_N^T \end{bmatrix}_{N*m}$$

Equation (5) then turns into a linear system. Furthermore, the output weights β can be determined analytically by discovering a least square solution in the following way:

$\beta = H^{\dagger}\mathrm{T}$

Where $H^{\dagger}$is represents the Moore–Penrose generalised inverse for H. Thus, the output weights are calculated via a mathematical transformation. This makes sure that the lengthy training phrase when network parameters are iteratively adjusted with some suitable learning parameters (like iterations and learning rate) is done away with.

Huang et al. (2006) enumerated the variables, where H represents the output matrix of the neural network's hidden layer; in H, the ith column is used to describe the ith hidden layer nodes in terms of the input nodes. If L ≤ N represents the desired number of hidden nodes, the activation function g becomes infinitely differentiable.

VARIANTS OF ELM

This section summarises and briefly introduces several typical variants of ELM.

Incremental ELM

(Huang et al. 2006) developed an Incremental Extreme Learning Machine (I-ELM) to create a feedforward network that is incremental. I-ELM added nodes randomly to the hidden layer. This addition was done one by one. It then froze the existing hidden nodes' output weights during the addition of a new hidden node. I-ELM is efficient not only for SLFN having continuous activation functions (as well as differentiable), but they are also efficient for SLFNs that have piecewise continuous activation functions (like threshold). Given this context of I-ELM, Huang et al. presented the convex I-ELM (CI-ELM) and enhance I-ELM (EI-ELM). Unlike I-ELM, CI-ELM (Huang & Chen 2007) recalculates the existing hidden nodes' output weights when a new hidden node is added. Compared to I-ELM, CI-ELM is able to achieve more compact network architectures and faster convergence rates while retaining the efficiency and simplicity of I-ELM. EI-ELM (Huang & Chen 2008) allows for the maximum amount of hidden nodes. Furthermore, users need not set control parameters manually. Unlike the original I-ELM, EI-ELM chooses the optimal hidden node. This means that the smallest residual error is obtained at every learning step among several of the hidden nodes that were randomly generated. EI-ELM is able to achieve a much more compact network architecture and faster convergence rate. Moreover, (Huang et al. 2008) proposed an improved I-ELM that possessed fully complex hidden nodes. This improvement extended the I-ELM from its real domain and all the way to the complex domain.

Pruning ELM

The use of too few/many hidden nodes may result into issues of underfitting/overfitting in pattern classification.(Rong et al. 2008) developed a pruned-ELM (P-ELM) algorithm that provides an automated and systematic way to design an ELM network. P-ELM starts with a large amount of hidden nodes before it gets rid of the lowly relevant or irrelevant hidden nodes by taking into account their relevance to the class labels in the learning process. As a result, one can automate the architectural design of ELM. Simulation results revealed that the P-ELM resulted in compact network classifiers that are capable of generating robust prediction accuracy and fast response on unseen data compared to the standard BP, ELM, and MRAN. P-ELM is mostly suitable for pattern classification problems. Given the fact that too few/many hidden nodes can result into issues of underfitting/overfitting in pattern classification, Rong et al. (2008) proposed a pruned-ELM (P-ELM) algorithm. This algorithm is a systematic and automated way to design an ELM network. P-ELM starts with a large amount of hidden nodes before it gets rid of the lowly relevant or irrelevant hidden nodes by taking into account their relevance to the class labels in the learning process. As a result, one can automate the architectural design of ELM. Simulation results revealed that the P-ELM resulted in compact network classifiers that are capable of generating robust prediction accuracy and fast response on unseen data compared to the standard BP, ELM, and MRAN. P-ELM is mostly suitable for pattern classification problems.

Error-minimised ELM

(Feng et al. 2009) developed an error-minimisation-based method that can be used for ELM (EM-ELM). This method is able to grow hidden nodes group by group or one by one and automatically know the amount of hidden nodes that can be found in generalised SLFNs. During network growth, updating of the output weights is done incrementally, which pointedly lowers the computational complexity. For sigmoid type hidden nodes, the simulation results revealed that this technique could significantly lower ELM's computational complexity and help formulate an efficient ELM implementation.

Two-stage ELM

To achieve a parsimonious solution for the preliminary ELM's network structure, (Lan et al. 2010) proposed a systematic two-stage algorithm (named TS-ELM). During the first stage, they applied a forward recursive algorithm to choose the hidden nodes from the randomly generated candidates in each step. These hidden nodes were then added to the network until they met the stopping criterion. Consequently, each hidden node's significance was determined by their net contribution after being added to the network. During the second stage, a review of the selected hidden nodes was done in order to get rid of the unimportant nodes in the network. This step significantly reduced the complexity of the network. The empirical studies conducted on the six cases revealed that TS-ELM having a significantly smaller network structure may be able to achieve similar or better performance than EM-ELM.

Online sequential ELM

When using the conventional ELM, all of the training data must be available for training purposes. However, the training data in real applications may be obtained one by one or chunk by chunk. (Liang et al. 2006) proposed a sequential learning algorithm called the online sequential extreme learning machine (OS-ELM). This algorithm is able to work with both RBF and additive nodes in a unified framework. OSELM that has additive nodes randomly generates the input weights that connect the hidden nodes and biases to the input nodes. It then analytically determines the output weights based on the hidden nodes' output. Unlike the other kinds of sequential learning algorithms, the OS-ELM only needs the specification for the number of hidden nodes. This is also similar to the conventional ELM. To enhance the OSELM's performance and acquaint the ensemble networks with the sequential learning mode, (Lan et al. 2009) developed an integrated network structure that is referred to as the ensemble of online sequential extreme learning machine (EOS-ELM). EOS-ELM is made up of several OS-ELM networks. For these networks, the network performance's final measurement is computed based on the average value of the outputs for every OS-ELM in the ensemble. Furthermore, to show the training data's timeliness in the learning process,(Zhao et al. 2012) developed an improved EOS-ELM called the online sequential extreme learning machine with forgetting mechanism (FOS-ELM). This algorithm can retain EOS-ELM's advantages and enhance its learning effects by quickly getting rid of the out-dated data during the learning process in order to lessen their bad affection to the next learning process.

Evolutionary ELM

Typically, the amount of hidden neurons is randomly determined during the application of ELM. However, ELM may require higher amounts of hidden neurons as a result of the random determination of the hidden biases and input weights. (Zhu et al. 2005) proposed a novel learning algorithm called the evolutionary extreme learning machine (E-ELM) to optimize the hidden biases and input weights and determine the weights of the output. In E-ELM, the input weights and hidden biases were optimised using the modified differential evolutionary (DE) algorithm. The output weights were analytically determined using the Moore– Penrose (MP) generalised inverse. Experimental results revealed that E-ELM was capable of achieving good generalisation performance that has more compact networks and which is superior to other algorithms like GALS, BP, and the original ELM.

Voting-based ELM

Since the hidden nodes in ELM's learning parameters are randomly assigned and stay the same during the training process, ELM may not achieve the optimal classification boundary. As such, the samples nearest to the classification boundary run the risk of being misclassified. Therefore, Cao et al. (2012) developed an improved algorithm referred to as the voting-based extreme learning machine (V-ELM). The aim of this algorithm is to lessen the amount of misclassified samples that are found near the classification boundary. In V-ELM, the main idea is to conduct multiple independent ELM trainings rather than just performing a single ELM training and taking a final decision that is based on the results of the majority voting method (Cao et al. 2012). V-ELM was able to improve the classification performance, lessen the amount of misclassified samples, and reduce variance among the different realisations. Based on the simulations conducted on numerous real-world classification datasets, it was observed that V-ELM generally performed better than the original ELM algorithm and even other recent classification algorithms.

Ordinal ELM

To study the ELM algorithm further for ordinal regression problems, (Deng et al. 2010) introduced three ELM-based ordinal regression algorithms and an encoding-based ordinal regression framework. The paper developed an encoding-based framework that can be used for ordinal regression and that contained three encoding schemes: multiple binary classifications having a one-against-all decomposition method, single multi-output classifier, and one-against-one method. The framework was used as the basis for the redesigning of the SLFN for ordinal regression problems. Extreme learning machine was then used to train the algorithms. Experiments conducted on the three types of datasets revealed that ordinal ELM is capable of achieving good generalisation ability and extremely rapid training speed.

Fully complex ELM

To extend the ELM algorithm's application,(Li et al. 2005) developed a fully complex extreme learning algorithm called the C-ELM. In this algorithm, the ELM algorithm's reach was extended from being in the real domain all the way to the complex domain. Like ELM, the hidden layer biases and input weights of C-ELM were selected randomly on the basis of some continuous distribution probability. Afterwards, the output weights were computed analytically instead of it turning iteratively. Then, C-ELM was utilised to equalise a complex nonlinear channel using QAM signals.

Symmetric ELM

(Liu et al. 2013) proposed a modified ELM algorithm known as the symmetric ELM (S-ELM). This algorithm turned the hidden neurons' original activation function into a symmetric function in terms of the samples' input variables. Theoretically, S-ELM is capable of preserving the capacity to approximate N arbitrary distinct samples with any errors. It was shown by the simulation results that S-ELM can obtain faster learning speed, better generalisation performance, and more compact network architecture by using the prior knowledge for symmetry.

EXPERIMENTAL STUDY OF ELM AND APPLICATIONS

Comprehensive empirical studies have been performed about the performance of ELM and ELM's variants. In the past years, it has also been compared with other advanced learning algorithms. Typically, ELM is easy in implementation, fast and stable in training, and accurate in prediction and modelling.

Comparison with SVM and its variant

A detailed empirical study about the generalisation performance and training efficiency of ELM regression and classification was given in Huang, Zhou, et al. (2012). Comparisons were made with LS-SVM and classical SVM on over forty data sets. Moreover, they studied various kinds of activation functions in ELM. The experimental results verified that ELM was able to exhibit similar or better generalisation performance for binary class classification and regression, and that it showed pointedly better generalisation performance in terms of multiclass classification data sets. Furthermore, ELM had a significantly faster learning speed (reaching up to several orders of magnitude) and better scalability.

Comparison with deep learning

Kasun et al. (2013) developed an ELM-based auto encoder (ELM-AE) that can be used for classification and representational learning. Their experiments tested a 784-700-700-15000-10 multi-layer ELM network using the popular MNIST data set. This data set contains 10,000 images for testing and 60,000 images of handwritten digits that are used for training. The results showed that compared to other state-of-the-art deep learning techniques, the multi-layer ELMAE achieved matchable precision and was significantly faster in training (see Table 2). Kasun et al. (2013) also examined by which the ELM auto encoder acquires and learns feature representations. They made 10 mini data sets that contained digits 0–9 taken from the MNIST data set. Each mini data set was then sent through an ELM AE (network structure: 784-20-784). They observed that the output weights β of the ELM-AE were able to actually obtain valuable information from the original images. Additionally, Huang et al. (2014) demonstrated that the unsupervised ELM performed better than the deep auto-encoder in terms of clustering and embedding tasks.

Table 2: Performance comparison of ELM-AE with state-of-the-art deep networks on MNIST data set (Kasun et al. 2013)

Algorithms	Testing accuracy (%)	Training time
ELM-AE	99.03	444.655 s
Deep belief network (DBN)	98.87	20,580 s
Deep Boltzmann machine (DBM)	99.05	68,246 s
Stacked auto-encoder (SAE)	98.6	>17 h
Stacked denoising auto-encoder (SDAE)	98.72	>17 h

Extreme Learning Machine for Speech Application

Comprehensive empirical studies have been performed in language identification. Furthermore, several attempts have been conducted to build an ELM-based language classifier as a replacement for the classical SVM. (Xu et al. 2015) formulated a new type of extreme learning machine which they then applied on language identification. They called this algorithm the Regularised Minimum Class Variance Extreme Learning Machine (RMCVELM). The algorithm's core goal is to lessen the structural risk, empirical risk, and the intra-class variance. The authors assessed it based on execution time and accuracy. They discovered that it performed better than SVM in terms of execution time. It was also able to achieve comparable classification accuracy. (Lan et al. 2013) also tried to apply extreme learning machine for speaker recognition. They applied LM on speaker that has text independent data. They then compared the results obtained with that of SVM. They found out that ELM has higher accuracy and faster execution. (Han et al. 2014) also attempted to identify the speaker's emotion using extreme learning machine as classifier and DNN as feature extractor. They found that ELM and Kernel ELM (KELM), when combined with DNN, have the highest accuracies compared to all the other baseline approaches. (Muthusamy et al. 2015) utilised ELM with another classifier on various types of audio-related classification problems. They also addressed emotion recognition based on the speaker's audio. They used GMM model features as inputs for the classifier. The authors stress the power of GMM-based features in offering discriminative factors that can be used to classify emotions.

ELM for medical/biomedical applications

Medical or biomedical data typically have high dimensional features or a large amount of samples. Thus, medical or biomedical data analysis often utilise advanced machine learning techniques like SVM, Since ELM offers many advantages compared to other learning algorithms, its application in this area could be an interesting thing to see. Indeed, many encouraging results on the application of ELM to predict protein–protein interactions (You et al. 2013), EEG-based vigilance estimation(Shi & Lu 2013),epileptic EEG patterns recognition (Yuan et al. 2011; Song et al. 2012; Song & Zhang 2013), transmembrane beta-barrel chains detection (Savojardo et al. 2011), an eye-control method for eye-based computer interaction (Barea et al. 2012), spike sorting with overlap resolution that is based on a hybrid noise-assisted methodology (Adamos et al. 2010), lie detection (Gao et al. 2013), an electrocardiogram ECG (Karpagachelvi et al. 2012), liver parenchyma segmentation (Huang et al. 2012), diagnosis of hepatitis (Kaya & Uyar 2013), thyroid (Li et al. 2012), arrhythmia classification in ECG (Kim et al. 2009), detection of mycobacterium tuberculosis in tissue sections (Osman et al. 2012), protein secondary structure prediction (Saraswathi et al. 2012), and metagenomics taxonomic classification (Rasheed & Rangwala 2012) have been observed in recent years.

ELM for computer vision

ELM has had successful applications in various computer vision tasks, such as human action recognition (Minhas et al. 2010; Minhas et al. 2012), face recognition (Mohammed et al. 2011; Zong & Huang 2011; Choi et al. 2012; Baradarani et al. 2013; Marqués & Graña 2013; He et al. 2014), terrain-based navigation (Kan et al. 2013), and matching of fingerprints (Yang et al. 2013).

ELM for image processing

ELM is also considered an attractive technique for image processing. For instance, (An & Bhanu 2012) introduced an efficient image super resolution method that has its basis in ELM. The aim of their approach is to generate high resolution images from inputs with low-resolution. In the training process, the input was extracted from the image features. Furthermore, the high frequency components that were taken from the original images with high-resolution were utilized as the target values. Then, ELM learns a model that is capable of mapping the interpolated image and imposing it on the high-frequency components. Once training is done, the learned model can predict the high-frequency components using low-resolution images. (Li et al. 2013) used ELM to burn state recognition of rotary kiln. (Chang et al. 2010) used ELM for change detection of land cover and land use. Moreover, ELM was utilised for image classification by (Cao et al. 2013; Bazi et al. 2014). ELM was used to assess the perceived image quality (Decherchi et al. 2013; Suresh et al. 2009). ELM was also utilised in the detection of semantic concept for videos (Lu et al. 2013). Image deblurring can also be done using filters that are learned by ELM (Wang et al. 2011). SAE-ELM was utilised in the coal mine water inrush's multi-level forecasting model (Zhao & Hu 2014).

ELM for system modelling and prediction

Because traditional neural networks have had wide uses in system prediction and modelling, ELM also has great potential in the development of accurate and efficient models for these applications. (Xu et al. 2013) proposed an ELM-based predictor that can be used in the actual frequency stability assessment (FSA) of power systems. The predictor's inputs are the power system operational parameters, while the output is set as the frequency stability margin. This margin measures the power system's stability degree, subject to a contingency. Using off-line training and a frequency stability database, one can apply the predictor online for real-time FSA. They tested the predictor on New England's 10-generator 39-bus test system. The results of this simulation revealed that it is capable of accurately and efficiently predicting the frequency stability. ELM was also utilised for electricity price forecasting (Chen et al. 2012), sales forecasting (Wong & Guo 2010; Chen & Ou 2011), temperature prediction of molten steel (Tian & Mao 2010), security assessment of wind power system (Xu et al. 2012; Xu et al. 2012), drying system modelling (Balbay et al. 2012), etc. Because of its notable advantages, many other applications have adopted ELM. Based on past literature, we can witness its successful applications in control system design, text analysis, chemical process monitor, feature selection, clustering, ranking, and representational learning. Furthermore, ELM can be applied to more potential fields as well.

STRENGTH AND WEAKNESSES OF ELM

Majority of the literature reviewed considers ELM to be a good learning machine tool. ELM's major strength is that the hidden layers' learning parameters, including the biases and input weights, do not have to be iteratively tuned like in SLFN (Huang et al. 2004; Huang et al. 2006; Huang et al. 2012; Ding et al. 2014; Lin et al. 2015; Liu et al. 2015; Ebtehaj et al. 2016). Because of this, the ELM is capable of achieving faster speeds and lower costs (Huang et al. 2015). Furthermore, it is the most favoured in machine learning compared to its predecessors. Some of the other commendable attributes of ELM include good generalisation accuracy and performance (Huang et al. 2015), simple learning algorithm (Zhang et al. 2016), improved efficiency (Huang et al. 2006; Ding et al. 2014), non-linear transformation during its training phase, possession of a unified solution to different practical applications (Huang 2015), lack of local minimal and overfitting (Huang et al. 2006; Huang 2015), the need for fewer optimisations compared to SVM and LS-SVM, and its similar computational cost with SVM (Zhang et al. 2016). More importantly, ELM is able to bridge the gap between biological learning machines and conventional learning machines (Huang et al. 2015),which is the goal of (Huang et al. 2004) (cited in(Huang et al. 2015)), a pioneer in the study of ELM.

Despite the many advantages that ELM possesses, it still has some flaws. For example, it was observed that the classification boundary of the hidden layers' learning parameters may not be optimal since they remain the same during training (Cao et al. 2012; Ding et al. 2014). Furthermore, ELM is not capable of managing large high dimensional data (Huang et al. 2015; Zhang et al. 2016) since it needs more hidden nodes compared to the conventional tuning algorithms. It also is not suitable for being parallelised because it goes through pseudo-inverse circulation (Oneto et al. 2016). A number of these challenges are already being addressed by modifications, optimisations, and hybridisations. However, many of the current literature in ELM still have the following recommendations for further research: (i) theoretical proof and application of the optimal amount of hidden nodes, (ii) approximation of generalisation performance (iii) generalisation capability so that it can manage high dimensional data (Deng et al. 2016; Liu et al. 2016; Oneto et al. 2016; Wang et al. 2016),(iv) modification of ELM algorithm for distributed and parallel computation (Ding et al. 2014; Bodyanskiy et al. 2016). (Huang 2014) also stressed on the need to examine the connection between ELM and other algorithms related to it, such as the random forest algorithm.

CONCLUSION

This paper presented a comprehensive review of the ELM algorithm, with emphasis on its applications and variants. The aim of the paper is to show that ELM is a valuable tool for research applications, as it can provide more accurate results and save time in terms of the calculation time during classification, regression and other similar problems. However, some of the ELM algorithm's open problems have to be solved as well. The following concerns are still open and may be worth studying in the future: (i) theoretical proof and application of the optimal amount of hidden nodes, (ii) approximation of generalisation performance (iii) generalisation capability so that it can manage high dimensional data (Deng et al. 2016; Liu et al. 2016; Oneto et al. 2016; Wang et al. 2016), (iv) modification of ELM algorithm for distributed and parallel computation (Ding et al. 2014; Bodyanskiy et al. 2016). (Huang 2014) also stressed on the need to examine the connection between ELM and other algorithms related to it, such as the random forest algorithm.

ACKNOWLEDGEMENTS

This project is funded by Malaysian government under research code FRGS/1/2016/ICT02/UKM/01/14.

REFERENCE

[1] Adamos, D. A., N. A. Laskaris, E. K. Kosmidis & G. Theophilidis 2010. NASS: an empirical approach to spike sorting with overlap resolution based on a hybrid noise-assisted methodology. *Journal of neuroscience methods*190(1): 129-142.

[2] An, L. & B. Bhanu 2012. Image super-resolution by extreme learning machine. *Image processing (ICIP), 2012 19th IEEE International Conference on*. pp. 2209-2212.

[3] Avci, E. & R. Coteli 2012. A new automatic target recognition system based on wavelet extreme learning machine. *Expert Systems with Applications*39(16): 12340-12348.

[4] Bai, Z., G.-B. Huang, D. Wang, H. Wang & M. B. Westover 2014. Sparse extreme learning machine for classification. *IEEE transactions on cybernetics*44(10): 1858-1870.

[5] Balbay, A., Y. Kaya & O. Sahin 2012. Drying of black cumin (Nigella sativa) in a microwave assisted drying system and modeling using extreme learning machine. *Energy*44(1): 352-357.

[6] Baradarani, A., Q. J. Wu & M. Ahmadi 2013. An efficient illumination invariant face recognition framework via illumination enhancement and DD-DTCWT filtering. *Pattern Recognition*46(1): 57-72.

[7] Barea, R., L. Boquete, S. Ortega, E. López & J.

Rodríguez-Ascariz 2012. EOG-based eye movements codification for human computer interaction. *Expert Systems with Applications*39(3): 2677-2683.

[8] Bazi, Y., N. Alajlan, F. Melgani, H. AlHichri, S. Malek & R. R. Yager 2014. Differential evolution extreme learning machine for the classification of hyperspectral images. *IEEE Geoscience and Remote Sensing Letters* **11**(6): 1066-1070.

[9] Bengio, Y. 2009. Learning deep architectures for AI. *Foundations and trends® in Machine Learning*2(1): 1-127.

[10] BenoíT, F., M. Van Heeswijk, Y. Miche, M. Verleysen & A. Lendasse 2013. Feature selection for nonlinear models with extreme learning machines. *Neurocomputing*102: 111-124.

[11] Bodyanskiy, Y., O. Vynokurova, I. Pliss, G. Setlak & P. Mulesa 2016. Fast learning algorithm for deep evolving GMDH-SVM neural network in data stream mining tasks. *Data Stream Mining & Processing (DSMP), IEEE First International Conference on*. pp. 257-262.

[12] Branke, J. 1995. Evolutionary algorithms for neural network design and training. *In Proceedings of the First Nordic Workshop on Genetic Algorithms and its Applications*.

[13] Cao, F., B. Liu & D. S. Park 2013. Image classification based on effective extreme learning machine. *Neurocomputing* **102**: 90-97.

[14] Cao, J., Z. Lin & G.-b. Huang 2010. Composite function wavelet neural networks with extreme learning machine. *Neurocomputing*73(7): 1405-1416.

[15] Cao, J., Z. Lin, G.-B. Huang & N. Liu 2012. Voting based extreme learning machine. *Information Sciences*185(1): 66-77.

[16] Chang, N.-B., M. Han, W. Yao, L.-C. Chen & S. Xu 2010. Change detection of land use and land cover in an urban region with SPOT-5 images and partial Lanczos extreme learning machine. *Journal of Applied Remote Sensing*4(1): 043551.

[17] Chen, F. & T. Ou 2011. Sales forecasting system based on Gray extreme learning machine with Taguchi method in retail industry. *Expert Systems with Applications*38(3): 1336-1345.

[18] Chen, S., C. F. Cowan & P. M. Grant 1991. Orthogonal least squares learning algorithm for radial basis function networks. *IEEE Transactions on neural networks*2(2): 302-309.

[19] Chen, X., Z. Y. Dong, K. Meng, Y. Xu, K. P. Wong & H. Ngan 2012. Electricity price forecasting with extreme learning machine and bootstrapping. *IEEE Transactions on Power Systems*27(4): 2055-2062.

[20] Choi, K., K.-A. Toh & H. Byun 2012. Incremental face recognition for large-scale social network services. *Pattern Recognition*45(8): 2868-2883.

[21] Cortes, C. & V. Vapnik 1995. Support vector machine. *Machine learning*20(3): 273-297.

[22] Daliri, M. R. 2012. A hybrid automatic system for the diagnosis of lung cancer based on genetic algorithm and fuzzy extreme learning machines. *Journal of medical systems*36(2): 1001-1005.

[23] Decherchi, S., P. Gastaldo, R. Zunino, E. Cambria & J. Redi 2013. Circular-ELM for the reduced-reference assessment of perceived image quality. *Neurocomputing*102: 78-89.

[24] Deng, J., K. Li & G. W. Irwin 2011. Fast automatic two-stage nonlinear model identification based on the extreme learning machine. *Neurocomputing*74(16): 2422-2429.

[25] Deng, W.-Y., Z. Bai, G.-B. Huang & Q.-H. Zheng 2016. A fast SVD-hidden-nodes based extreme learning machine for large-scale data analytics. *Neural Networks*77: 14-28.

[26] Deng, W.-Y., Q.-H. Zheng, S. Lian, L. Chen & X. Wang 2010. Ordinal extreme learning machine. *Neurocomputing*74(1): 447-456.

[27] Ding, S., X. Xu & R. Nie 2014. Extreme learning machine and its applications. *Neural computing & applications*25.

[28] Du, D., K. Li, G. W. Irwin & J. Deng 2013. A novel automatic two-stage locally regularized classifier construction method using the extreme learning machine. *Neurocomputing*102: 10-22.

[29] Ebtehaj, I., H. Bonakdari & S. Shamshirband 2016. Extreme learning machine assessment for estimating sediment transport in open channels. *Engineering with Computers*32(4): 691-704.

[30] Feng, G., G.-B. Huang, Q. Lin & R. Gay 2009. Error minimized extreme learning machine with growth of hidden nodes and incremental learning. *IEEE Transactions on Neural Networks*20(8): 1352-1357.

[31] Fernández-Delgado, M., E. Cernadas, S. Barro, J. Ribeiro & J. Neves 2014. Direct Kernel Perceptron (DKP): Ultra-fast kernel ELM-based classification with non-iterative closed-form weight calculation. *Neural Networks*50: 60-71.

[32] Gao, J., Z. Wang, Y. Yang, W. Zhang, C. Tao, J. Guan & N. Rao 2013. A novel approach for lie detection based on F-score and extreme learning machine. *PloS one*8(6): e64704.

[33] Hagan, M. T. & M. B. Menhaj 1994. Training feedforward networks with the Marquardt algorithm. *IEEE transactions on Neural Networks*5(6): 989-993.

[34] Han, K., D. Yu & I. Tashev 2014. Speech emotion recognition using deep neural network and extreme learning machine. *Interspeech*. pp. 223-227.

[35] He, B., D. Xu, R. Nian, M. van Heeswijk, Q. Yu, Y. Miche & A. Lendasse 2014. Fast Face Recognition

Via Sparse Coding and Extreme Learning Machine. *Cognitive Computation*6(2): 264-277.

[36] He, Q., C. Du, Q. Wang, F. Zhuang & Z. Shi 2011. A parallel incremental extreme SVM classifier. *Neurocomputing*74(16): 2532-2540.

[37] Horata, P., S. Chiewchanwattana & K. Sunat 2013. Robust extreme learning machine. *Neurocomputing*102: 31-44.

[38] Huang, G.-B. 2014. An insight into extreme learning machines: random neurons, random features and kernels. *Cognitive Computation*6(3): 376-390.

[39] Huang, G.-B. 2015. What are extreme learning machines? Filling the gap between Frank Rosenblatt's dream and John von Neumann's puzzle. *Cognitive Computation*7(3): 263-278.

[40] Huang, G.-B. & L. Chen 2007. Convex incremental extreme learning machine. *Neurocomputing*70(16): 3056-3062.

[41] Huang, G.-B. & L. Chen 2008. Enhanced random search based incremental extreme learning machine. *Neurocomputing*71(16): 3460-3468.

[42] Huang, G.-B., L. Chen & C. K. Siew 2006. Universal approximation using incremental constructive feedforward networks with random hidden nodes. *IEEE Trans. Neural Networks*17(4): 879-892.

[43] Huang, G.-B., M.-B. Li, L. Chen & C.-K. Siew 2008. Incremental extreme learning machine with fully complex hidden nodes. *Neurocomputing*71(4): 576-583.

[44] Huang, G.-B., D. H. Wang & Y. Lan 2011. Extreme learning machines: a survey. *International journal of machine learning and cybernetics*2(2): 107-122.

[45] Huang, G.-B., H. Zhou, X. Ding & R. Zhang 2012. Extreme learning machine for regression and multiclass classification. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*42(2): 513-529.

[46] Huang, G.-B., Q.-Y. Zhu & C.-K. Siew 2004. Extreme learning machine: a new learning scheme of feedforward neural networks. *Neural Networks, 2004. Proceedings. 2004 IEEE International Joint Conference on*. 2 pp. 985-990.

[47] Huang, G.-B., Q.-Y. Zhu & C.-K. Siew 2006. Extreme learning machine: theory and applications. *Neurocomputing*70(1): 489-501.

[48] Huang, G., G.-B. Huang, S. Song & K. You 2015. Trends in extreme learning machines: A review. *Neural Networks*61: 32-48.

[49] Huang, G., S. Song, J. N. Gupta & C. Wu 2014. Semi-supervised and unsupervised extreme learning machines. *IEEE transactions on cybernetics*44(12): 2405-2417.

[50] Huang, W., Z. Tan, Z. Lin, G.-B. Huang, J. Zhou, C. Chui, Y. Su & S. Chang 2012. A semi-automatic approach to the segmentation of liver parenchyma from 3D CT images with Extreme Learning Machine. *Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE*. pp. 3752-3755.

[51] Kan, E. M., M. H. Lim, Y. S. Ong, A. H. Tan & S. P. Yeo 2013. Extreme learning machine terrain-based navigation for unmanned aerial vehicles. *Neural computing & applications*: 1-9.

[52] Karpagachelvi, S., M. Arthanari & M. Sivakumar 2012. Classification of electrocardiogram signals with support vector machines and extreme learning machine. *Neural Computing & Applications*21(6): 1331-1339.

[53] Kasun, L. L. C., H. Zhou, G.-B. Huang & C. M. Vong 2013. Representational learning with ELMs for big data.

[54] Kaya, Y. & M. Uyar 2013. A hybrid decision support system based on rough set and extreme learning machine for diagnosis of hepatitis disease. *Applied Soft Computing*13(8): 3429-3438.

[55] Kim, J., H. S. Shin, K. Shin & M. Lee 2009. Robust algorithm for arrhythmia classification in ECG using extreme learning machine. *Biomedical engineering online*8(1): 31.

[56] Lahoz, D., B. Lacruz & P. M. Mateo 2013. A multi-objective micro genetic ELM algorithm. *Neurocomputing*111: 90-103.

[57] Lan, Y., Z. Hu, Y. C. Soh & G.-B. Huang 2013. An extreme learning machine approach for speaker recognition. *Neural Computing and Applications*22(3-4): 417-425.

[58] Lan, Y., Y. C. Soh & G.-B. Huang 2009. Ensemble of online sequential extreme learning machine. *Neurocomputing*72(13): 3391-3395.

[59] Lan, Y., Y. C. Soh & G.-B. Huang 2010. Two-stage extreme learning machine for regression. *Neurocomputing*73(16): 3028-3038.

[60] Le, Q., T. Sarlós & A. Smola 2013. Fastfood-approximating kernel expansions in loglinear time. *Proceedings of the international conference on machine learning*. 85.

[61] Li, B., Y. Li & X. Rong 2013. The extreme learning machine learning algorithm with tunable activation function. *Neural computing & applications*: 1-9.

[62] Li, K., J.-X. Peng & G. W. Irwin 2005. A fast nonlinear model identification method. *IEEE Transactions on Automatic Control*50(8): 1211-1216.

[63] Li, L.-N., J.-H. Ouyang, H.-L. Chen & D.-Y. Liu 2012. A computer aided diagnosis system for thyroid disease using extreme learning machine. *Journal of medical systems*36(5): 3327-3337.

[64] Li, M.-B., G.-B. Huang, P. Saratchandran & N.

Sundararajan 2005. Fully complex extreme learning machine. *Neurocomputing*68: 306-314.

[65] Li, W., D. Wang & T. Chai 2013. Burning state recognition of rotary kiln using ELMs with heterogeneous features. *Neurocomputing*102: 144-153.

[66] Liang, N.-Y., G.-B. Huang, P. Saratchandran & N. Sundararajan 2006. A fast and accurate online sequential learning algorithm for feedforward networks. *IEEE Transactions on neural networks*17(6): 1411-1423.

[67] Lin, S., X. Liu, J. Fang & Z. Xu 2015. Is extreme learning machine feasible? A theoretical assessment (Part II). *IEEE Transactions on Neural Networks and Learning Systems*26(1): 21-34.

[68] Liu, D., Y. Wu & H. Jiang 2016. FP-ELM: An online sequential learning algorithm for dealing with concept drift. *Neurocomputing*207: 322-334.

[69] Liu, X., C. Gao & P. Li 2012. A comparative analysis of support vector machines and extreme learning machines. *Neural Networks*33: 58-66.

[70] Liu, X., P. Li & C. Gao 2013. Symmetric extreme learning machine. *Neural computing & applications*: 1-8.

[71] Liu, X., S. Lin, J. Fang & Z. Xu 2015. Is extreme learning machine feasible? A theoretical assessment (Part I). *IEEE Transactions on Neural Networks and Learning Systems*26(1): 7-20.

[72] Lu, B., G. Wang, Y. Yuan & D. Han 2013. Semantic concept detection for video based on extreme learning machine. *Neurocomputing*102: 176-183.

[73] Malathi, V., N. Marimuthu & S. Baskar 2010. Intelligent approaches using support vector machine and extreme learning machine for transmission line protection. *Neurocomputing*73(10): 2160-2167.

[74] Malathi, V., N. Marimuthu, S. Baskar & K. Ramar 2011. Application of extreme learning machine for series compensated transmission line protection. *Engineering Applications of Artificial Intelligence*24(5): 880-887.

[75] Man, Z., K. Lee, D. Wang, Z. Cao & C. Miao 2011. A new robust training algorithm for a class of single-hidden layer feedforward neural networks. *Neurocomputing*74(16): 2491-2501.

[76] Marqués, I. & M. Graña 2013. Fusion of lattice independent and linear features improving face identification. *Neurocomputing*114: 80-85.

[77] MartíNez-MartíNez, J. M., P. Escandell-Montero, E. Soria-Olivas, J. D. MartíN-Guerrero, R. Magdalena-Benedito & J. GóMez-Sanchis 2011. Regularized extreme learning machine for regression problems. *Neurocomputing*74(17): 3716-3721.

[78] Miche, Y., A. Sorjamaa, P. Bas, O. Simula, C. Jutten & A. Lendasse 2010. OP-ELM: optimally pruned extreme learning machine. *IEEE transactions on neural networks*21(1): 158-162.

[79] Minhas, R., A. Baradarani, S. Seifzadeh & Q. J. Wu 2010. Human action recognition using extreme learning machine based on visual vocabularies. *Neurocomputing*73(10): 1906-1917.

[80] Minhas, R., A. A. Mohammed & Q. J. Wu 2012. Incremental learning in human action recognition based on snippets. *IEEE Transactions on Circuits and Systems for Video Technology*22(11): 1529-1541.

[81] Mohammed, A. A., R. Minhas, Q. J. Wu & M. A. Sid-Ahmed 2011. Human face recognition based on multidimensional PCA and extreme learning machine. *Pattern Recognition*44(10): 2588-2597.

[82] Muthusamy, H., K. Polat & S. Yaacob 2015. Improved emotion recognition using Gaussian Mixture Model and extreme learning machine in speech and glottal signals. *Mathematical Problems in Engineering*2015.

[83] Oneto, L., F. Bisio, E. Cambria & D. Anguita 2016. Statistical learning theory and ELM for big social data analysis. *ieee CompUTATionAl inTelliGenCe mAGAzine*11(3): 45-55.

[84] Osman, M. K., M. Y. Mashor & H. Jaafar 2012. Performance comparison of extreme learning machine algorithms for mycobacterium tuberculosis detection in tissue sections. *Journal of Medical Imaging and Health Informatics*2(3): 307-312.

[85] Qu, Y., C. Shang, W. Wu & Q. Shen 2011. Evolutionary Fuzzy Extreme Learning Machine for Mammographic Risk Analysis. *International Journal of Fuzzy Systems*13(4).

[86] Rahimi, A. & B. Recht 2008. Random features for large-scale kernel machines. *Advances in neural information processing systems*. pp. 1177-1184.

[87] Rahimi, A. & B. Recht 2008. Uniform approximation of functions with random bases. *Communication, Control, and Computing, 2008 46th Annual Allerton Conference on*. pp. 555-561.

[88] Rahimi, A. & B. Recht 2009. Weighted sums of random kitchen sinks: Replacing minimization with randomization in learning. *Advances in neural information processing systems*. pp. 1313-1320.

[89] Rasheed, Z. & H. Rangwala 2012. Metagenomic taxonomic classification using extreme learning machines. *Journal of bioinformatics and computational biology*10(05): 1250015.

[90] Rong, H.-J., G.-B. Huang, N. Sundararajan & P. Saratchandran 2009. Online sequential fuzzy extreme learning machine for function approximation and classification problems. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*39(4): 1067-1072.

[91] Rong, H.-J., Y.-S. Ong, A.-H. Tan & Z. Zhu 2008. A

fast pruned-extreme learning machine for classification problem. *Neurocomputing*72(1): 359-366.

[92] Rumelhart, D. E., G. E. Hinton & R. J. Williams 1988. Learning representations by back-propagating errors. *Cognitive modeling*5(3): 1.

[93] Saraswathi, S., J. L. Fernández-Martínez, A. Koliński, R. L. Jernigan & A. Kloczkowski 2012. Fast learning optimized prediction methodology (FLOPRED) for protein secondary structure prediction. *Journal of molecular modeling*: 1-15.

[94] Savojardo, C., P. Fariselli & R. Casadio 2011. Improving the detection of transmembrane β-barrel chains with N-to-1 extreme learning machines. *Bioinformatics*27(22): 3123-3128.

[95] Saxe, A., P. W. Koh, Z. Chen, M. Bhand, B. Suresh & A. Y. Ng 2011. On random weights and unsupervised feature learning. *Proceedings of the 28th international conference on machine learning (ICML-11)*. pp. 1089-1096.

[96] Shi, L.-C. & B.-L. Lu 2013. EEG-based vigilance estimation using extreme learning machines. *Neurocomputing*102: 135-143.

[97] Song, Y., J. Crowcroft & J. Zhang 2012. Automatic epileptic seizure detection in EEGs based on optimized sample entropy and extreme learning machine. *Journal of neuroscience methods*210(2): 132-146.

[98] Song, Y. & J. Zhang 2013. Automatic recognition of epileptic EEG patterns via extreme learning machine and multiresolution feature extraction. *Expert Systems with Applications*40(14): 5477-5489.

[99] Suresh, S., R. V. Babu & H. Kim 2009. No-reference image quality assessment using modified extreme learning machine classifier. *Applied Soft Computing*9(2): 541-552.

[100] Suykens, J. A. & J. Vandewalle 1999. Least squares support vector machine classifiers. *Neural processing letters*9(3): 293-300.

[101] Tian, H.-X. & Z.-Z. Mao 2010. An ensemble ELM based on modified AdaBoost. RT algorithm for predicting the temperature of molten steel in ladle furnace. *IEEE Transactions on Automation Science and Engineering*7(1): 73-80.

[102] Wang, H., Z. Xu & W. Pedrycz 2016. An overview on the roles of fuzzy set techniques in big data processing: Trends, challenges and opportunities. *architecture*2: 14.

[103] Wang, L., Y. Huang, X. Luo, Z. Wang & S. Luo 2011. Image deblurring with filters learned by extreme learning machine. *Neurocomputing*74(16): 2464-2474.

[104] Wang, N., M. J. Er & M. Han 2014. Parsimonious extreme learning machine using recursive orthogonal least squares. *IEEE transactions on neural networks and learning systems*25(10): 1828-1841.

[105] Widrow, B., A. Greenblatt, Y. Kim & D. Park 2013. The No-Prop algorithm: A new learning algorithm for multilayer neural networks. *Neural Networks*37: 182-188.

[106] Wilamowski, B. M. & H. Yu 2010. Neural network learning without backpropagation. *IEEE Transactions on Neural Networks*21(11): 1793-1803.

[107] Wong, W. & Z. Guo 2010. A hybrid intelligent model for medium-term sales forecasting in fashion retail supply chains using extreme learning machine and harmony search algorithm. *International Journal of Production Economics*128(2): 614-624.

[108] Xi-Zhao, W., S. Qing-Yan, M. Qing & Z. Jun-Hai 2013. Architecture selection for networks trained with extreme learning machine using localized generalization error model. *Neurocomputing*102: 3-9.

[109] Xu, J., W.-Q. Zhang, J. Liu & S. Xia 2015. Regularized minimum class variance extreme learning machine for language recognition. *EURASIP Journal on Audio, Speech, and Music Processing*2015(1): 22.

[110] Xu, Y., Y. Dai, Z. Y. Dong, R. Zhang & K. Meng 2013. Extreme learning machine-based predictor for real-time frequency stability assessment of electric power systems. *Neural computing & applications*: 1-8.

[111] Xu, Y., Z. Y. Dong, Z. Xu, K. Meng & K. P. Wong 2012. An intelligent dynamic security assessment framework for power systems with wind power. *IEEE Transactions on industrial informatics*8(4): 995-1003.

[112] Xu, Y., Z. Y. Dong, J. H. Zhao, P. Zhang & K. P. Wong 2012. A reliable intelligent system for real-time dynamic security assessment of power systems. *IEEE Transactions on Power Systems*27(3): 1253-1263.

[113] Yang, J., S. Xie, S. Yoon, D. Park, Z. Fang & S. Yang 2013. Fingerprint matching based on extreme learning machine. *Neural computing & applications*: 1-11.

[114] Yang, Y., Y. Wang & X. Yuan 2012. Bidirectional extreme learning machine for regression problem and its learning effectiveness. *IEEE Transactions on Neural Networks and Learning Systems*23(9): 1498-1505.

[115] Yang, Y., Y. Wang & X. Yuan 2013. Parallel chaos search based incremental extreme learning machine. *Neural Processing Letters*: 1-25.

[116] Yao, X. 1993. A review of evolutionary artificial neural networks. *International journal of intelligent systems*8(4): 539-567.

[117] Ye, Y., S. Squartini & F. Piazza 2013. Online sequential extreme learning machine in nonstationary

environments. *Neurocomputing*116: 94-101.

[118] You, Z.-H., Y.-K. Lei, L. Zhu, J. Xia & B. Wang 2013. Prediction of protein-protein interactions from amino acid sequences with ensemble extreme learning machines and principal component analysis. *BMC bioinformatics*14(8): S10.

[119] Yu, Q., Y. Miche, E. Eirola, M. Van Heeswijk, E. SéVerin & A. Lendasse 2013. Regularized extreme learning machine for regression with missing data. *Neurocomputing*102: 45-51.

[120] Yuan, Q., W. Zhou, S. Li & D. Cai 2011. Epileptic EEG classification based on extreme learning machine and nonlinear features. *Epilepsy research*96(1): 29-38.

[121] Zhang, J., L. Feng & B. Wu 2016. Local extreme learning machine: local classification model for shape feature extraction. *Neural Computing and Applications*27(7): 2095-2105.

[122] Zhang, L., J. Li & H. Lu 2016. Saliency detection via extreme learning machine. *Neurocomputing*218: 103-112.

[123] Zhang, W. & H. Ji 2013. Fuzzy extreme learning machine for classification. *Electronics Letters*49(7): 448-450.

[124] Zhao, J., Z. Wang & D. S. Park 2012. Online sequential extreme learning machine with forgetting mechanism. *Neurocomputing*87: 79-89.

[125] Zhao, Z. & M. Hu 2014. Multi-level forecasting model of coal mine water inrush based on self-adaptive evolutionary extreme learning machine. *Appl. Math. Inf. Sci. Lett*2(3): 103-110.

[126] Zhu, Q.-Y., A. K. Qin, P. N. Suganthan & G.-B. Huang 2005. Evolutionary extreme learning machine. *Pattern recognition*38(10): 1759-1763.

[127] Zong, W. & G.-B. Huang 2011. Face recognition based on extreme learning machine. *Neurocomputing*74(16): 2541-2551.

[128] Zong, W., G.-B. Huang & Y. Chen 2013. Weighted extreme learning machine for imbalance learning. *Neurocomputing*101: 229-242.