

Design of Improved Version of Sigmoidal Function with Biases for Classification Task in ELM Domain

S. R. Mugunthan,

Associate Professor,
Department of Computer Science and Engineering,
Sriindu College of Engineering and Technology,
Hyderabad, India.
srmugunth@gmail.com

Dr. T. Vijayakumar,

Professor,
Department of ECE,
Guru Nanak Institute of Technology,
Hyderabad, India.
vishal_16278@yahoo.co.in

Abstract- Extreme Learning Machine (ELM) is one of the latest trends in learning algorithm, which can provide a good recognition rate within less computation time. Therefore, the algorithm can sustain for a faster response application by utilizing a feed-forward neural network. In this research article, the ELM method has been designed with the presence of sigmoidal function of biases in the hidden nodes to perform the classification task. The classification task is very challenging with the existing learning algorithm and increased computation time due to the huge amount of dataset. While handling of the stochastic matrix for hidden layer, output provides the lower performance for learning rate and robustness in the determination. To address these issues, the modified version of ELM has been developed to obtain better accuracy and minimize the classification error. This research article includes the mathematical proof of sigmoidal activation function with biases of the hidden nodes present in the networks. The output matrix maintains the column rank in order to improve the speed of the training output weights (β). The proposed improved version of ELM leverages better accuracy and efficacy in classification and regression problems as well. Due to the inclusion of matrix

column ranking in mathematical proof, the proposed improved version of ELM solves the slow training speed and over-fitting problems present in the existing learning approach.

Keywords: *Extreme Learning Machine*

1. INTRODUCTION

Recently, more developments of ELM are observed with the introduction of a new class of SLFN. Various patterns are available to modify the network structure by using input weights and biases. This is the basic idea involved in selecting the random hidden nodes from the layers before undergoing the good training procedure [1]. ELM algorithm addresses many classification problems, recognition issues, and also the image classification challenges. This ELM algorithm has the ability to provide an efficient learning framework for classification and regression tasks [2].

The study of the feed-forward neural network is universal and also about the compact input handling for approximation finite set distribution across layer parameters. The parameters are validated throughout the layer and further it is used to determine the layer-to-layer connection arbitrarily [3]. The Back propagation algorithm is used to tune the weight vector and biases of the network in order to minimize the classification error. Besides, appending this algorithm with ELM will provide a better convergence rate in the network during its training time [4].

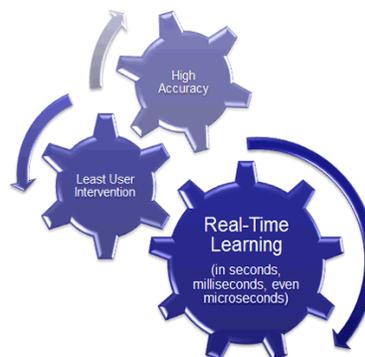


Figure 1 ELM Framework with Targets

The learning speed is increasing than the existing method and its training process become very simple. Based on stochastic characteristics, the connecting weight vectors are present in the structured network to perform the learning process [5]. Many learning algorithms are unstable to perform the training procedure. ELM procedure is combined with ensemble techniques. The multiple machine learning algorithms are organized through the ensemble procedure [6] [7]. The application-wise traditional network is combined with a real-time learning system, which is shown in figure 1.

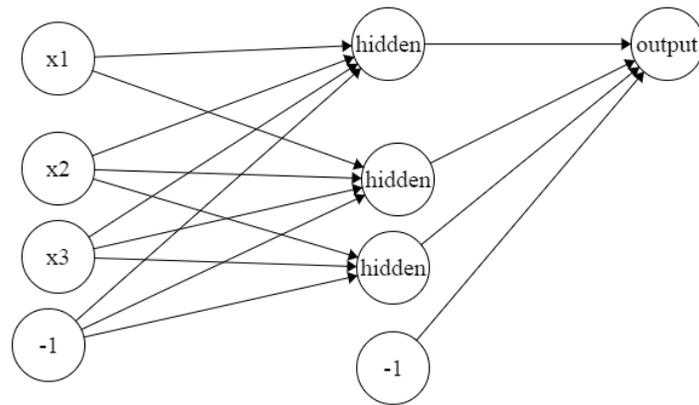


Figure 2 Simplified Input Output Weights

Recently, the Internet of Things (IoT) model is used to predict or detect many applications with a traditional network that is connected to several sensing devices through the internet [8]. The IoT application benefits are growing day by day in the world market. Besides, the online security remains as a challenging task with widespread attention [9]. Figure 2 shows the simplified input and output weight vectors. The learning algorithm requires an optimized algorithm to reduce the computation time and enhance the processing of the neural network. The feed-forward neural network algorithm includes input, hidden, and output layer weight vectors with an implementation [10]. ELM should be extended with a faster processing algorithm to solve the aforementioned issues and it should be implemented in IoT devices [11].

To increase the generalization ability, ELM provides fast processing time for the network. Initially, the ELM chooses the input weight vector randomly with matrix computation to determine the output weight vectors. The utilization of optimal hyper-plane in the network

provides reliable classification results in any applications [12]. Based on the above-integrated concepts, this research article modifies the sigmoidal function with an optimized parameter.

2. ORGANIZATION OF THE RESEARCH

The research article has been organized as several sections. Section 3 is containing related works of ELM-based classification, section 4 gives the design details of modified ELM based on various parameters. Section 5 discusses the description of the results tested and obtained. Finally, the proposed research work is concluded in section 6.

3. PRELIMINARIES

The neural networks with feed-forward model have uniform capabilities on the input weight vector of the finite set. Cybenko and Funahashi introduce the compact set of topology to approximate feed-forward neural networks. This continuous function is combined with the sigmoidal activation function for performing any classification task [13] [14]. Hornik provides the measurable function for a single-layer feed-forward neural network (SLFN) along with a comprehensive analysis. Besides, many research articles propose new capabilities to handle the multivariate functions for SLFN approximation [15]. Lan et al used complex domain mapping techniques from the input samples of the large class to handle the conventional parameter issues [16]. Several challenges are raised through theoretical sources of neural network algorithm. Almost all are related to the learning capabilities of networks present within the datasets [17]. Huang et al deal with SLFN for nonlinear complex function. They can learn N nodes for N distinct samples. The back propagation is used for training multilayer feed-forward neural networks. Despite the hype, performance-wise these algorithms are relatively slow learning algorithms during the training period and a problem arises with the local minima. The support vector machine algorithm is used for increasing the learning capabilities within the short time duration on the trained dataset. This time-intensive process is very suitable for SLFN with its parameters [18].

Huang et al proposed many traditional approaches with ELM to tune their network to achieve a faster response. The architecture is required to tune the input weights and biases to concise it. Besides, the ELM contains the smallest training error based on the norms of weights [19] [20]. Bartlett proposed a better generalization performance for feed-forward networks due to the least training error observed in the networks. A good generalization performance has been achieved in the network [21]. Cheng et al focus on classifying the imbalanced classes based on a weighted factor of ELM. Generally, these weighted factors of ELM are used in many applications such as intrusion detection, and also in any type of crime classification. They achieved the improved accuracy in intrusion detection applications [22]. Singh et al believe the feed-forward network includes three stages such as feature selection, network tuning with calculation, and classification decision. They monitor the hidden nodes and change based on their hidden layer activities. This comprehensive approach reduces the training period for many applications [23]. Singh et al worked on intrusion detection algorithms by using online sequential ELM (OS-ELM) for classification purpose. They solved the high false-positive rate in detection problems due to the large data volume. They implemented an improved version of linear discriminant analysis based ELM classification for the detection algorithm. The spatial similarity function is used for performing spatial separation and dimensionality reduction [24].

Manzoor et al introduce a dimension reduction algorithm with an artificial neural network. They have also proposed to determine the feature extraction for correlation function, which can be combined to compute features and further information for the classification task. They have eliminated redundancy in the datasets to optimize the algorithm [25].

Zhang et al applied the proposed algorithm into the application by using a probabilistic neural network (PNN). They believe in a feed-forward neural network and further it has been compared with a backpropagation (BP) neural network algorithm. They have shown good training time when compared with the BP neural networks [26]. Brown et al proposed many performance detection metrics to determine the superiority of the algorithm among the neural networks. They were used detection rate determination, false detection rate, computational complexity parameters, accuracy, and efficacy of the proposed architectures. Also, they have achieved good accuracy among the neural networks [27].

4. METHODOLOGIES

The sigmoidal modification can be made possible through Moore Penrose generalization inverse. The activation function changes through the orthogonal projection method in order to determine the output weight vectors in the matrix form. Figure 3 shows internal architecture of ELM.

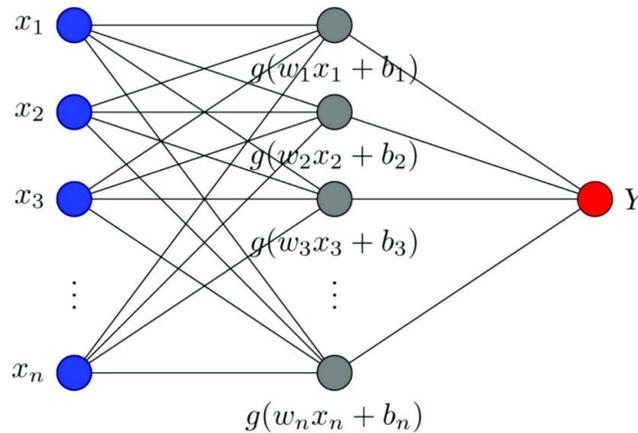


Figure 3 Internal Architecture of ELM

Design of Improved Version Algorithm

Step 1

Initial condition on ELM algorithm,

Randomly assign input weight (w_i) and bias (b_i) where $i=1,2,\dots,N$

Step 2

Calculate the hidden layer input / output matrix H.

Step 3

Calculate the output weight β by $\beta = H^\dagger T$, here H^\dagger is inverse matrix of H $T = (t_1, t_2, \dots, t_n)^T$

Step 4

The hidden layer output = $g(w_l + x_m + b_n)$ where, $l = 1,2,\dots,N$; $m = 1,2,\dots,N$; $n = 1,2,\dots,N$

Step 5

For “n” distinct vectors,

$$x_1 < x_2 < \dots < x_n \in \mathbb{R}^d (d > 1)$$

This exists vector $w \in \mathbb{R}^d$ such that,

$$w \cdot x_1 < w \cdot x_2 < \dots < w \cdot x_n$$

Step 6

$\sigma(x)$ is a bounded sigmoidal function

$x_i \in \mathbb{R}^d, b_i \in \mathbb{R}, i = 1, 2 \dots N$ Such that the matrix,

$$H = \begin{pmatrix} g(w_1 \cdot x_1 + b_1) & \dots & g(w_N \cdot x_1 + b_N) \\ \vdots & \ddots & \vdots \\ g(w_1 \cdot x_N + b_1) & \dots & g(w_N \cdot x_N + b_N) \end{pmatrix}$$

Modification:

Given a training data should be improved through $g(x)$,

$$\mathcal{N} = \{(x_i^*, t_i^*) \mid x_i^* \in \mathbb{R}^d, t_i^* \in \mathbb{R}, i = 1, 2 \dots N\}$$

Activation function of sigmoidal function for instance,

$$g(x) = \frac{1}{1 + e^{-x}}$$

Remarks:

The target function is to measure the performance,

$$y = f(x) = \begin{cases} \sin(x)/x & x \neq 0, \\ 1 & x = 0 \end{cases}$$

These basic remarks change in the similarity measures function of high dimensional data in the datasets to measure the weights and bias of the hidden nodes. This basic idea helps to maximize the class and create an optimal transformation matrix to reduce the computation time and dimension of the dataset [28] [29]. This can be incorporated with the ELM classification procedure to provide better accuracy and computation time.

5. RESULTS & DISCUSSION

The improved ELM can be performed well in the classification task. The UCI repository datasets of the machine learning databases are used in the proposed examination. The testing speculation is tabulated in table 1. From the dataset, an 80-20% ratio is used for training and testing respectively in every validation trial [30] [31].

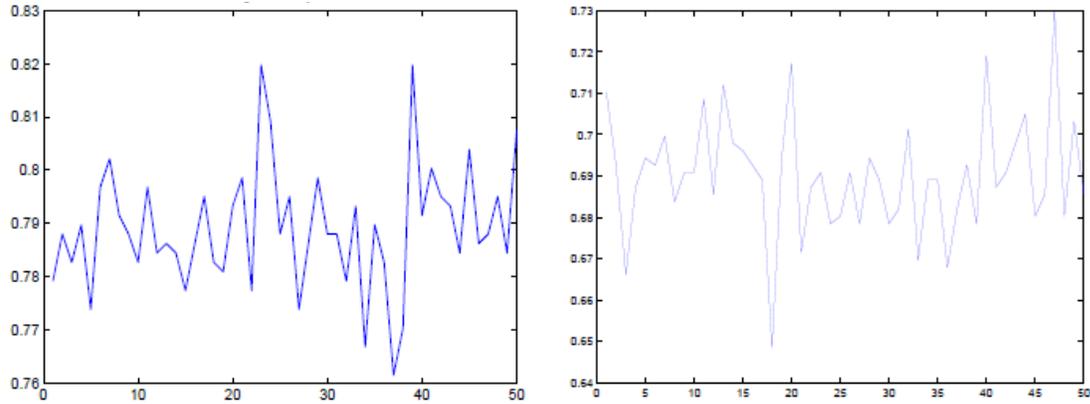


Figure 4 Training Accuracy of ELM and Optimized Improved ELM

This validation has reduced the random error in the hidden layer present in the network structure. The normal result of OS-ELM is varied from the proposed ELM algorithm in the obtained simulation results. The training accuracy of the proposed version of ELM is shown in the figure 4. Figure 5 shows the sigmoidal function of the ELM algorithm.

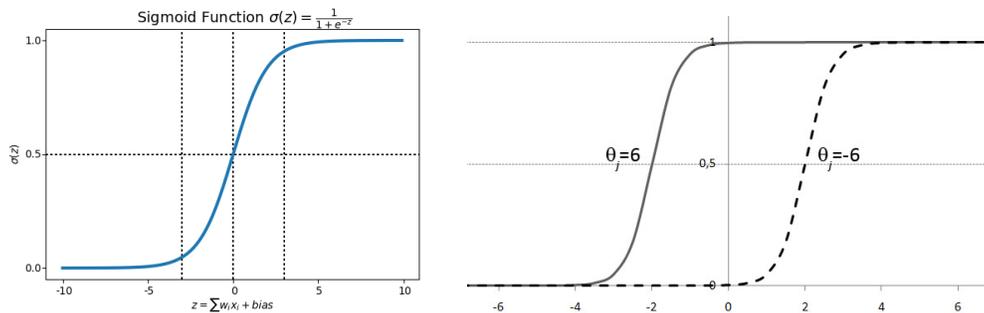


Figure 5 Sigmoidal Function of the ELM Algorithm

More trials are conducted between the two algorithms that have similar database. Figure 4 shows the training accuracy graph of ELM and the improved version of ELM at the huge amount of hidden nodes present during the examination. From this figure 4, it is observed that the training accuracy has improved and the error has minimized. The modified training and testing results is shown in the figure 6.

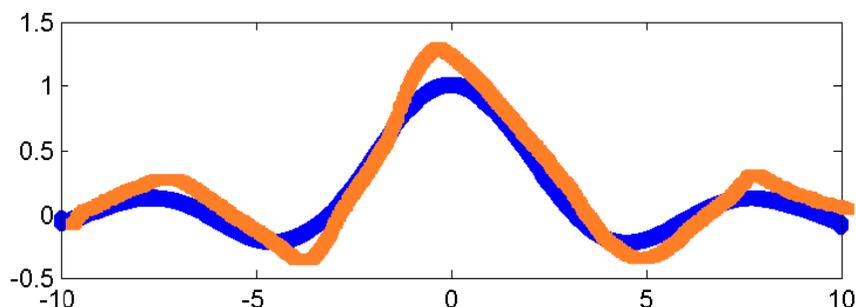


Figure 6 Training and testing results of ELM

The simulation results are reported in the table. Though they both have better learning rates; the proposed ELM algorithm is found to be superior to another one in terms of learning rate and accuracy.

Table 1 Overall Performance of Proposed Algorithm

S.NO	MODEL	TOTAL NODES	TIME		ACCURACY	
			Training	Testing	Training	Testing
1	Stacked Auto Encoder	20	20532	19345	65%	63%
2	ELM algorithm		2567.2	1905	81%	76%
3	Improved ELM algorithm		700.1	600	95.2%	93%
4	Stacked Auto Encoder	10	15932	14021	75%	72%
5	ELM algorithm		512.2	450	90%	86%
6	Improved ELM algorithm		447.1	399	97.2%	95%

The computation time has also been reduced based on the network complexity along with the input weight vectors. But the proposed ELM addresses the basic regression problem with good training accuracy results. The overall performance measure of the proposed work is shown in the figure 7.

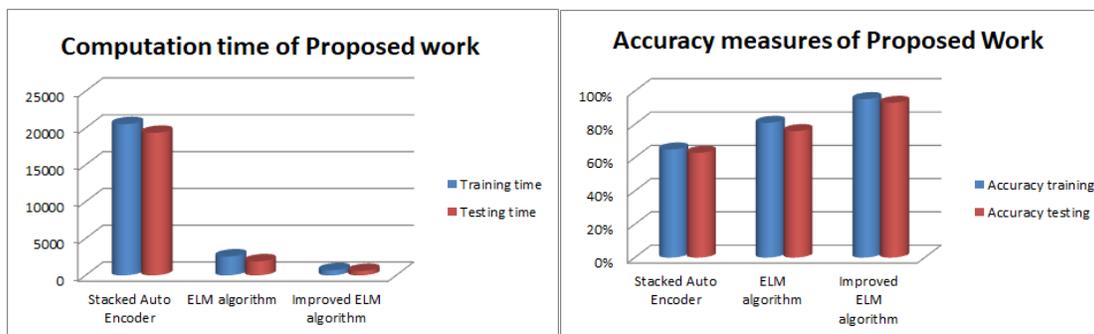


Figure 7 Performance Measures of Proposed Work

6. CONCLUSION

Thus, the proposed framework has solved the slow training and over-fitting problems. Besides, it provides better model accuracy and efficiency for the classification task. The basic regression problems are solved by using least-square minimization, which provides more effective output weight vectors for layer-to-layer. The selection of input weights and biases is random in an improved version of ELM. Also, the proposed model consumes less time for training time of weight vectors. This random selection of input weights and biases is avoided by using the singular column rank matrix. This phenomenon can calculate the speedy process of training procedure with orthogonal projection. This function can be ensemble with various function activities to train the networks, which give better stability and speed of the process. The number of samples in the network structure is combined with the feature vectors of the sample inputs to extend the modified ELM architecture. The dimension reduction still exists as our future work with high-dimensional data. The training time is very high since the proposed system specification is very limited. Therefore, our future work is implemented in good specification system. The effectiveness of the proposed system can be improved to further process the classification task. Besides, the proposed system will perform well in the IoT application due to its less training time. The implementation of the proposed system can be carried out in the IoT devices as future work.

REFERENCES

- [1] G. B. Huang, L. Chen, “Enhanced random search based incremental extreme learning machine,” *Neurocomputing*, vol. 71, pp. 3460–3468, 2008.
- [2] G. B. Huang, L. Chen, C. K. Siew, “Universal approximation using incremental constructive feedforward networks with random hidden nodes,” *IEEE Transactions on Neural Networks*, vol. 17, pp. 879–892, 2006.
- [3] Hou, H.R.; Meng, Q.H.; Zhang, X.N. “A voting-near-extreme-learning-machine classification algorithm” In *Proceedings of the 2018 24th International Conference on Pattern Recognition (ICPR)*, Beijing, China, 20–24 August 2018; pp. 237–241.

- [4] J. Zhang, W. Xiao, and Y. Li, "Residual compensation extreme learning machine for regression," *Neurocomputing*, vol. 311, pp. 126_136, Oct. 2018.
- [5] Z. L. Sun, T. M. Choi, and K. F. Au, "Sales forecasting using extreme learning machine with applications in fashion retailing," *Decis. Support Syst.*, vol. 46, no. 1, pp. 411_419, 2009.
- [6] Y. Lan, Y. C. Soh, and G.-B. Huang, "Ensemble of online sequential extreme learning machine," *Neurocomputing*, vol. 72, nos. 13_15, pp. 3391_3395, Aug. 2009.
- [7] J. Cao, Z. Lin, and G. B. Huang, "Voting based extreme learning machine," *Inf. Sci.*, vol. 185, no. 1, pp. 66_77, 2012.
- [8] M. Heeswijk, Y. Miche, and T. Lindh-Knuutila, "Adaptive ensemble models of extreme learning machines for time series prediction," in *Proc. Int. Conf. Artif. Neural Netw.*, Berlin, Germany: Springer-Verlag, 2009, pp. 305_314.
- [9] S. M. Yang, Y. L. Wang, and B. Sun, "ELM weighted hybrid modelling and its online modification," in *Proc. Control Decis. Conf.*, 2016, pp. 3443_3448.
- [10] J. Cao, S. Kwong, and R. Wang, "Class-specific soft voting based multiple extreme learning machines ensemble," *Neurocomputing*, vol. 149, pp. 275_284, Feb. 2015.
- [11] M. Rahhal, Y. Bazi, and N. Alajlan, "Classification of AAMI heartbeat classes with an interactive ELM ensemble learning approach," *Biomed. Signal Process., Control*, vol. 19, pp. 56_67, May 2015.
- [12] Xu, X.; Li, S.; Liang, T.; Sun, T. Sample selection-based hierarchical extreme learning machine. *Neurocomputing* 2020, 377, 95–102.
- [13] G. Cybenko, "Approximation by superposition of sigmoidal function," *Mathematics of Control, Signals and Systems*, vol. 2, pp. 303–314, 1989.
- [14] K. I. Funahashi, "On the approximate realization of continuous mappings by neural networks," *Neural Networks*, vol. 2, pp. 183–192, 1989.
- [15] K. Hornik, "Approximation capabilities of multilayer feedforward networks," *Neural Networks*, vol. 4, pp. 251–257, 1991.
- [16] Y. Lan, Y. C. Soh, G. B. Huang, "Random search enhancement of error minimized extreme learning machine," *European Symposium on Artificial Neural Networks*, pp. 327–332, April 2010.

- [17] F. L. Cao, S. B. Lin, Z. B. Xu, "Approximation capabilities of interpolation neural networks," *Neurocomputing*, vol. 74, pp. 457–460, 2010.
- [18] G. B. Huang, H. A. Babri, "Upper bounds on the number of hidden neurons in feedforward networks with arbitrary bounded nonlinear activation functions," *IEEE Transactions on Neural Networks*, vol. 9, pp. 224–229, 1998.
- [19] G. B. Huang, Q. Y. Zhu, C. K. Siew, "Extreme learning machine: theory and applications," *Neurocomputing*, vol. 70, pp. 489–501, 2006.
- [20] G. B. Huang, Q. Y. Zhu, C. K. Siew, "Extreme learning machine: a new learning scheme of feedforward neural networks," *Proceedings of 2004 IEEE International Joint Conference on Neural Networks*, vol. 2, pp. 985–990, July 2004.
- [21] P. L. Bartlett, "The sample complexity of pattern classification with neural networks: the size of the weights is more important than the size of the network," *IEEE Transactions on Information Theory*, vol. 44, pp. 525–536, 1998.
- [22] Cheng, C.; Tay, W.P.; Huang, G.B. Extreme learning machines for intrusion detection. In *Proceedings of the 2012 International Joint Conference on Neural Networks (IJCNN)*, Brisbane, QLD, Australia, 10–15 June 2012; pp. 1–8.
- [23] Singh, D.; Bedi, S.S. Multiclass ELM based smart trustworthy IDS for MANETs. *Arab. J. Sci. Eng.* **2016**, 41, 3127–3137.
- [24] Singh, R.; Kumar, H.; Singla, R.K. Performance analysis of an intrusion detection system using Panjab university intrusion dataSet. In *Proceedings of the 2015 2nd International Conference on Recent Advances in Engineering & Computational Sciences (RAECS)*, Chandigarh, India, 21–22 December 2015; pp. 1–6.
- [25] Manzoor, I.; Kumar, N. A feature reduced intrusion detection system using ANN classifier. *Expert Syst. Appl.* **2017**, 88, 249–257.
- [26] Zhang, M.; Guo, J.; Xu, B.; Gong, J. Detecting network intrusion using probabilistic neural network. In *Proceedings of the 2015 11th International Conference on Natural Computation (ICNC)*, Zhangjiajie, China, 15–17 August 2015; pp. 1151–1158.
- [27] Brown, J.; Anwar, M.; Dozier, G. "An evolutionary general regression neural network classifier for intrusion detection". In *Proceedings of the 2016 25th International Conference on*

Computer Communication and Networks (ICCCN), Waikoloa, HI, USA, 1–4 August 2016; pp. 1–6.

[28] M. Eshtay, H. Faris, and N. Obeid, “Improving Extreme Learning Machine by Competitive Swarm Optimization and its application for medical diagnosis problems,” *Expert Syst. Appl.*, vol. 104, pp. 134–152, 2018.

[29] P. P. Das, R. Bisoi, and P. K. Dash, “Data decomposition based fast reduced kernel extreme learning machine for currency exchange rate forecasting and trend analysis,” *Expert Syst. Appl.*, vol. 96, pp. 427– 449, 2018.

[30] P. Yuan, D. Chen, T. Wang, S. Cao, Y. Cai, and L. Xue, “A compensation method based on extreme learning machine to enhance absolute position accuracy for aviation drilling robot,” *Adv. Mech. Eng.*, vol. 10, no. 3, p. 1687814018763411, 2018.

[31] G.-B. Huang, H. Zhou, X. Ding, and R. Zhang, “Extreme learning machine for regression and multiclass classification” *IEEE Trans. Syst. Man. Cybern. B. Cybern.*, 2012.