UNIVERSITI PUTRA MALAYSIA

ENHANCED EXTREME LEARNING MACHINE FOR GENERAL REGRESSION AND CLASSIFICATION TASKS

SAIF F MAHMOOD

FK 2020 97

ENHANCED EXTREME LEARNING MACHINE FOR GENERAL REGRESSION AND CLASSIFICATION TASKS

By

SAIF F MAHMOOD

Thesis Submitted to the School of Graduate Studies, Universiti Putra Malaysia, in Fulfilment of the Requirements for the Degree of Doctor of Philosophy

September 2020

DEDICATION

Dedicated to:

My family.

Abstract of thesis presented to the Senate of Universiti Putra Malaysia in fulfilment of the requirement for the degree of Doctor of Philosophy

ENHANCED EXTREME LEARNING MACHINE FOR GENERAL REGRESSION AND CLASSIFICATION TASKS

By

SAIF F MAHMOOD

September 2020

Chairman : Professor Mohammad Hamiruce Marhaban, PhD
Faculty : Engineering

Extreme Learning Machine (ELM) is a single hidden layer feedforward neural network which randomly chooses hidden nodes and analytically determines the output weights using least square method. Despite its popularity, ELM has a number of challenges worth to investigating for improving the usability of ELM in a more advanced application. This thesis focusses on challenges namely design architecture and learning technique. The first challenge is to select the optimal number of hidden nodes for ELM in different application. To address this problem, a new approach referred to SVM-ELM is proposed, which utilizes 1-norm support vector machine (SVM) to the hidden layer matrix of ELM in order to automatically discover the optimal number of hidden nodes. The method is developed for regression task by using mean/ median of ELM training errors which is then used as threshold for separating the training data and converting the continuous targets to binary. This will allow projection to 1-norm SVM dimension in order to find the best number of nodes as support vectors. Second problem in ELM, is the restriction in performance of ELM in terms of training time and model generalization, due to the complexity of singular value decomposition (SVD) for computing the Moore-Penrose generalized inverse of the hidden layer matrix, especially on a large matrix. To address this issue, a fast adaptive shrinkage/thresholding algorithm ELM (FASTA-ELM) which uses an extension of forward-backward splitting (FBS) to compute the smallest norm of the output weights in ELM is presented. The proposed FASTA-ELM replaces the analytical step usually solved by SVD with an approximate solution through proximal gradient method, which dramatically speeds up the training time and improves the generalization ability in classification task. The performance of FASTA-ELM is evaluated on face gender recognition problem and the result is comparable to other state-of-the-art methods, with significantly reduced training time. For instance, the training

time of 1000 nodes ELM is 18.11 s, while FASTA-ELM completed in 1.671 s. The proposed modification to the ELM shows significant improvement to the conventional ELM in terms of training time and accuracy, and provide good generalization performance in regression and classification tasks.

Abstrak tesis yang dikemukakan kepada Senat Universiti Putra Malaysia
sebagai memenuhi keperluan untuk ijazah Doktor Falsafah

MESIN PEMBELAJARAN EKSTRIM DIPERTINGKAT UNTUK TUGASAN REGRESI UMUM DAN PENGELASAN

Oleh

SAIF F MAHMOOD

September 2020

Pengerusi : Profesor Mohammad Hamiruce Marhaban, PhD
Fakulti : Kejuteraan

Mesin pembelajaran ekstrim (*Extreme Learning Machine*: ELM) adalah satu rangkaian neural lapisan tersembunyi tunggal suap ke depan, yang memilih nod tersembunyi secara rawak dan menentukan pemberat keluaran secara analitik dengan menggunakan kaedah kuasa dua terkecil. Walaupun begitu, ELM mempunyai cabaran yang perlu dikaji untuk meningkatkan kebolehgunaan ELM dalam aplikasi termaju. Thesis ini memberi tumpuan kepada dua cabaran utama iaitu reka bentuk seni bina dan teknik pembelajaran. Cabaran pertama adalah memilih bilangan nod tersembunyi yang optimum untuk ELM dalam pelbagai aplikasi. Untuk mengatasi masalah ini, pendekatan baru yang dinamakan sebagai SVM-ELM dicadangkan, yang menggunakan mesin vektor sokongan (*Support Vector Machine*: SVM) 1-norma terhadap lapisan tersembunyi matriks ELM untuk secara automatik mendapatkan bilangan nod tersembunyi yang optimum. Kaedah ini dibangunkan untuk tugasan regresi dengan menggunakan purata atau median ralat latihan ELM yang mana seterusnya ia digunakan sebagai titik ambang untuk memisahkan data latihan dan menukar sasaran berterusan ke binari. Ini membolehkan unjuran ke dimensi SVM 1-norma untuk mencari bilangan nod terbaik sebagai vektor sokongan. Masalah kedua dalam ELM adalah prestasi ELM yang terhad dari segi tempoh latihan dan generalisasi model, disebabkan kerumitan penguraian nilai tunggal (*Singular Value Decomposition*: SVD) untuk mengira pembalikan umum Moore-Penrose dari matriks lapisan tersembunyi, terutama bagi matriks yang besar. Untuk mengatasi masalah tersebut, algoritma penyusutan / titik ambang tersuai pantas (*Fast Adaptive Shrinkage/Tresholding Algorithm*: FASTA) ELM (FASTA-ELM) yang menggunakan lanjutan pemisahan maju-mundur (*Forward-Backward Splitting*: FBS) untuk mengira norma terkecil dari pemberat keluaran ELM dipersembahkan. FASTA-ELM yang dicadangkan menggantikan langkah analitik yang biasanya diselesaikan oleh SVD dengan penyelesaian hampiran melalui kaedah kecerunan proksimal, yang terbukti mempercepat tempoh latihan dan meningkatkan kemampuan generalisasi dalam tugasan pengkelasan. Prestasi FASTA-ELM dinilai berdasarkan masalah pengecaman jantina dan hasilnya setanding dengan kaedah terkini, dengan tempoh latihan yang lebih pantas. Sebagai contoh tempoh latihan untuk 1000 nod ELM adalah 18.11 s, sementara FASTA-ELM selesai dalam 1.671 s. Pengubahsuaian ELM yang dicadangkan menunjukkan peningkatan signifikan ELM konvensional dari segi tempoh latihan dan kejituan, dan memberi prestasi generalisasi yang baik dalam tugasan regresi dan pengkelasan.

ACKNOWLEDGEMENTS

First and foremost, I would like to thank almighty ALLAH (S.W.T), the Most Beneficent and the Most Merciful, for giving me the strength, courage, and his blessed guidance, during my post graduate period.

I am grateful to my supervisor, Mohammad Hamiruce Marhaban, for affording the opportunity to embark on this research. His consistent motivation, support, and guidance have been very crucial in the completion of this thesis. This research would not have been possible without her insightful conversations and recommendations.

I would like to express my sincere gratitude to my supervisory committee members, Fakhrul Zaman Rokhani and Khairulmizam Samsudin for spending their time in reviewing this thesis. Also, their valuable suggestions and comments have been very helpful in modifying this thesis.

Finally, I would like to thank my parents and my wife who deserve my utmost appreciation. I am immensely indebted to them for everything they have provided for me in life. I also thank my brothers and sisters for their love and affection. May ALLAH bless and protect my family, and may they live long with sound health.

This thesis was submitted to the Senate of the Universiti Putra Malaysia and has been accepted as fulfilment of the requirement for the degree of Doctor of Philosophy. The members of the Supervisory Committee were as follows:

Mohammad Hamiruce Marhaban, PhD
Professor
Faculty of Engineering
Universiti Putra Malaysia
(Chairman)

Fakhrul Zaman Rokhani, PhD
Associate Professor
Faculty of Engineering
Universiti Putra Malaysia
(Member)

Khairulmizam Samsudin, PhD
Senior Lecturer
Faculty of Engineering
Universiti Putra Malaysia
(Member)

ZALILAH MOHD SHARIFF, PhD
Professor and Dean
School of Graduate Studies
Universiti Putra Malaysia

Date:12 November 2020

Declaration by graduate student

I hereby confirm that:

- this thesis is my original work;
- quotations, illustrations and citations have been duly referenced;
- this thesis has not been submitted previously or concurrently for any other degree at any institutions;
- intellectual property from the thesis and copyright of thesis are fully-owned by Universiti Putra Malaysia, as according to the Universiti Putra Malaysia (Research) Rules 2012;
- written permission must be obtained from supervisor and the office of Deputy Vice-Chancellor (Research and innovation) before thesis is published (in the form of written, printed or in electronic form) including books, journals, modules, proceedings, popular writings, seminar papers, manuscripts, posters, reports, lecture notes, learning modules or any other materials as stated in the Universiti Putra Malaysia (Research) Rules 2012;
- there is no plagiarism or data falsification/fabrication in the thesis, and scholarly integrity is upheld as according to the Universiti Putra Malaysia (Graduate Studies) Rules 2003 (Revision 2012-2013) and the Universiti Putra Malaysia (Research) Rules 2012. The thesis has undergone plagiarism detection software

Signature:________________________________ Date:________________

Name and Matric No: Saif F Mahmood, GS35200

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ELM	Extreme Leaning Machine
LBP	Local Binary Pattern
HOG	Histogram of Oriented Gradients
WLD	Weber Law Descriptor
SVM	Support Vector Machine
TP	True Positive
TN	True Negative
SLFN	Single Layer Feedforward Network
FASTA	Fast Adaptive Shrinkage / Thresholding Algorithm
S-ELM	Standard Extreme Leaning Machine
EM-ELM	Error minimized Extreme Leaning Machine
R-ELM	Regularized Extreme Leaning Machine
SVD	singular Value Decomposition
BP	Back Propagation
RBF	Radial Basis Function
OP-ELM	Optimally Pruned Extreme Learning Machine
BIP-ELM	Batch Intrinsic Plasticity Extreme Learning Machine
EAs	Evolutionary Algorithms
RVFL	Random Vector functional-link
VC dimension	Vapnik–Chervonenkis dimension
LS-SVM	Least Square Support Vector Machine

CHAPTER 1

INTRODUCTION

In general, this chapter introduce background knowledge on the domain of machine learning approach. More insight is given into Artificial Neural Network (ANN) method, with highlights to the tendency of initial weight by random process in a method knowing as Extreme Learning Machine (ELM). In addition, this chapter points out the problem statement, objectives that are to be achieved at the end of this study, and, the scope of the research.

1.1 Machine Learning Approach

In recent years, in line with technological advances, machine learning has exhibited several promising algorithms in various real world applications. Rather than inference based rules, machine learning makes sense of the data through building a predication model from seen data in phase named training process, then valid the model in unseen data in second stage called testing phase. Whether the output is labelled entirely or semi and neither without labeling, data-driven tasks are learned as supervised, semi supervised and unsupervised learning respectively. Distinguished machine learning approaches lies in two important aspects include learning speed and performance. Whereas learning speed is typically depend on computational burden in term of execution time, the performance of learning algorithm is represented by generalization ability measured by training and testing error. And with incorporation of learnable functions to undertake the universality capability (universal function approximation). In addition to former two metrics, reliability of the model possess a crucial role in data-driven model as explicitly studied by Neumann (2013). To build a compact model, model parameter can be involve in optimization problem with some objective function. From particular point of view, many challenges are arisen in machine learning starting with the unpredictable model come out with complex environments and not ending up with the enormous growth in amount of data. For these, machine learning model that can be sorted include probabilistic model, geometric model and logic model (Flach 2012).

The importance of probabilistic in machine learning lies in alleviating the tendency of uncertainty by learning process. Motivated by geometric concept such as line, plane and distance, geometric model has been developed especially with high dimension with prefixed “Hyper”. In additional, logic model has been used based on logical relation. Features also called attributes are contributed in the model assessment based on scale and ordering through those with numerical scale called quantitative feature and those without scale named ordinal feature as well categorical features that are without scale or ordering (Flach 2012). Basically, machine learning scheme can be represented as shown in Figure1.1. Where the model in task box perform a mapping from date defined by features to output. Mapping is obtained using training data in learning process.

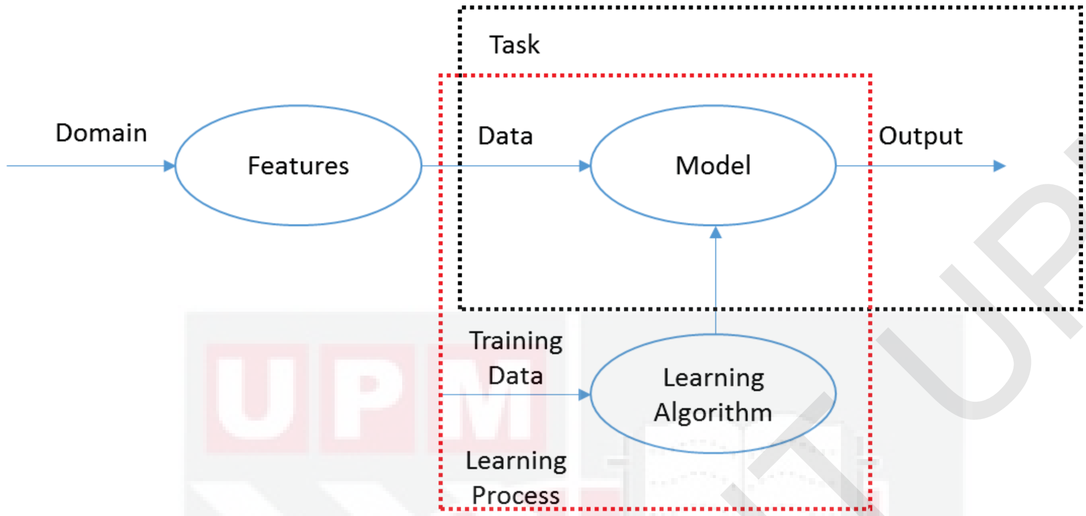

Figure 1.1 : Machine learning scheme

1.2 Artificial Neural Networks

The underlying function of human in brain has received considerable attention by neurologists, physiologists and psychologists. Brain architecture constitutes of immense number of computing units (neurons) that are interconnected (synapses) and operating in parallel manner. Inspired by theoretical and empirical aspects of human brain, well-known artificial neural network has been acquired as artificial brain models to solving specific problem. Basically, the main research that interested on artificial neural networks are: (i) mathematic at intrinsic of neuron and synapses; and (ii) using artificial neural networks to effectively build a machine learning model to perform data processing, inferring decisions and interpretable by simulate the biological nervous system. Indeed, two important features of neural network as a computational method (Daniel Graupe 2013):

(i) It employs simple computational operations to solve mathematical problems with ill-define nonlinear and stochastic approach.
(ii) It possesses a self-organizing feature with learning ability that implies the capability of solving wide range of problems.

Furthermore, White (1989) suggested QuickNet algorithm which inserted randomness of hidden neurons in the single of hidden layer feedforward networks (SLFNs) and the output of hidden layer weights can be calculated of the current hidden neurons after adding new hidden neuron. Randomness clearly exploited in hidden neurons by Schmidt and Kraaijveld (1992).

And they concluded that output weight has significantly contributed in learning process rather than hidden layer weights. The main constraint with it is required a output neuron bias, and cannot prove the universal capability of the method. For same purpose, a new approach named Random Vector Functional Link (RVFL) which is depend on randomness that is partial technique in neurons that are hidden (Pao et al 1994) (Igelnik. B and Pao. Y. H, 1995). Although it is remarkable contribution, there is no universal capability prove of the single layer feedforward. In addition, RVFL network employed the traditional gradient descent mechanism to learn the weights of output layer, in summarize, RVFL don't exploit random features and random hidden layer neurons.

1.3 Extreme Learning Machines

The advantage of random process approach has involved a lot of consideration in many challenging of application fields, incorporation of random projections, efficient processing can be obtained for large and high-dimensional data sets (Miche et al 2009). Inspired by vast literature of these methods, researchers explore the usefulness of generate the initial parameters randomly and restrict learning phase to be formalized as a linear model. A prominent machine learning method with single hidden layer feedforward neural network (SLFN) and high-dimensional random process named extreme learning machine (ELM) is suggest by Guang Bin Huang (2014a).

Compared to conventional backpropagation training methods, ELM is trained much faster with capability to be trapped at global optimum. As theoretical aspects have reveal that although that random process of hidden nodes parameters, ELM can be maintained the universal approximation capability of SLFNs. Unlike other learning process involve with random projection for /network, the indecency of hidden nodes in ELM are not only for training data but also on each other. Moreover, the generation of hidden node parameters in ELM can be obtained even before training data has been shown (Ding, Xu, & Nie, 2013). In ELM theory the meaning of random hidden nodes imply that parameters linked the hidden layer with input layer are selected by random process and without depending on the training samples, e.g., input weights are randomly generated weight and biases for additive hidden nodes such sigmoid, or both center and impact factor for RBF networks. The main aspect of ELM is the hidden nodes do not essential to be tuned (Guang-Bin Huang, Hongming Zhou, Xiaojian Ding, & Rui Zhang, 2012a). In summary, the randomness of ELM clarify in two folds:

(a) The randomness of hidden nodes parameters.
(b) Although the hidden nodes are randomly found, they do no need to be explicitly tuned.

For instance, a hidden node in the follow layer can be basically linear sum or nonlinear transform of some random nodes in the earlier layer. In this example, some neuron are selected randomly and some are not, but no one of them have being tuned. Different from Schmidt et al (1992) and Pao et al (1994) that each neuron is either sigmoid or RBF node only, each hidden node in ELM can be any bounded nonconstant piecewise continuous functions. As shown in Figure 1.2 ELM can be a subnetwork of other nodes in which feature learning allow to be used considerably. As well, ELM can cooperate with compression, feature learning, clustering, regression or classification. thus, homogeneous ordered blocks of ELM can be constructed. For instance, one ELM as feature learning process, the following ELM mechanism is as a classifier. In this instance, there are two hidden layers of ELM, in general it is not randomly generated and it is ordered, but hidden nodes in each layer do not need to be tuned (e.g., randomly chosen or obviously given/calculate) (Guang-bin Huang, 2015) Based on ELM theory, ELM SLFNs, activation function contains but are not restricted to:

1. Sigmoid activation function networks
2. RBF activation function networks
3. Threshold function networks
4. Trigonometric function networks
5. Inference Fuzzy systems
6. Fully complex neural networks
7. High-order networks
8. Ridge function polynomial networks
9. Wavelet function networks
10. Fourier series

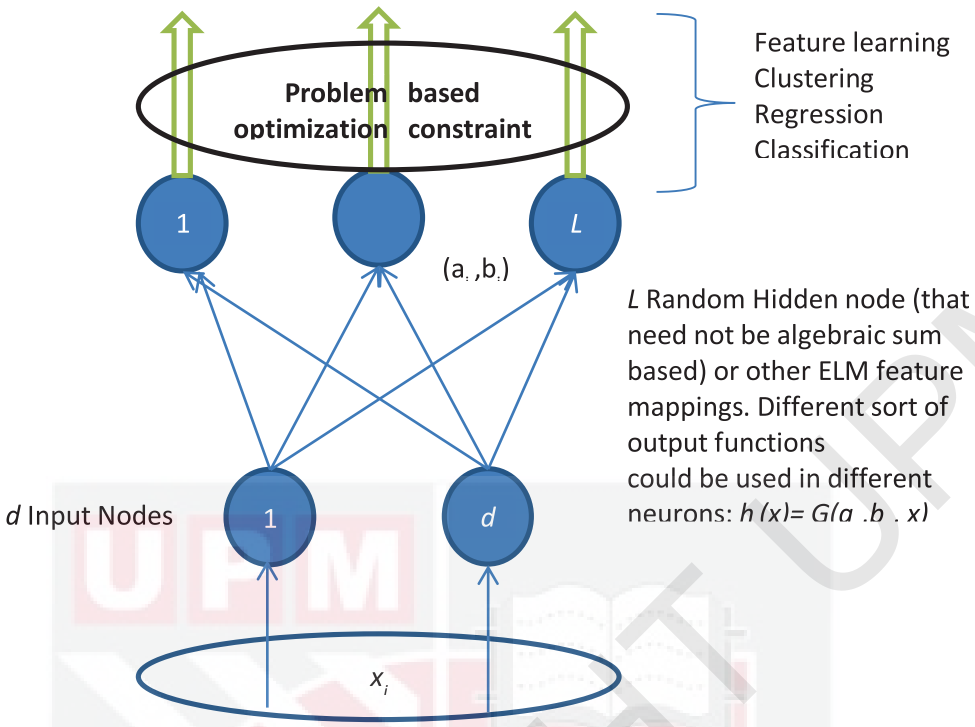

Figure 1.2 : ELM Structure

Furthermore, there are distinction and similarity among ELM and other well-known algorithms, such as Deep Learning and support vector machine (SVM) least square support vector machine (LS-SVM). Unlike deep Learning, ELM in its logic that hidden layer node of the whole ELM need not to be tuned. Because of ELM's altered roles of feature learning and clustering, ELM can be executed as the earlier layers in multilayer networks in which the late layers are trained by other learning algorithms such as deep learning.

Compared with SVM and LS-SVM, SVM was firstly found to interact with multilayer feedforward networks by Cortes and Vapnik (1995) which shows that there is no method to train a multilayer network. Different from ELM and deep learning which investigate feature representations in each layer, SVM and LS-SVM do not consider the feature representation and functioning roles of each inner hidden layer. Moreover, SVM and LS-SVM can be able as single hidden- layer networks associated with hidden layer output function. For this regard, while ELM and SVM/LSSVM possess a single hidden layers, but ELM explicitly possess hidden layer mapping (suitable for feature representations) and SVM/LS-SVM has unknown hidden layer mapping (inconvenient for feature representations). In different machine learning tasks, ELM represents feature mapping, clustering based regression and pattern classification with ridge regression optimization criteria, whereas SVM/LS-SVM basically perform for binary pattern classification based on Vapnik–Chervonenkis dimension (VC dimension) through maximal margin. However, it is hard for

SVM/LSSVM to have feature mapping because anonymous mapping as SVM and LS-SVM generally offer suboptimal results.

Moreover, the universal approximation capability of ELM can be contributed in a varied sort of functions that are nonlinear piecewise continuous, and it obvious that not need to bias existing in the output layer. For that, ELM is employed wide range of activation functions in both real and complex domains (Huang et al 2006)

Following the Bartlett's concept (Bartlett 1998), smallest training error can be associated with improving of generalization performance for the network, the proposed solution of ELM in this thesis is to solve the linear equation and reach solution of norm to be smallest. Three vital aspects making it stimulating solution (Huang et al 2006):

1. The minimum training error can be obtained as it is basically linear equation that use least-squares method as a solution.
2. Among the least-squares solutions, ELM can reach the smallest norm.
3. The smallest norm solution is unique among the least-squares solutions.

1.4 Problems and Motivations

Artificial Neural Networks of (ANNs) are common approaches for pattern recognition tasks. They have the ability to handle the problems in practical applications such as non-linear and noisy data. Nevertheless, it suffer from slowing of the learning in feedforward neural networks and generally far slower than necessary. Thus, it has been a main restriction for their implementations. Two important reasons could be (Zhang et al., 2008):

(i) The slowing of the training in error back-propagation (BP) process and other gradient descent learning methods has been extensively used.
(ii) Iteratively tune the synaptic parameters of the neural networks by using gradient descent learning algorithms.

To cope with the shortcomings mentioned above, in 2006 Huang et al (2006) suggest a novel learning algorithm called ELM for single hidden layer feedforward neural network (SLFN). ELM mitigates the iterative procedures of adjusting weights by randomly select hidden layer neurons and determine the output weights analytically. From theory point of view, ELM incline to include perfect generalization performance associated with extremely fast learning speed. Despite of many aspects on ELM, there are some basic matters that need to be studied to understand the feasibility of ELM in real-world implementations. One of these issues is how to find the optimal and the compact network architecture, which can finally result in a robust model and applicable to many learning tasks (X. Wang & Cao, 2018). It is obvious that the choosing of an optimal number of nodes is thoroughly related to the problem of curve fitting using polynomials (Kwok & Yeung, 1997)(Rocha & Ananda, 2017). Such as too many coefficients will encounter overfitting the model and therefore poor performance would be occurred, while few coefficients cannot learn well the underlying function (MLD Dias et al. 2018). Practically, a certain number of the hidden nodes in such networks may have a minor effect on the output of the network and theoretically increase the network complexity (Guang Bin Huang et al 2004).

Moreover, ELM may meet the singularity problem in case of the number of nodes are greater than the training samples which eventually lead to instability of the system (Alcin et al., 2015). For that reason, empirical method such as trial-and-error is used to find a compact network structure for a specific task without any prior knowledge (Guang-Bin Huang et al., 2012a). ELM methods generally need a lot of hidden nodes as mentioned in (Guang-Bin Huang et al (2012a) in order to find an optimal number of neurons, publications are suggested using optimization algorithms (Sheela & Deepa, 2013).

These works also imply that identification of hidden nodes can be considered as a kind of optimization process (Rong et al., 2008). If some optimal neurons are selected in the learning process, the learning effectiveness will be improved dramatically. While reducing the number of hidden nodes can be further occur without disturbing learning performance for large-size or high dimension data sets. Guang-Bin Huang et al (2012b) has stated that good performance can be obtained as long as number of node is large enough.

Another shortcoming of ELM is the computational time during training phase due to singular value decomposition (SVD) for calculating the Moore-Penrose generalized inverse of the hidden layer matrix that has high complexity burden. Especially with large dimension on hidden layer matrix and the effect of the ill conditioned of the hidden layer matrix that make ELM suffer from quite weak numerical stability that lead to waning of the generalization performance (Tang & Han, 2009).

According to previous explanation, the problem statement can be summarized as follow:

1. The existing methods to determine the optimal number of hidden neurons in ELM is trial-and-error approach that it is time consuming and there is no guarantee of fixing the hidden neuron.

2. The output weight of ELM is found out by using SVD which has high burden of computational complexity $O(4Nn^2 + 8n^3)$, when number hidden neurons became large the computational complexity of SVD will significantly rise.

1.5 Objectives(s) of the study

The main objective of this research work is to design, develop, and enhance ELM based neural network models that capitalize the advantage of ELM, and at the same time, avoid their limitation. To attain the main objectives, sub-objectives are defined as below:

1. To develop novel architectures of ELM based SVM to select the optimal number of nodes
2. To test and evaluate the procedure of an SVM-ELM algorithm by mitigating the influence of large number of node in regression task.
3. To develop a novel method named Fast Adaptive Shrinkage /Thresholding Algorithm (FASTA) to calculate the output weights of ELM with the smallest norm using a forward backward splitting.
4. To investigate the feasibility and efficiency of FASTA-ELM algorithm by applying it in face gender recognition task.

1.6 Research Scopes

With a single hidden layer feedforward, ELM consider the mainstay of this thesis that establish by developing a novel pruning network architectures method named ELM-SVM. As well, new learning algorithms called fast adaptive shrinkage/thresholding algorithm (FASTA-ELM) that attempt to disregards the need of employ SVD to compute the hidden layer of matrix of ELM. The scope of this research is motived on two areas:

1. Development of effective novel pruning strategy of network architectures systems of ELM based 1-norm SVM that are capable of acquiring the active nodes with as little supervision as possible to data regression tasks.
2. Development a new method that calculating the output weights of ELM with the smallest norm using a forward backward splitting method named (FASTA) for pattern classification tasks.
3. To demonstrate the performance of the proposed method, the dataset is used in this thesis is as follow:

For regression task, the proposed SVM–ELM is evaluated on 10 real-world regression problems from an UCI ML (C. Blake & C. Merz 1998). For classification task, the experiments in this thesis are conducted on 5 benchmarked face gender recognition datasets as follow:

- 2 unconstrained datasets: Cropped Labeled Faces in the Wild (G. B. Huang et al., 2008) and Image of Groups(Gallagher & Chen, 2009)
- 3 constrained datasets: PAL (Minear & Park, 2004), FERET (Phillips et al., 1999), and CAS-PEAL R1 (Gao et al., 2008)

4. All the simulations for ELM and the new ELM algorithms are both carried out in the Matlab 2016 environment running in Intel Core i5 6200u CPU.With the speed 2.40GHz.

It is very clear that the hidden node choosing is a stimulating problem and is still a very much open problem of research. Our study in this thesis is motivated by the model suggested by Han and Yin (2008), where a 1-norm support vector machine (SVM) is used to pre-select the hidden neurons, followed by a stepwise selection method based on ridge regression for selecting the optimal hidden nodes of wavelet networks. The key objective of the proposed method is to remove the issue of including a great number of inactive neurons in wavelet networks.

Another problem in ELM that handle by this thesis that has been formulated as objective function of ELM to minimizing compound optimization problem solved with the forward backward splitting (gradient descent method) in instead of computing the hidden layer matrix analytically using SVD. However, adaptive stepsize selection method and non-monotonic line search are used in the suggested FASTA-ELM method to speed up the convergence rate. Thus, if we have a large ill-conditioned matrix, rather than searching through all possible steps of the objective, the step-size adaptively choose such as that the dimension of the gradient descent problem becomes very small, which in turn makes the algorithm converge very fast. As well, FASTA-ELM further employ a backward descent step to check objective in each step in order to select the sub-gradient which ensures convergence.

1.7 Thesis Outline

The thesis is structured with 5 main chapters. Chapter 1 introduces the concept of machine learning, which is divided into neural network and extreme learning machine. The chapter also states the problem statement, objectives, and scope of the study. In Chapter 2, in order to point out the advantages and disadvantages of previous techniques. Also, to highlight the techniques that are adopted in this study and the areas the thesis contributes to knowledge, literature review on extreme learning machine by several approaches is introduced. In chapter 3 the proposed framework on FATSA-ELM and SVM-ELM is presented, and the methodologies utilized are clarified in detail. Chapter 4 presents comprehensive experiment results and discuss on the analysis of findings attained. To conclude, chapter 5 provides the conclusion and recommendation for future studies.

REFERENCES

Ahila, R., Sadasivam, V., & Manimala, K. (2015). An integrated PSO for parameter determination and feature selection of ELM and its application in classification of power system disturbances. *Applied Soft Computing Journal*, *32*, 23–37. https://doi.org/10.1016/j.asoc.2015.03.036

Alcin, O. F., Sengur, A., Qian, J., & Ince, M. C. (2015). OMP-ELM: orthogonal matching pursuit-based extreme learning machine for regression. Journal of Intelligent Systems, 24(1), 135-143.

Alpaydin. (2020). Introduction to machine learning. *MIT Press.*

Arigbabu, O.A., Ahmad, S. M. S., Adnan, W. A. W., Yussof, S., & Mahmood, S. (2015). Soft biometrics: Gender recognition from unconstrained face images using local feature descriptor. *Journal of Information and Communication Technology*, *14*(1).

Arigbabu, Olasimbo Ayodeji, Ahmad, S. M. S., Adnan, W. A. W., Yussof, S., Iranmanesh, V., & Malallah, F. L. (2014). Gender recognition on real world faces based on shape representation and neural network. *Proceedings of 2014 IEEE International Conference on Computer and Information Sciences (ICCOINS)*, 1–5. Retrieved from http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6868361

Bartlett, P. L. (1998b). 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*, *44*(2), 525–536. https://doi.org/10.1109/18.661502

Barzilai, J., & Borwein, J. M. (1988). Two-point step size gradient methods. IMA journal of numerical analysis, 8(1), 141-148.

Berg, E. V. A. N. D. E. N., & Friedlander, M. P. (2008). PROBING THE PARETO FRONTIER FOR BASIS PURSUIT SOLUTIONS, *31*(2), 890–912.

Bertsekas, D. P. (1976). On the Goldstein-Levitin-Polyak gradient projection method. IEEE Transactions on automatic control, 21(2), 174-184.

Birgin, E. G., Martínez, J. M., & Raydan, M. (2000). Nonmonotone spectral projected gradient methods on convex sets. SIAM Journal on Optimization, 10(4), 1196-1211.

Blanz, V., Schölkopf, B., Bülthoff, H. C. B. V. V., Burges, C., Vapnik, V., & Vetter, T. (1996, July). Comparison of view-based object recognition algorithms using realistic 3D models. In International Conference on Artificial Neural Networks (pp. 251-256). Springer, Berlin, Heidelberg.

Boser, B. E., Guyon, I. M., & Vapnik, V. N. (1992, July). A training algorithm for optimal margin classifiers. In Proceedings of the fifth annual workshop on Computational learning theory (pp. 144-152).

Brereton Richard G, & Lloyd Gavin R. (2010). Support vector machines for classification and regression. *Analyst*, *135*(2), 230–267. https://doi.org/10.1039/b918972f

Bueno-Crespo, A., García-Laencina, P. J., & Sancho-Gómez, J. L. (2013). Neural architecture design based on extreme learning machine. Neural Networks, 48, 19-24.

Blake, C. (1998). UCI repository of machine learning databases. http://www. ics. uci. edu/~ mlearn/MLRepository. html.

C. J. C. Burges. (1998). A Tutorial on Support Vector Machines for Pattern Recognition. *Data Mining and Knowledge Discovery, 2(2), 121-167*

Calvetti, D., & Reichel, L. (2003). Tikhonov regularization of large linear problems. BIT Numerical Mathematics, 43(2), 263-283.

Candes, E. J., Strohmer, T., & Voroninski, V. (2013). Phaselift: Exact and stable signal recovery from magnitude measurements via convex programming. Communications on Pure and Applied Mathematics, 66(8), 1241-1274.

Cao, W., Wang, X., Ming, Z., & Gao, J. (2018). A review on neural networks with random weights. *Neurocomputing*, *275*, *278–2*.

Cao, J., Lin, Z., & Huang, G. B. (2012). Self-adaptive evolutionary extreme learning machine. Neural processing letters, 36(3), 285-305.

Castaño, a., Fernández-Navarro, F., & Hervás-Martínez, C. (2012). PCA-ELM: A Robust and Pruned Extreme Learning Machine Approach Based on Principal Component Analysis. *Neural Processing Letters*, *37*(3), 377–392. https://doi.org/10.1007/s11063-012-9253-x

Cervellera, C., & Macciò, D. (2016). Brief Papers Low-Discrepancy Points for Deterministic Assignment of Hidden Weights in Extreme Learning Machines, *27*(4), 891–896.

Chambolle, A. (2004). An algorithm for total variation minimization and applications. Journal of Mathematical imaging and vision, 20(1-2), 89-97.

Chen, K., Lv, Q., Lu, Y., & Dou, Y. (2017). Robust regularized extreme learning machine for regression using iteratively reweighted least squares. *Neurocomputing*, *230*(June), 345–358. https://doi.org/10.1016/j.neucom.2016.12.029

Combettes, P. L., & Wajs, V. R. (2005). Signal recovery by proximal forward-backward splitting. Multiscale Modeling & Simulation, 4(4), 1168-1200.

Corinna, C., & Vladimir, V. (1995). Support-Vector Networks. *Machine Learning*, *20*(3), 273–297. https://doi.org/10.1007/BF00994018

Cortes, C., & Vapnik, V. (1995). Supprot-Vector Networks. *Machine Learning*, *297*(20), 273–297. https://doi.org/10.1111/j.1747-0285.2009.00840.x

Cotter, S. F., Rao, B. D., Engan, K., & Kreutz-Delgado, K. (2005). Sparse solutions to linear inverse problems with multiple measurement vectors. IEEE Transactions on Signal Processing, 53(7), 2477-2488.

Courrieu, P. (2002). Straight monotonic embedding of data sets in Euclidean spaces. Neural Networks, 15(10), 1185-1196.

Courrieu, P. (2008). Fast computation of Moore-Penrose inverse matrices. arXiv preprint arXiv:0804.4809.

Cristianini, N., & Shawe-Taylor, J. (2000). An introduction to support vector machines and other kernel-based learning methods. Cambridge university press.

Deng, W., Zheng, Q., & Chen, L. (2009, March). Regularized extreme learning machine. In 2009 IEEE symposium on computational intelligence and data mining (pp. 389-395). IEEE.

Dias, M. L., de Sousa, L. S., Neto, A. R. R., & Freire, A. L. (2018). Fixed-Size Extreme Learning Machines Through Simulated Annealing. Neural Processing Letters, 48(1), 135-151.

Ding, S., Ma, G., & Shi, Z. (2014). A Rough RBF Neural Network Based on Weighted Regularized Extreme Learning Machine. *Neural Processing Letters*, *40*(3), 245–260. https://doi.org/10.1007/s11063-013-9326-5

Ding, S., Xu, X., & Nie, R. (2013). Extreme learning machine and its applications. *Neural Computing and Applications*, *25*(3–4), 549–556. https://doi.org/10.1007/s00521-013-1522-8

Ding, S., Zhao, H., Zhang, Y., Xu, X., & Nie, R. (2015). Extreme learning machine: algorithm, theory and applications. *Artificial Intelligence Review*, *44*(1), 103–115. https://doi.org/10.1007/s10462-013-9405-z

Duchi, J., Shalev-Shwartz, S., Singer, Y., & Chandra, T. (2008, July). Efficient projections onto the l 1-ball for learning in high dimensions. In Proceedings of the 25th international conference on Machine learning (pp. 272-279).

Duchi, J., & Singer, Y. (2009). Efficient online and batch learning using forward backward splitting. The Journal of Machine Learning Research, 10, 2899-2934.

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

Figueiredo, E. M., & Ludermir, T. B. (2014). Investigating the use of alternative topologies on performance of the PSO-ELM. Neurocomputing, 127, 4-12.

Figueiredo, M. A., Nowak, R. D., & Wright, S. J. (2007). Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems. IEEE Journal of selected topics in signal processing, 1(4), 586-597.

Flach, P. (2012). Machine learning: the art and science of algorithms that make sense of data. Cambridge University Press.

R. Fletcher. (1987). Practical Methods of Optimization. *John Wiley and Sons*.

Fung, G., & Mangasarian, O. L. (2003). Finite Newton method for Lagrangian support vector machine classification. *Neurocomputing*, *55*(1–2), 39–55.

Furey, T. S., Cristianini, N., Duffy, N., Bednarski, D. W., Schummer, M., & Haussler, D. (2000). Support vector machine classification and validation of cancer tissue samples using microarray expression data. Bioinformatics, 16(10), 906-914.

Jégou, H., Furon, T., & Fuchs, J. J. (2012, March). Anti-sparse coding for approximate nearest neighbor search. In 2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 2029-2032). IEEE.

Gallagher, A. C., & Chen, T. (2009, June). Understanding images of groups of people. In 2009 IEEE Conference on Computer Vision and Pattern Recognition (pp. 256-263). IEEE.

Gao, W., Cao, B., Shan, S., Chen, X., Zhou, D., Zhang, X., & Zhao, D. (2007). The CAS-PEAL large-scale Chinese face database and baseline evaluations. IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans, 38(1), 149-161.

Ghosh, R., & Verma, B. (2003). A hierarchical method for finding optimal architecture and weights using evolutionary least square based learning. International Journal of Neural Systems, 13(01), 13-24.

Girosi, F., & Poggio, T. (1990). Networks and the best approximation property. Biological cybernetics, 63(3), 169-176.

Goldstein, T., & Setzer, S. (2010). High-order methods for basis pursuit. UCLA CAM Report, 10-41.

Goldstein, T., Studer, C., & Baraniuk, R. (2014). A field guide to forward-backward splitting with a FASTA implementation. arXiv preprint arXiv:1411.3406.

Graupe, D. (2013). Principles of artificial neural networks (Vol. 7). World Scientific.

Gregor, K., & LeCun, Y. (2010, June). Learning fast approximations of sparse coding. In Proceedings of the 27th international conference on international conference on machine learning (pp. 399-406).

Huang, G. B., Zhou, H., Ding, X., & Zhang, R. (2011). Extreme learning machine for regression and multiclass classification. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 42(2), 513-529.

Han, F., Yao, H. F., & Ling, Q. H. (2013). An improved evolutionary extreme learning machine based on particle swarm optimization. Neurocomputing, 116, 87-93.

Han, M., & Yin, J. (2008). The hidden neurons selection of the wavelet networks using support vector machines and ridge regression. Neurocomputing, 72(1-3), 471-479.

Hesterberg, T., Choi, N. H., Meier, L., & Fraley, C. (2008). Least angle and ℓ1 penalized regression: A review. Statistics Surveys, 2, 61-93.

Horata, P., Chiewchanwattana, S., & Sunat, K. (2011, October). A comparative study of pseudo-inverse computing for the extreme learning machine classifier. In The 3rd International Conference on Data Mining and Intelligent Information Technology Applications (pp. 40-45). IEEE.

Horata, P., Chiewchanwattana, S., & Sunat, K. (2013). Robust extreme learning machine. Neurocomputing, 102, 31-44.

Hu, L., Wang, H., Zhao, B., & Yang, W. (2007). A hybrid reconstruction algorithm for electrical impedance tomography. Measurement Science and Technology, 18(3), 813.

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

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

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

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

Huang, G. B., Li, M. B., Chen, L., & Siew, C. K. (2008). Incremental extreme learning machine with fully complex hidden nodes. Neurocomputing, 71(4-6), 576-583.

Huang, G. B., Zhu, Q. Y., & Siew, C. K. (2004, July). Extreme learning machine: a new learning scheme of feedforward neural networks. In 2004 IEEE international joint conference on neural networks (IEEE Cat. No. 04CH37541) (Vol. 2, pp. 985-990). IEEE.

Huang, G. B., & Chen, L. (2008). Enhanced random search based incremental extreme learning machine. Neurocomputing, 71(16-18), 3460-3468.

Huang, G. B., Ramesh, M., Berg, T., and Learned-Miller, E. Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments. Technical Report 07-49, University of Massachusetts, Amherst, 2007

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.

Igelnik, B., & Pao, Y. H. (1995). Stochastic choice of basis functions in adaptive function approximation and the functional-link net. IEEE transactions on Neural Networks, 6(6), 1320-1329.

Israel, A., & Greville, T. N. (2003). Generalized inverses: theory and applications (Vol. 15). Springer Science & Business Media.

Jaiganesh, V., & Sumathi, P. (2012). Kernelized extreme learning machine with levenberg-marquardt learning approach towards intrusion detection. International Journal of Computer Applications, 54(14).

Jenatton, R., Mairal, J., Obozinski, G., & Bach, F. R. (2010, January). Proximal methods for sparse hierarchical dictionary learning. In ICML.

Joachims, T. (1998, April). Text categorization with support vector machines: Learning with many relevant features. In European conference on machine learning (pp. 137-142). Springer, Berlin, Heidelberg.

John C. Platt. (1999). Fast Training of Support Vector Machines using Sequential Minimal Optimization. *In A. J. Smola, P. L. Bartlett, B. Schölkopf, and D. Schuurmans, Eds., Advances in Large Margin*

Classifiers, Pp. 185-208, MIT Press, Cambridge, MA, U.S.A.,.

Karaboga, D., & Basturk, B. (2007). A powerful and efficient algorithm for numerical function optimization: artificial bee colony (ABC) algorithm. Journal of global optimization, 39(3), 459-471.

Keshavan, R. H., Montanari, A., & Oh, S. (2009, September). Low-rank matrix completion with noisy observations: a quantitative comparison. In 2009 47th Annual Allerton Conference on Communication, Control, and Computing (Allerton) (pp. 1216-1222). IEEE.

Kilmer, M. E., & O'Leary, D. P. (2001). Choosing regularization parameters in iterative methods for ill-posed problems. SIAM Journal on matrix analysis and applications, 22(4), 1204-1221.

Kwok, T. Y., & Yeung, D. Y. (1997). Objective functions for training new hidden units in constructive neural networks. IEEE Transactions on neural networks, 8(5), 1131-1148.

Lan, Y., Soh, Y. C., & Huang, G. B. (2010, April). Random search enhancement of error minimized extreme learning machine. in: European Symposium on Artificial Neural Networks (ESANN) 2010, Bruges, Belgium, 2010, pp. 327–332.

Lan, Y., Soh, Y. C., & Huang, G. B. (2010). Constructive hidden nodes selection of extreme learning machine for regression. Neurocomputing, 73(16-18), 3191-3199.

Levenberg, K. (1944). A method for the solution of certain non-linear problems in least squares. Quarterly of applied mathematics, 2(2), 164-168.

Li, S., You, Z. H., Guo, H., Luo, X., & Zhao, Z. Q. (2015). Inverse-free extreme learning machine with optimal information updating. IEEE transactions on cybernetics, 46(5), 1229-1241.

Liao, S., & Feng, C. (2014). Meta-ELM: ELM with ELM hidden nodes. Neurocomputing, 128, 81-87.

Liu, H., Gao, Y., & Wang, C. (2014, October). Gender identification in unconstrained scenarios using self-similarity of gradients features. In 2014 IEEE International Conference on Image Processing (ICIP) (pp. 5911-5915). IEEE.

Liu, T., Yang, Y., Huang, G. B., Lin, Z., Klanner, F., Denk, C., & Rasshofer, R. H. (2015, September). Cluster regularized extreme learning machine for detecting mixed-type distraction in driving. In 2015 IEEE 18th International Conference on Intelligent Transportation Systems (pp. 1323-1326). IEEE.

Mangasarian, O. L. (2006). Exact 1-norm support vector machines via unconstrained convex differentiable minimization. Journal of Machine Learning Research, 7(Jul), 1517-1530.

Marquardt, D. W. (1963). An Algorithm for Least-Squares Estimation of Nonlinear Parameters. *Journal of the Society for Industrial and Applied Mathematics*, *431441*.

Miche, Y., Sorjamaa, A., Bas, P., Simula, O., Jutten, C., & Lendasse, A. (2009). OP-ELM: optimally pruned extreme learning machine. *IEEE transactions on neural networks*, *21*(1), 158-162.

Miche, Y., Van Heeswijk, M., Bas, P., Simula, O., & Lendasse, A. (2011). TROP-ELM: a double-regularized ELM using LARS and Tikhonov regularization. Neurocomputing, 74(16), 2413-2421.

Miche, Y., Schrauwen, B., & Lendasse, A. (2010, April). Machine Learning Techniques based on Random Projections. In ESANN.

Minear, M., & Park, D. C. (2004). A lifespan database of adult facial stimuli. Behavior Research Methods, Instruments, & Computers, 36(4), 630-633.

Neumann, K. (2013). Reliability of Extreme Learning Machines, (October). phD thesis

Neumann, K., & Steil, J. J. (2013). Optimizing extreme learning machines via ridge regression and batch intrinsic plasticity. Neurocomputing, 102, 23-30.

Osuna, E., Freund, R., & Girosit, F. (1997, June). Training support vector machines: an application to face detection. In Proceedings of IEEE computer society conference on computer vision and pattern recognition (pp. 130-136). IEEE.

Oztoprak, F., Nocedal, J., Rennie, S., & Olsen, P. A. (2012). Newton-like methods for sparse inverse covariance estimation. In Advances in neural information processing systems (pp. 755-763).

Pao, Y. H., Park, G. H., & Sobajic, D. J. (1994). Learning and generalization characteristics of the random vector functional-link net. Neurocomputing, 6(2), 163-180.

Phillips, P. J., Moon, H., Rizvi, S. A., & Rauss, P. J. (2000). The FERET evaluation methodology for face-recognition algorithms. IEEE Transactions on pattern analysis and machine intelligence, 22(10), 1090-1104.

Platt, J. (1999). Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. Advances in large margin classifiers, 10(3), 61-74.

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

Rocco, C. M., & Moreno, J. A. (2002). Fast Monte Carlo reliability evaluation using support vector machine. Reliability Engineering & System Safety, 76(3), 237-243.

Rong, H. J., Ong, Y. S., Tan, A. H., & Zhu, Z. (2008). A fast pruned-extreme learning machine for classification problem. Neurocomputing, 72(1-3), 359-366.

Rudin, L. I., Osher, S., & Fatemi, E. (1992). Nonlinear total variation based noise removal algorithms. Physica D: nonlinear phenomena, 60(1-4), 259-268.

Schmidt, W. F., Kraaijveld, M. A., & Duin, R. P. (1992, August). Feed forward neural networks with random weights. In International Conference on Pattern Recognition (pp. 1-1). IEEE COMPUTER SOCIETY PRESS.

Smola, A. J., & Schölkopf, B. (2004). A tutorial on support vector regression. Statistics and computing, 14(3), 199-222.

Shan, C. (2012). Learning local binary patterns for gender classification on real-world face images. Pattern recognition letters, 33(4), 431-437.

Shao, H., & Japkowicz, N. (2012, May). Applying least angle regression to ELM. In Canadian Conference on Artificial Intelligence (pp. 170-180). Springer, Berlin, Heidelberg.

Sheela, K. G., & Deepa, S. N. (2013). Review on methods to fix number of hidden neurons in neural networks. Mathematical Problems in Engineering, 2013.

Silva, D. N., Pacifico, L. D., & Ludermir, T. B. (2011, June). An evolutionary extreme learning machine based on group search optimization. In 2011 IEEE Congress of Evolutionary Computation (CEC) (pp. 574-580). IEEE.

Similä, T., & Tikka, J. (2005, September). Multiresponse sparse regression with application to multidimensional scaling. In International Conference on Artificial Neural Networks (pp. 97-102). Springer, Berlin, Heidelberg.

Simon, H. D., & Zha, H. (2000). Low-rank matrix approximation using the Lanczos bidiagonalization process with applications. SIAM Journal on Scientific Computing, 21(6), 2257-2274.

Tang, X., & Han, M. (2009). Partial Lanczos extreme learning machine for single-output regression problems. Neurocomputing, 72(13-15), 3066-3076.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society: Series B (Methodological), 58(1), 267-288.

Vapnik, V. (2006). Estimation of dependences based on empirical data. Springer Science & Business Media.

Vapnik, V. N. (1999). An overview of statistical learning theory. IEEE transactions on neural networks, 10(5), 988-999.

Van Den Berg, E., & Friedlander, M. P. (2009). Probing the Pareto frontier for basis pursuit solutions. SIAM Journal on Scientific Computing, 31(2), 890-912.

Viola, P., & Jones, M. (2001, December). Rapid object detection using a boosted cascade of simple features. In Proceedings of the 2001 IEEE computer society conference on computer vision and pattern recognition. CVPR 2001 (Vol. 1, pp. I-I). IEEE.

Wang, X., & Cao, W. (2018). Non-iterative approaches in training feed-forward neural networks and their applications.

Wang, Y., Cao, F., & Yuan, Y. (2011). A study on effectiveness of extreme learning machine. Neurocomputing, 74(16), 2483-2490.

Wen, Z., Yin, W., Zhang, H., & Goldfarb, D. (2012). On the convergence of an active-set method for ℓ1 minimization. Optimization Methods and Software, 27(6), 1127-1146.

White, H. (1989, June). An additional hidden unit test for neglected nonlinearity in multilayer feedforward networks. In Proceedings of the international joint conference on neural networks (Vol. 2, pp. 451-455).

Wright, S. J., Nowak, R. D., & Figueiredo, M. A. (2009). Sparse reconstruction by separable approximation. IEEE Transactions on signal processing, 57(7), 2479-2493.

Xi-Zhao, W., Qing-Yan, S., Qing, M., & Jun-Hai, Z. (2013). Architecture selection for networks trained with extreme learning machine using localized generalization error model. Neurocomputing, 102, 3–9. https://doi.org/10.1016/j.neucom.2011.12.053

Xiao, J., Zhou, J., Li, C., Xiao, H., Zhang, W., & Zhu, W. (2014). Multi-fault classification based on the two-stage evolutionary extreme learning machine and improved artificial bee colony algorithm. Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical

Engineering Science, 228(10), 1797-1807.

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

Xue, H., Chen, S., & Yang, Q. (2011). Structural regularized support vector machine: a framework for structural large margin classifier. IEEE Transactions on Neural Networks, 22(4), 573-587.

Yu, H., & Wilamowski, B. M. (2011). Levenberg-marquardt training. Industrial electronics handbook, 5(12), 1.

Yuan, Y., Wang, Y., & Cao, F. (2011). Optimization approximation solution for regression problem based on extreme learning machine. Neurocomputing, 74(16), 2475-2482.

Zhai, J., Shao, Q., & Wang, X. (2016). Architecture selection of ELM networks based on sensitivity of hidden nodes. Neural processing letters, 44(2), 471-489.

Zhang, C., Wu, W., Chen, X. H., & Xiong, Y. (2008). Convergence of BP algorithm for product unit neural networks with exponential weights. Neurocomputing, 72(1-3), 513-520.

Zhao, G., Shen, Z., Miao, C., & Man, Z. (2009, December). On improving the conditioning of extreme learning machine: a linear case. In 2009 7th International Conference on Information, Communications and Signal Processing (ICICS) (pp. 1-5). IEEE.

Zhao, Y. P., Wang, K. K., & Li, Y. B. (2015). Parsimonious regularized extreme learning machine based on orthogonal transformation. Neurocomputing, 156, 280-296.

Zhou, B., Gao, L., & Dai, Y. H. (2006). Gradient methods with adaptive step-sizes. Computational Optimization and Applications, 35(1), 69-86.

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

BIODATA OF STUDENT

Saif F. Mahmood received the B.Sc. degree in Industrial Engineering from University of Technology, Iraq, in 2006, and the M.Sc degree in Mechanical Engineering from Universiti Tenaga Nasional, Malaysia (UNITEN), in 2012. He received a Ph.D. in Computer and Communication Systems Engineering from Universiti Putra Malaysia (UPM), in 2020. His research interests include machine learning, optimization, and computer vision.

LIST OF PUBLICATIONS

Saif F. Mahmood, M. H. Marhaban, F. Z. Rokhani, K. Samsudin and O. A. Arigbabu: SVM-ELM: Pruning of Extreme Learning Machine with Support Vector Machines for Regression, Journal of Intelligent Systems, pp. 1-12, 2015.

Mahmood, S. F., Marhaban, M. H., Rokhani, F. Z., Samsudin, K., & Arigbabu, O. A. (2017). FASTA-ELM: a fast adaptive shrinkage/thresholding algorithm for extreme learning machine and its application to gender recognition. Neurocomputing, 219, 312-322.

Saif F. Mahmood and Keem Siah Yap: Tuning of Fuzzy Controller for Antenna Tracking Satellite Using Ge netic Algorithm, presented at the National Graduate Conference 2012 (NatGrad2012), Nov 8-10, 2012.

Olasimbo Ayodeji Arigbabu, Sharifah Mumtazah Syed Ahmad, Wan Azizun Wan Adnan, Salman Yussof, Saif Mahmood: Soft biometrics: gender recognition from unconstrained face images using local feature descriptor. arXiv preprint arXiv:1702.02537

Olasimbo Ayodeji Arigbabuc, Saif Mahmood, Sharifah Mumtazah Syed Ahmad, and Abayomi A. Arigbabu. "Smile detection using hybrid face representation." Journal of Ambient Intelligence and Humanized Computing 7, no. 3 (2016):415-426.