A Review of Advances in Extreme Learning Machine Techniques and Its Applications

Oyekale Abel Alade[1,2], Ali Selamat[1(✉)], and Roselina Sallehuddin[1]

[1] Faculty of Computing, Universiti Teknologi Malaysia, Skudai, Johor, Malaysia
kaleabel@gmail.com, {aselamat, roselina}@utm.my
[2] Computer Science Department, Federal Polytechnic, Bida, Nigeria

Abstract. Feedforward neural networks (FFNN) has been used for machine learning researches, and it really has a wide acceptance. It was noted in the recent time that feedforward neural network is far slower than required. This has created a serious bottleneck in its applications. Extreme Learning Machines (ELM) had been proposed as alternative learning algorithm to FFNN, which is characterized by single-hidden layer feedforward neural networks (SLFN). It randomly chooses hidden nodes and determines their output weight analytically. This paper review is to provide a roadmap for ELM as an efficient research tool in machine learning with the aim of finding research gap into further study. It was discovered through this study that research publications in ELM continues to grow yearly from 16.20% in 2013 to 40.83% in 2016.

Keywords: Activation functions · Classification · Extreme learning machines · Single layer feedforward neural network

1 Introduction

Machine learning is a science of making computers to respond to changes without being explicitly programmed. It is a branch of artificial intelligence that focuses on automatic program execution change when exposed to new data. Machine learning studies patterns in data and adjusts program actions accordingly. Machine learning has been used over the year to accomplish many natural tasks like self-driving cars, speech recognition, effective web search, understanding human genome, and many more. Machine learning has employed the capabilities of multilayer feedforward neural network (MLFN). Single layer feedforward neural network has been adjudged by researchers as having the ability to learn effectively with tolerable errors using N hidden neurons and any activation function [1]. In almost all the learning algorithms of feedforward neural network, the input weights and biases need to be iteratively tuned using gradient descent training algorithm. Feedforward neural network is however far slower than required, and this has created a serious bottleneck in its applications.

Extreme learning machines (ELM) are naturally feedforward neural networks. The word "Extreme" is synonymous to a move from conventional learning technique towards brain-like learning [2, 3]. They are three-steps training algorithms that are simple and they require no tuning like the normal single layer feed forward neural

F. Saeed et al. (eds.), *Recent Trends in Information and Communication Technology*, Lecture Notes on Data Engineering and Communications Technologies 5, DOI 10.1007/978-3-319-59427-9_91

network [4]. ELMs randomly assign the connection weights of the hidden neurons and then tune the connections with output neurons ones. The machines usually have single layer of hidden nodes, and when the weights between the input and the hidden nodes are randomly assigned, they are never updated. The weights between the hidden nodes and the output neuron(s) are learned in a single step. Therefore, there is no more dependency between the parameters (weights and biases) of the hidden layer as it was in the traditional feedforward networks that necessitated the tuning. The principle of learning is essentially a linear model. The focus of ELM theories is to address the tuning problem that has plagued previous machine learning and neuroscience research work and applications. ELM models have produced good generalization performance, and learn thousands of times faster than networks trained using backpropagation (BP).

This paper review is to provide a roadmap for ELM as an efficient research tool in machine learning with the aim of finding research gaps into further study. The rest of this paper is organised as follows: Sect. 2 discussed classical ELM in brief. Section 3 discussed the research method. In Sect. 4, results and discussion of findings in response to research questions were presented. It includes the various ELM techniques, the strengths and weaknesses of ELM, and its applications. Section 5 concluded the paper.

2 Classical Extreme Learning Machines

ELM came on board to address the problem of iterative tuning in SLFNN which was associated with earlier gradient descent techniques of machines learning. The hidden layer may not necessarily be neurons alike [3]. The connections weights between the input and the hidden neurons can be called input weights, and the connection weights between the hidden neurons and the output neuron(s) as the output weight(s).

Let a set of input data be defined as $D = \{(x_k, y_k)\}$; $k = 1, 2, \ldots, N$. (N and the size).

Where $x_k = [x_{k1}, x_{k2}, \ldots x_{kM}]$ is the k^{th} set of input vector of dimension $_M$, and $y_k = [y_{k1}, y_{k2}, \ldots, y_{kL}]$ is a set of output vector of dimension $_L$.

The model for SLFN with n hidden neurons is given in Eq. (1)

$$\tilde{y} = \sum_{i=1}^{n} v_i g_i(x_k) = \sum_{i=1}^{n} v_i g(w_i x_k + b_i); k = 1, 2, \ldots, N \quad (1)$$

The vector $w_i = [w_{i1}, w_{i2}, \ldots, w_{iM}]^T$ in Eq. (1) is the i^{th} hidden neuron's input weights, i.e. weights that define the connections between M input neurons and the i^{th} hidden neuron, the vector $v_i = [v_{i1}, v_{i2}, \ldots, v_{iM}]^T$ represents the i^{th} hidden neuron output weights, i.e. the weights that define the connections between i^{th} hidden neuron and the L output neurons, b_i represents the bias of the hidden neurons i, $g_i(x_i)$ represents a suitable activation function of the i^{th} hidden neuron, $w_i x_k$ is the inner product of the input vector x_k and the connection weights w_i between the input layer and the hidden neuron.

Activations Functions
During the initialization, ELM maps the input data into the "ELM future space" [3] by nonlinear mapping functions. These functions can be nonlinear piecewise continuous function or any of those in Table 1.

Table 1. Some mapping functions in ELM

Sigmoid function	$G(a,b,x) = \frac{1}{1+\mathbf{exp}(-a.x+b))}$
Hyperbolic tangent function	$G(a,b,x) = \frac{1-\mathbf{exp}(-a.x+b))}{1+\mathbf{exp}(-a.x+b))}$
Gaussian function	$G(a,b,x) = exp(-b\,\|x-a\|)$
Multiquadric function	$G(a,b,x) = \sqrt{((\|x-a\|^2 + b^2))}$
Hard limit function	$G(a,b,x) = \begin{cases} 1, \text{ if } a.x \leq 0 \\ 0, \text{ otherwise} \end{cases}$
Cosine function/Fourier basis	$G(a,b,x) = \mathbf{Cos}(\mathbf{a}.\mathbf{x}+\mathbf{b})$

3 Research Method

3.1 Research Questions

The aim of this study is to empirically identify the research gaps in ELM as a roadmap into further research. To achieve this aim, the following research questions are set:
RQ1: What are the existing techniques used in ELM?
RQ2: What are the major application areas of ELM?
RQ3: What are the strength and weaknesses of ELM?

3.2 Research Strategy

To attend to research question one RQ1, a research strategy was mapped out thus, (i) search strings were composed, (ii) the scope of the search was defined, and (iii) the search method was set. ELM was viewed in two ways, (a) as a single machine and (b) as combination of machines. Therefore, the string terms used are (i) Machine learning (ii) Extreme learning machine and (iii) Extreme learning machines, Table 2

Table 2. Review search strings

Concept	Search String
Extreme learning machine	Extreme learning **OR**
	Extreme learning machine **OR**
	Extreme learning machines

The search strings were extracted from titles, abstracts and keywords. The search was limited to only Web of Science Citation index services online for papers published between January 2013 and December 2016. Only journal articles and conference proceedings related to this research interest were considered, that is, computer science field.

3.3 Search Based on Strings and Scopes

Manual and automatic methods were adopted in our search for literature of interest on ELM and its variance. Web of science citation index service was used as our citation index service engine. We carried out automatic search for relevant literature on ELM within the scope of time frame set –2013 to 2016. To narrow down the automatic search through refinement, the citation indexes were limited to (i) science citation index expanded (SCI), (ii) conference proceedings citation index (CPCI-S) and (iii) emerging sources citation index (ESCI). The search category was broadly on computer science. By inclusion and exclusion, the selected area of computer science were (i) artificial intelligence (AI), (ii) theory and methods (T & M), (iii) information science (IS) and (iv) software engineering (SE). The schematic diagram of the search technique is shown in Fig. 1.

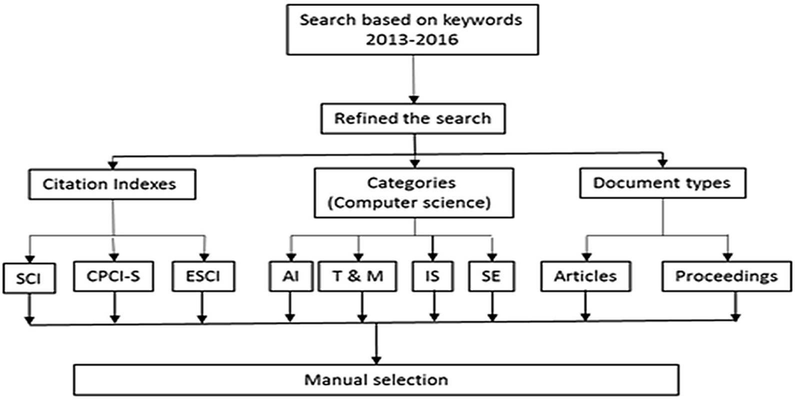

Fig. 1. Literature search framework

The various literature retrieved were screened with key terms in the search strings from the titles, abstracts and keywords in the journal articles; where the abstracts and keywords are poor to reflect the research focus of these papers, the introductions and methodologies of the papers were explored. Then, the references of the selected papers were also used to retrieve more relevant papers that fall within the scope of this review. After this, we refined the search papers based on exclusion and inclusion criteria defined thus:

Inclusion:

- Papers that discussed the trends of theories and algorithms of ELM and survey developments
- Papers that discussed major ELM applications
- Papers that have implemented ELM in one way or the other in a computer science domain.
- Papers that criticized or defended the criticism of ELM as learning theory.

Exclusion:

- Papers that only mentioned ELM search string in abstracts and keywords, and have no sufficient discussions on the subject.
- And papers that are not available in full text.
- Papers that contain the search strings but are not related to computer science.

3.4 Analysis of Search Result

Based on the inclusion and exclusions constraints set above, a total of 605 articles was retrieved. Out of this number of filtered records, fifty (51) papers were selected manually through critical reading to further trim down the size for this study.

The analysis of our findings from the literature retrieved for this study are discussed thus: Research interests continue to grow yearly as shown in Fig. 2. The percentage growth is on the increase from 16.20% in 2013 to in 2016 40.83%. Figure 3 shows that more papers were published as journal articles than conference proceedings (69% as against 31%). For interested researchers to know the leading authorities in this research domain, the names of authors who have published ten (10) papers or more are shown in Table 3. Lendasse led authors with 25 articles on ELM. This would equally enhance research collaboration in this field. Table 4 shows that Neuro-computing, with 247 publications which constituted 40.826% of papers retrieved has the highest number of publications on ELM. This will help authors to know where to publish as well as have quick access to research literature on ELM.

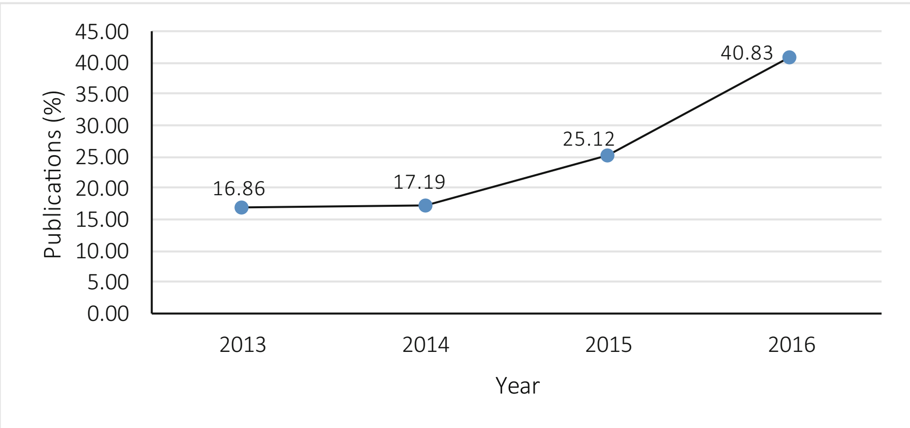

Fig. 2. Yearly publications on ELM

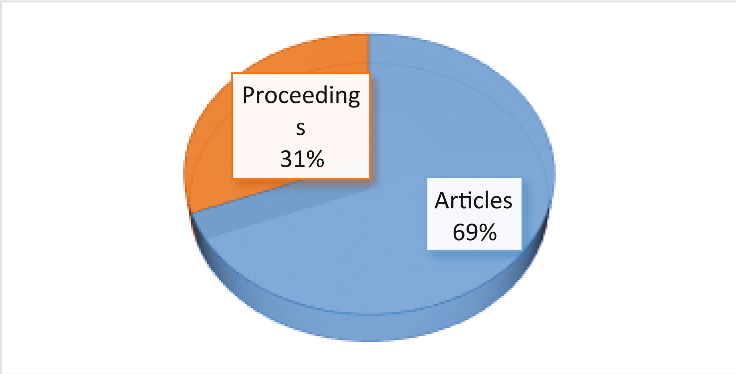

Fig. 3. Distribution of the retrieved papers by document types

Table 3. The 10 top leading authors in the retrieved papers on ELM

Major Authors	Records	%
LENDASSE A	25	4.1
WANG GR	18	3.0
MICHE Y	18	3.0
VONG CM	12	2.0
NIAN R	12	2.0
HUANG GB	12	2.0
WONG PK	11	1.8
IOSIFIDIS A	11	1.8
GRANA M	11	1.8

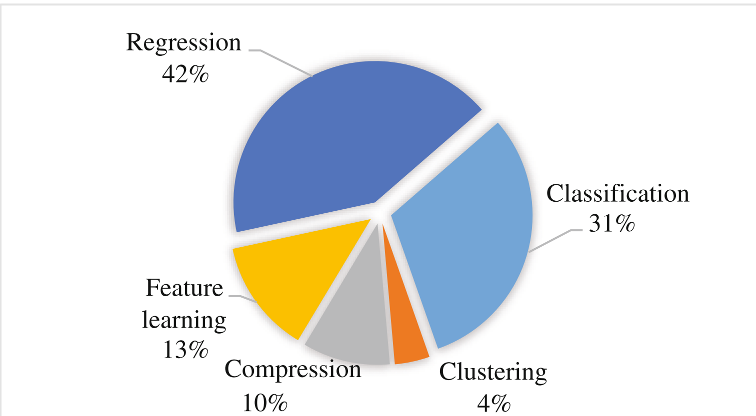

Fig. 4. Chart showing the distribution of application areas of ELM

Table 4. Source title records for 10 or more publications on ELM

Source	Records.	%
Neurocomputing	247	40.826
Neural computing applications	69	11.405
Proceedings of ELM theory algorithms and applications	40	6.612
Proceedings in adaptation learning and optimization	40	6.612
Lecture notes in computer science	31	5.124
Neural networks	22	3.636
Cognitive computation	22	3.636
Neural processing letters	17	2.810
International journal of uncertainty fuzziness and knowledge based systems	14	2.314
Pattern recognition letters	10	1.653

4 Results and Discussion

The review is discussed below in response to the research questions in Sect. 3.1.

4.1 ELM Techniques

As set in research question RQ1, the usefulness of ELM has brought about a number of improvements to the original algorithm. ELM principal focus is not on generalized single hidden SLFN only, but also on MLFN in which it may consists of sub-nodes of another nodes [2, 5]. This has led to many variants of ELM. Some variants of ELM in the literature reviewed are OP-ELM [6], Wavelet ELM [2, 7], KBELM [6], H-ELM, [8], FASTS-ELM [9], FP-ELM [10], Graph embedded-ELM (GE-ELM) [11], Incremental-ELM [12], Improved convex incremental-ELM [7], KB-ELM [6], Multiple-Kernel-ELM [13], DrELM [14], SMOTE based OS-ELM [15], Inverse-free-ELM [16] and many others. Most of these ELM techniques and their algorithms are discussed in more details in [3]. The variants have improved its stability and compactness, to the extent that it had enabled it to be used for online sequential data [17], imbalanced data, noisy/missing data and to solve the problem of over-fitting [3]

4.2 Major Applications of ELM

Answering the second research question RQ2; it was observed that the efficiency of ELM in machine learning has arouse the interest of many researchers and industries to venture into its various theories and applications in the recent years. Although there were various applications of ELM in literature, and various variants of it exist as improved or hybrid versions, five widely reported applications of ELM are considered in this review. In [18, 19], and [20] the theory and applications of ELM was extensively discussed. ELM has been used majorly for compression [2, 3, 21], feature learning [2, 8, 22], clustering, regression [16] and classification [5, 22, 23]. For the purpose of this study, a review into applications of ELM based on this broad classification is shown in Fig. 4. The result shows that regression has the highest attraction of research interests with 42%, while clustering is the least favoured researched area with a percentage of only 4. ELM applications, among many others, are Big data [8], sediment transportation [24], streamflow forecasting [25], rainfall runoff [17, 26] and many more. Other applications of ELM contained in [3] are (i) computational vision, (ii) image/video understanding, (iii) text classification and understanding, (iv) system modelling prediction, (v) control and robotics, (vi) fault detection and diagnosis, (vii) chemical process, (viii) time series analysis and (ix) remote sensing. While [27] mentioned only three ELM application areas as (i) classification, (ii) regression and (iii) function approximation. [27] also listed some variants of ELM, their descriptions and applications.

4.3 Strength and Weaknesses of ELM

In response to research question Q3, it was observed that a good number of literature reviewed have high regard for ELM as a good learning machine tool. The major strength of ELM is that the learning parameters of the hidden layer, including the input weights and biases, need not be tuned iteratively as in SLFN [1, 18–20, 24, 27–29]. This major characteristic feature offers ELM a faster speed and cost reduction [3], and it is more favour in machine learning than any of its predecessors. Other commendable attributes of ELM are simple learning algorithm [30], good generalization performance and accuracy [3], improved efficiency [18, 27], a unified solution to different practical applications [2], non-linear transformation in its training phase, devoid of local minimal and overfitting [2, 18, 29], requires fewer optimization when compared with SVM and LS-SVM, though it has the same computational cost with SVM [30]. More importantly, ELM is filling the gap between conventional learning machines and biological learning machine [3] which is the objective of Hang [1] (cited in [3]) who is a pioneer researcher of ELM.

Despite all the numerous advantages that ELM has to its credit, there exist some issues of imperfection to researchers. It has been observed that the classification boundary for learning parameters of hidden layer may not be optimal because they do not change during training [21, 27]. ELM cannot handle Big high dimensional data [3, 23], it requires more hidden nodes than conventionally tuning algorithms, and cannot be parallelized due to the presence of pseudo-inverse circulation [31].

Although, some of these challenging issues had been attempted in some variants of modifications, hybridizations and optimizations of ELM, many current literature still have the following recommended area for further research: (i) theoretical prove and implementation of optimal number of hidden nodes, (ii) estimation of generalization performance (iii) generalization capability to handle high dimensional data [8, 10, 31, 32] (iv) adjustment of ELM algorithm for parallel and distributed computation [27, 33]. Hang in [34] emphasized the need to study the link between ELM and other related algorithm like random forest algorithm.

5 Conclusion

This paper review has gleaned into some literature as a window into research gaps in Extreme Learning Machines (ELM). To achieve the aim of the review, ELM theory has been briefly introduced, and literature were carefully selected in some specific areas of computer science, and it was clearly shown that there is a growing interest in ELM. Research work has grown steadily from 16.20% in 2013 to 40.83% in 2016 as clearly shown in the number of papers being published yearly. Various applications of ELM were observed. The study revealed that about 42% of the total research work are concentrated on regression while clustering has only 4%. The strengths and weaknesses were pointed out. It is observed that the relevance of ELM in artificial intelligent has brought about a great research interest to it in recent years. Despite these research efforts in ELM, there are still call on researchers for more generalization of ELM technique, viz-viz its handling of high dimensional data.

Acknowledgements. The authors wish to thank Universiti Teknologi Malaysia (UTM) under Research University Grant Vot-02G31 & Vot-15H17 and Ministry of Higher Education Malaysia (MOHE) under the Fundamental Research Grant Scheme (FRGS Vot-4F551) for the completion of the research.

References

1. Huang, G., Zhu, Q., Siew, C.: Extreme learning machine: a new learning scheme of feedforward neural networks. In: IEEE International Joint Conference Neural Networks, vol. 2, pp. 985–990 (2004). doi:10.1109/IJCNN.2004.1380068
2. Bin, H.G.: What are extreme learning machines? Filling the gap between Frank Rosenblatt's dream and John von Neumann's puzzle. Cognit. Comput. **7**, 263–278 (2015). doi:10.1007/s12559-015-9333-0
3. Huang, G., Bin, H.G., Song, S., You, K.: Trends in extreme learning machines: a review. Neural Networks **61**, 32–48 (2015). doi:10.1016/j.neunet.2014.10.001
4. Lazarevska, L.: Wind speed prediction with extreme learning machine, pp. 154–159 (2016)
5. Yang, Y., Wu, Q.M.J., Member, S.: Extreme learning machine with subnetwork hidden nodes for regression and classification. IEEE Trans. Cybern. **46**, 2885–2898 (2016)
6. Balasundaram, S., Gupta, D.: Knowledge-based extreme learning machines. Neural Comput. Appl. **27**, 1629–1641 (2016). doi:10.1007/s00521-015-1961-5
7. Musikawan, P., Sunat, K., Chiewchanwattana, S., et al.: Improved convex incremental extreme learning machine based on ridgelet and PSO algorithm (2016)
8. Deng, W.Y., Bai, Z., Bin, H.G., Zheng, Q.H.: A fast SVD-hidden-nodes based extreme learning machine for large-scale data analytics. Neural Networks **77**, 14–28 (2016). doi:10.1016/j.neunet.2015.09.003
9. Mahmood, S.F., Marhaban, M.H., Rokhani, F.Z., et al.: FASTA-ELM: a fast adaptive shrinkage/thresholding algorithm for extreme learning machine and its application to gender recognition. Neurocomputing (2016). doi:10.1016/j.neucom.2016.09.046
10. Liu, D., Wu, Y.X., Jiang, H.: FP-ELM: an online sequential learning algorithm for dealing with concept drift. Neurocomputing **207**, 322–334 (2015). doi:10.1016/j.neucom.2016.04.043

11. Iosifidis, A., Tefas, A., Pitas, I.: Graph embedded extreme learning machine. IEEE Trans. Cybern. **46**, 311–324 (2016). doi:10.1109/TCYB.2015.2401973
12. Zhang, J., Ding, S., Zhang, N., Shi, Z.: Incremental extreme learning machine based on deep feature embedded. Int. J. Mach. Learn. Cybern. **7**, 111–120 (2016). doi:10.1007/s13042-015-0419-5
13. Liu, X., Wang, L., Huang, G.-B., et al.: Multiple kernel extreme learning machine. Neurocomputing **149**, 253–264 (2015). doi:10.1016/j.neucom.2013.09.072
14. Yu, W., Zhuang, F., He, Q., Shi, Z.: Learning deep representations via extreme learning machines. Neurocomputing **149**, 308–315 (2015). doi:10.1016/j.neucom.2014.03.077
15. Mao, W., Wang, J., Wang, L.: Online sequential classification of imbalanced data by combining extreme learning machine and improved SMOTE algorithm. In: Proceeding of the International Joint Conference on Neural Networks (2015). doi:10.1109/IJCNN.2015.7280620
16. Li, S., You, Z., Guo, H., et al.: Inverse-Free extreme learning machine with optimal information updating. IEEE Trans. Cybern. **46**, 1229–1241 (2016)
17. Yadav, B., Ch, S., Mathur, S., Adamowski, J.: Discharge forecasting using an online sequential extreme learning machine (OS-ELM) model: a case study in Neckar River, Germany. Meas. J. Int. Meas. Confed. **92**, 433–445 (2016). doi:10.1016/j.measurement.2016.06.042
18. Huang, G.-B., Zhu, Q., Siew, C., et al.: Extreme learning machine: theory and applications. Neurocomputing **70**, 489–501 (2006). doi:10.1016/j.neucom.2005.12.126
19. Liu, X., Lin, S., Fang, J., Xu, Z.: Is extreme learning machine feasible? A theoretical assessment (part I). IEEE Trans. Neural Netw. Learn. Syst. **26**, 7–20 (2015). doi:10.1109/TNNLS.2014.2335212
20. Liu, X., Lin, S., Fang, J., Xu, Z.: Is extreme learning machine feasible? A theoretical assessment (Part II). IEEE Trans. Neural Netw. Learn. Syst. **26**, 7–20 (2015). doi:10.1109/TNNLS.2014.2335212
21. Cao, J., Lin, Z., Bin, H.G., Liu, N.: Voting based extreme learning machine. Inf. Sci. (Ny) **185**, 66–77 (2012). doi:10.1016/j.ins.2011.09.015
22. Hu, X., Lin, H., Li, S., Sun, B.: Global and local features based classification for bleed-through removal. Sens. Imaging **17**, 9 (2016). doi:10.1007/s11220-016-0134-7
23. Zhang, J., Feng, L., Wu, B.: Local extreme learning machine: local classification model for shape feature extraction. Neural Comput. Appl. **27**, 2095–2105 (2016). doi:10.1007/s00521-015-2008-7
24. Ebtehaj, I., Bonakdari, H., Shamshirband, S.: Extreme learning machine assessment for estimating sediment transport in open channels. Eng. Comput. **32**, 1–14 (2016). doi:10.1007/s00366-016-0446-1
25. Mundher Yaseen, Z., Jaafar, O., Deo, R.C., et al.: Boost stream-flow forecasting model with extreme learning machine data-driven: a case study in a semi-arid region in Iraq. J. Hydrol. **542**, 603–614 (2016). doi:10.1016/j.jhydrol.2016.09.035
26. Badrzadeh, H., Sarukkalige, R., Jayawardena, A.W.: Hourly runoff forecasting for flood risk management: application of various computational intelligence models. J. Hydrol. (2015). doi:10.1016/j.jhydrol.2015.07.057
27. Ding, S.F., Xu, X.Z., Nie, R.: Extreme learning machine and its applications. Neural Comput. Appl. **25**, 549–556 (2014). doi:10.1007/s00521-013-1522-8
28. Huang, G.-B., Zhou, H., Ding, X., Zhang, R.: Extreme learning machine for regression and multiclass classification. IEEE Trans. Syst. Man Cybern. Part B Cybern. **42**, 513–529 (2012). doi:10.1109/TSMCB.2011.2168604
29. Sakakura, Y.: Extreme Learning Machine (ELM), pp. 1–14 (2013)

30. Zhang, L., Li, J., Lu, H.: Saliency detection via extreme learning machine. Neurocomputing **218**, 103–112 (2016). doi:10.1016/j.neucom.2016.08.066
31. Oneto, L., Bisio, F., Cambria, E., Anguita, D.: Statistical learning theory and ELM for big social data analysis. IEEE Comput. Intell. Mag. **11**, 45–55 (2016). doi:10.1109/MCI.2016.2572540
32. Wang, H., Xu, Z., Pedrycz, W.: An overview on the roles of fuzzy set techniques in big data processing Trends, challenges and opportunities. Knowledge-Based Syst **118**, 1–16 (2016). doi:10.1016/j.knosys.2016.11.008
33. Bodyanskiy, Y., Vynokurova, O., Pliss I, et al.: Fast learning algorithm for deep evolving GMDH-SVM neural network in data stream mining tasks, pp. 257–262 (2016)
34. Bin, H.G.: An insight into extreme learning machines: random neurons, random features and kernels. Cognit. Comput. **6**, 376–390 (2014). doi:10.1007/s12559-014-9255-2