HYBRID META-HEURISTIC ALGORITHM BASED PARAMETER OPTIMIZATION FOR EXTREME LEARNING MACHINES CLASSIFICATION

OYEKALE ABEL ALADE

A thesis submitted in fulfilment of the
requirements for the award of the degree of
Doctor of Philosophy

School of Computing
Faculty of Engineering
Universiti Teknologi Malaysia

JULY 2021

DEDICATION

This thesis is dedicated to my wife Ruth Adishetu Alade,
our daughter Eunice Oluwatomiwa Odewuyi
and the entire Alade's family.

ACKNOWLEDGEMENT

First and foremost, all praise and thanks are due to Almighty God. Next, I wish to express my sincere appreciation to my main supervisor, Associate Professor Dr. Roselina Salleh@Sallehuddin, for her encouragement, guidance, critiques, support and friendship from the first day till the end of this research work. I am also very thankful to my Co-supervisor in person of Dr. Nor Haizan Bt Mohamed Radzi for her patience and considerate nature that made her accessible whenever I needed her assistance. More so, I am indebted to Professor Dr. Ali Selamat for his contributions and support to this research. I indeed thank them for showing me how to identify interesting problems and how the research can be started and finished correctly.

I wish to take this opportunity to thank my ever-supportive wife Mrs. Ruth Adishetu Alade for her unending prayers and encouragements throughout this journey. Special appreciation goes to my parents: Pa Emmanuel Oyebode Alade (Father) and Mama Janet Tewogbade Alade (Mother). Also, I am indebted to Miss Eunice Oluwatomiwa Odewuyi, Mr. and Mrs Timothy Oyerinde Alade, Mr. & Mrs. Elijah ’Wale Oyebode, Mr and Mrs. Sunday Oyeniyi Alade, Mrs. Emily Oyejide, Mrs. Rodah Odewuyi, Mrs. Jumoke Opeola, my other siblings, the family of Dr. and Dr. (Mrs.) Okegbile and Bro James Samuel. It is my sincere prayer that the Almighty God will reward you abundantly in the name of Jesus Christ.

I also recognize and appreciate the support of the families of Mr. and Mrs. Jeremiah Peter Abiodun for holding forth for my family, Pastor Dr. and Mrs. Oladokun Olagoke and the entire family of the Redeemed Christian Church of God (Power Palace) Taman Teratai, Skudai Johor Bahru Malaysia for their encouragements, prayers and spiritual supports all through this journey. Also, worthy of appreciation are the Pastor and members of Shalom Baptist Church, Bida Nigeria. The Lord will reward your labour of love in Jesus name. My sincere appreciation also goes to all the people that have added values to my life in one way or the other. Finally, special appreciation goes to the Nigerian Tertiary Education Trust Fund (TETFund) for her support via the Federal Polytechnic, Bida Niger State, Nigeria, and the staff of the Department of Computer Science, FPB, Bida, Nigeria.

ABSTRACT

Most classification algorithms suffer from manual parameter tuning and it affects the training computational time and accuracy performance. Extreme Learning Machines (ELM) emerged as a fast training machine learning algorithm that eliminates parameter tuning by randomly assigning the input weights and biases, and analytically determining the output weights using Moore Penrose generalized inverse method. However, the randomness assignment, does not guarantee an optimal set of input weights and biases of the hidden neurons. This will lead to ELM instability and local minimum solution. ELM performance also is affected by the network structure especially the number of hidden nodes. Too many hidden neurons will increase the network structure complexity and computational time. While too few hidden neuron numbers will affect the ELM generalization ability and reduce the accuracy. In this study, a heuristic-based ELM (HELM) scheme was designed to secure an optimal ELM structure. The results of HELM were validated with five rule-based hidden neuron selection schemes. Then HELM performance was compared with Support Vector Machine (SVM), k-Nearest Neighbour (KNN), and Classification and Regression Tree (CART) to investigate its relative competitiveness. Secondly, to improve the stability of ELM, the Moth-Flame Optimization algorithm is hybridized with ELM as MFO-ELM. MFO generates moths and optimizes their positions in the search space with a logarithm spiral model to obtain the optimal values of input weights and biases. The optimal weights and biases from the search space were passed into the ELM input space. However, it did not completely solve the problem of been stuck in the local extremum since MFO could not ensure a good balance between the exploration and exploitation of the search space. Thirdly, a co-evolutionary hybrid algorithm of the Cross-Entropy Moth-Flame Optimization Extreme Learning Machines (CEMFO-ELM) scheme was proposed. The hybrid of CE and MFO metaheuristic algorithms ensured a balance of exploration and exploitation in the search space and reduced the possibility of been trapped in the local minima. The performances of these schemes were evaluated on some selected medical datasets from the *University of California, Irvine (UCI) machine learning repository*, and compared with standard ELM, PSO-ELM, and CSO-ELM. The hybrid MFO-ELM algorithm enhanced the selection of optimal weights and biases for ELM, therefore improved its classification accuracy in a range of 0.4914 - 6.0762%, and up to 8.9390% with the other comparative ELM optimized meta-heuristic algorithms. The convergence curves plot show that the proposed hybrid CEMFO meta-heuristic algorithm ensured a balance between the exploration and exploitation in the search space, thereby improved the stability up to 53.75%. The overall findings showed that the proposed CEMFO-ELM provided better generalization performance on the classification of medical datasets. Thus, CEMFO-ELM is a suitable tool to be used not only in solving medical classification problems but potentially be used in other real-world problems.

ABSTRAK

Kebanyakan algoritma-algoritma klasifikasi menghadapi masalah penalaan parameter secara manual, dan ia mempengaruhi masa pengkomputeran latihan dan prestasi ketepatan. Mesin Pembelajaran Extrim (ELM) muncul sebagai algoritma pembelajaran mesin pantas yang menghapuskan penalaan parameter dengan pengumpukan secara rawak pengumpukan pemberat dan bias input, dan secara analitik menentukan pemberat menggunakan kaedah songsang umum Moore Penrose. Walau bagaimanapun, umpukan parameter secara rawak, tidak dapat menjamin nilai optimun bagi input dan bias neuron tersembunyi dan ini akan mengakibatkan ketidakstabilan ELM dan penyelesaian minima setempat. Prestasi ELM juga dipengaruhi oleh struktur rangkaian terutamanya bilangan neuron tersembunyi. Terlalu banyak bilangan neuron tersembunyi akan meningkatkan kompleksiti struktur rangkaian dan masa pengkomputeran sementara bilangan neuron tersembunyi terlalu sedikit akan menjejaskan keupayaan pengitlakan ELM dan mengurangkan ketepatan. Dalam kajian ini, skema ELM berasaskan heuristik (HELM) direka untuk memastikan struktur ELM yang optimum. Hasil HELM disahkan dengan lima skema aturan-asas penentuan neuron tersembunyi. Seterusnya, prestasi HELM dibandingkan dengan Mesin Sokongan Vektor (SVM), *k-Nearest Neighbour* (KNN) dan Pohon Pengelasan dan Regresi (CART) untuk mengkaji kebolehsaing relatifnya. Kedua, bagi meningkatkan kestabilan ELM, algoritma *Moth-Flame Optimization* dilakukan dengan ELM sebagai MFO-ELM. MFO menjanakan rama-rama dan mengoptimunkan kedudukan mereka dalam ruang carian menggunakan model logaritma lingkaran untuk mendapatkan nilai optima bagi input pemberat dan bias. Pemberat dan bias input yang optimum dari ruang carian diteruskan ke ruang input ELM Walua bagaimanapun, ia tidak menyelesaikan sepenuhnya masalah terperangkap di kawasan setempat kerana MFO tidak dapat memastikan keseimbangan yang baik antara eksploitasi dan eksplorasi ruang carian. Ketiga, algoritma gabungan evolusi-bersama iaitu skema *Cross-Entropy Moth-Flame Optimization Extreme Learning Machines* (CEMFO-ELM) dicadangkan. Gabungan algoritma metaheuristik CE dan MFO memastikan keseimbangan eksplorasi dan eksploitasi id ruang carian dan mengurangkan kebarangkalian untuk terperangkap dalam minima setempat. Prestasi skema ini dinilai pada beberapa set data perubatan terpilih dari machine learning repository University of California, Irvine (UCI), dan dibandingkan dengan ELM piawai, PSO-ELM dan CSO-ELM. Gabungan algoritma MFO-ELM telah meningkatkan keupayaan pemilihan nilai optima pemberat dan bias ELM, oleh itu ketepatan pengelasan ELM telah ditingkatkan antara julat 0.4914-6.0762%, dan sehingga 8.9390% berbanding dengan algoritma ELM pengoptima metaheuristik yang lain. Plot keluk penumpuan menunjukkan bahawa algoritma gabungan CEMFO yang dicadangkan memastikan keseimbangan antara eksplorasi dan eksploitasi dalam ruang carian, sehingga meningkatkan kestabilan hingga 53.75%. Dapatan keseluruhan menunjukkan bahawa CEMFO-ELM yang dicadangkan menghasilkan prestasi pengitlakan yang lebih baik ke atas pengelasan data perubatan. Oleh itu, CEMFO-ELM adalah alat yang sesuai untuk digunakan bukan hanya dalam menyelesaikan masalah klasifikasi perubatan tetapi juga berpotensi digunakan dalam masalah dunia nyata yang lain.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AG	-	Adaptive Growth
AIS	-	Artificial Immune System
ANN	-	Artificial Neural Network
BO	-	Bionic Optimization
BP	-	Back Propagation
BSO	-	Bat Swarm Optimization
CART	-	Classification And Regression Tree
CE	-	Cross-Entropy
CEMFO	-	Cross-Entropy Moth-Flame Optimization
CMFO	-	Chaotic Moth-Flame Optimization
CMWAO	-	Chaotic Multiple-Swarm Whale Optimization Algorithm
CNN	-	Convolution Neural Network
CRS	-	Cellular Robotic Systems
CSO	-	Competitive Swarm Optimization
DE	-	Differential Evolution
DGO	-	Dynamic Group Optimization
DT	-	Decision Tree
EA	-	Evolutionary Algorithm
EELM	-	Efficient Extreme Learning Machine
ELM	-	Extreme Learning Machines
ELM-AE	-	Extreme Learning Machines Auto-Encoder
EMD	-	Euclidian Minimum Distance
ESA	-	Evolutionary Simulated Annealing
FA	-	FireFly Algorithm
GA	-	Genetic Algorithm
GP	-	Genetic Programming
GWO	-	Grey Wolf Optimization
HS	-	Harmony Search
KNN	-	K-Nearest Neighbour
MA	-	Meta-heuristic Algorithms

MFO	-	Moth-Flame Optimization
MSCA	-	Modified Sine Cosine Algorithm
NMF	-	Non-negative Matrix Factorization
PSO	-	Particle Swarm Optimization
RBF	-	Radial Bases Function
ReLU	-	Rectified Linear Unit
RF	-	Random Forest
RMSE	-	Root Mean Square Error
SB	-	Sparse Bayesian
SI	-	Swarm Intelligence
SLFN	-	Single Layer Feedforward Neural Network
SS	-	Semi-Supervised
SSOE	-	Semi-Supervised Online Elastic
SVM	-	Support Vector Machine
UCI	-	University of California, Irvine
WCA	-	Water Cycle Algorithm

LIST OF SYMBOLS

σ	-	Standard deviation
d	-	Dimension
μ	-	Mean
v	-	Velocity
α	-	Pruning parameter
β	-	Output weight
r	-	Radius
b	-	Hidden neuron biases
G	-	Mapping function
$H^{\dagger}$	-	More Penrose inverse matrix of H
η	-	Learning rate
D	-	Distance
w	-	Input weights
φ	-	Control parameter

LIST OF APPENDICES

CHAPTER 1

INTRODUCTION

1.1 Overview of the Study

The heart of medical science is clinical diagnosis of human ailments. In recent years, there has been a dramatic increase in the use of computational techniques to analyse medical datasets. The general approach falls under artificial intelligence or machine learning, in which a computer program “learns” important features of a dataset to make predictions about other data that are not part of the training set (Tharwat, 2020). A classifier separates instances in a dataset into (usually) two or (rarely) more classes based on the attributes measured in each subject. Analysis of medical data and diagnosis of diseases involve the use of classifiers (Foster, Koprowski, Skufca, et al., 2014).

A major challenge in medical science is the early detection and treatment of diseases in patients. Doctors make mistakes when analyzing the symptoms of diseases (Fathurachman and Kalsum, 2014). Wrong diagnosis results in wrong treatments, which may have adverse effects and sometimes death of patients. Several techniques are used for this over the ages (Tsanas, Little, and Mcsharry, 2013). Human experts have been employed as diagnostic agents: Some human agents apply the heuristic approach to medical diagnosis – as in “*trado-medical*”, then orthodox medicine – using laboratory analysis of symptoms specimen from the patients. The advances in modern technology have brought new dimensions to medical diagnosis, yet there are various challenges in these tools. Therefore, automated systems can assist the doctors in clinical diagnosis based on established symptoms in the systems’ repository. Such systems will reduce the rate of wrong diagnosis and mortality rate.

Many automated techniques have been used by researchers in medical classification (Yassin, Omran, El Houby, et al., 2018). There are two broad groups of such techniques - statistical and soft computing methods. The statistical methods draw inferences from relationships that exist among various features of datasets, while the soft computing methods use computational algorithms on datasets to achieve the desired results. The marriage between statistics and computer science poses a computational challenge to building statistical models that can handle massive data to run billions or trillions of data points (Kotsavasiloglou, Kostikis, Hristu-Varsakelis, et al., 2017) and the draw expected results in less time with little or no human intervention. Statistical techniques like Euclidian minimum distance (EMD), quadratic minimum distance (QMD), and K-nearest neighbor, Bayesian decision theory, are used in constructing classifiers (Mohapatra, Chakravarty, and Dash, 2015). The challenge of statistical techniques is that their performance depends on the correctness of some underline assumptions for successful application. Therefore, the accuracy of statistical based classifiers is generally less than the soft computing methods. The soft computing techniques are implemented in machine learning algorithms.

Several machine learning algorithms are used for the classification of datasets in medical research. Some of them are Artificial Neural Network (ANN) (Shahid, Rappon, and Berta, 2019), Random Forest (RF) (Alam, Rahman, and Rahman, 2019), Support Vector Machines (SVM) (Wang and Chen, 2020), Multilayer Perceptions (MLP), k-Nearest Neighbors (KNN) and Bagging, etc. (Mangesh Metkari, 2014). The machine learning algorithms have *hyper-parameters* whose values cannot be estimated directly from datasets (Saporetti, Duarte, Fonseca, et al., 2019). These hyper-parameters have a great influence on the performance of the classification algorithms. However, most of these machine learning parameters have some issues in their learning processes - they are characterised by parameter tuning which makes them slow and gets stuck in local minimal, they could not reach optimal performance.

Huang proposed Extreme Learning Machines (ELM) to address the problem of parameter tuning (Huang, Zhu, and Siew, 2004). The ELM learning principle is essentially a linear model (Wang et al., 2017). ELM is a three-step training algorithm. It is simple, and it requires no tuning like the gradient-descent algorithms (Baron and

Zhang, 2020). ELM randomly assigns the connection weights and biases to the hidden neurons, then computes the output of the hidden neurons and analytically determines the output weights. The weights between the hidden nodes and the output neuron(s) are learned in a single step. Therefore, the parameters (weights and biases) of the hidden layer no longer require iterative tuning as they were in the conventional learning machines (Chen, Kloft, Yang, et al, 2018; Huang et al., 2006). ELM gains popularity because of its fast learning speed, which is far superior to the gradient descent based algorithms.

In recent times, researchers have exploited the usage of ELM in many real-world applications. Some of the application areas of ELM are computer vision (Albadr and Tiun, 2017; Wang, Tianlei, Cao, et al., 2018), image processing (Cao and Lin, 2015; López-Úbeda, Díaz-Galiano, Martín-Noguerol, et al., 2021), time series analysis (Yayık, Kutlu, and Altan, 2019), biomedical applications (Duan, Li, Yang, et al., 2018; Raghuwanshi and Shukla, 2020). In such research areas, ELM proves to achieve good generalization performance, while maintaining low computational cost. The main idea of ELM is to randomly generate the input weights of a single hidden layer feedforward neural network (SLFN) and analytically determine the output weights. Considering the high value of health research to society, and how valuable the generalization performance of ELM is, its efficient classification of medical datasets will have a direct and indirect impact on diagnosis, treatment, the pattern of health care, and functionality of public health intervention. Therefore, this research focuses on improving ELM to further exploit its numerous advantages for better classification results especially in medical datasets.

1.2 Problem Background

The existing study shows that different artificial intelligence (AI) methods have been employed by medical experts to assist them in the medical diagnosis of patients in recent times (Yassin, 2018). Many of these machine learning algorithms are single hidden layer feedforward neural networks (SLFN) (Albadr, 2017). The algorithms are based on gradient descent methods, using backpropagations (BP), to train SLFN (Decherchi, Gastaldo, Zunino, et al., 2013; Eshtay, Faris, and Obeid, 2018a). They are challenged with turning of the parameters. SLFN has been well studied and applied in various areas of machine learning, and it is commended by researchers for its capabilities and fault tolerance abilities (Da Silva and Krohling, 2018; Nayak, Dash, Majhi, et al., 2018). Although these algorithms are popular, they have the difficulties that they are highly dependent on the hyper-parameters (initial weights, biases, learning rate, etc.) of the network; they can easily be stuck in local minimal; and they are usually slow in convergence (Akusok, 2016; Ling, Song, Han, et al., 2019; Song, Chunning, Feng, et al., 2014). This is because the parameters require iterative tuning (Li, Shuai, You, et al., 2016).

Huang (2004) proposed Extreme Learning Machine (ELM) to address the challenge of parameter turning in gradient-descent algorithms. Many researchers embraced ELM, and it has been applied in many areas including medical diagnosis (Toprak, 2018). However, the acceptance of ELM has opened up some gaps for the improvements of the algorithm. It also requires a higher number of processing nodes in the hidden layer (Faris, Mirjalili, Aljarah, et al., 2020b) – that is, a complex network structure; instability of the network output (Eshtay, Faris, and Obeid, 2018b); and eventually a low degree of accuracy. Filling these gaps is being responded to in many ways by researchers.

ELM's complex network structure affects its performance of ELM (Salam, Zawbaa, Emary, et al., 2016). It reduces its response to unknown data. The basic ELM requires $\tilde{N}$ hidden neurons to train N distinct datasets (Huang, Li, Chen, et al., 2008; Huang, Huang, Song, et al., 2015). Many hidden neurons have little contribution to the performance of the architecture (Rong, Ong, Tan, et al., 2008). More so, too large a network leads to over-fitting and high cost of the network. This affects its deployment for some time-sensitive applications.

Several variants of ELM were proposed to ensure the optimal number of hidden neurons. Lekamalage, Kasun, Yang, et al., (2016) employed ELM Auto-Encoder (ELM-AE) for dimensional reduction. Their research investigated linear and non-linear ELM-AE and sparse ELM-AE that are based on orthogonal and sparse input neurons without tuning. They showed that ELM-AE and SELM-AE learn the between-class scatter matrix, and distance points within a cluster are reduced. More so, the learning of their algorithms is robust to noise, and the normalized mean square error (NMSE) is reduced. However, the sparsity of the algorithms is low and the computational time is high.

Regularization parameter has also been used to improve the compactness of ELM network architecture. Some of these approaches are based on ridge regression theory and weighted least squares (Deng, Zheng, and Chen, 2009). Martínez-Martínez et al. (2011) improved the work of Deng et al. They proposed the use of ridge regression, elastic net, and lasso methods to prune the size of hidden neurons in ELM architecture. Their work was validated with some regression benchmark tasks, and it was proved to scale a more compact network with a competitive result when compared with ELM. However, Inaba et al. (2018) appraised the generalization of their algorithms but shows that the ridge regularized ELM requires large memory space, and since large matrix inversion is involved, the computational cost is high. Therefore, they proposed the generalized regularized ELM (GR-ELM) approach for multiclass classification tasks. The approach combined the Frobenius norm and $\ell_{2,1}$norm of output weights as ELM penalty. The R-ELM was maintained for binary classification tasks. The Alternating Direction Method for Multiplier (ADMM) was used for implementation. They came up with a more compact network structure. However, the approach becomes more complex and the issue of computational cost remains unresolved.

Some static rule based approaches have been used in literature. The rules relate the network size to the number of features and/or the number of output neurons (Eshtay, Faris, and Obeid, 2020; Hecht-Nielsen, 1968; Masters, 1993; Sheela and Deepa, 2013). However, the rule based approaches are static and have not been proven that the network size of the SLFN depends on the number of features or the output nodes.

Some other improvements on ELM network structures are online-sequential ELM (Linag, Huang, Saratchandran, et al., 2006), Incremental-ELM (Huang, Li, Chen, et al, 2008), pruning-ELM (Rong, 2008), Voting based ELM (Huang et al., 2008), two-stage ELM (Zhao, Wang, and Park, 2012), Ensemble ELM (Albadr, 2017) and many more. Multiple tests of error comparisons are required by these methods to determine the optimal number of hidden neurons. These approaches are time-consuming (Tian, Ren, and Wang, 2019). More so, they could not adequately address the problem of how the optimal number of the hidden neuron is determined. The problem of over-fitting persists, and some of the approaches have little or no effect on the output. In this research, a randomized heuristic determination of optimum network structure for ELM classification is proposed.

ELM's random assignment of input weights and hidden neuron biases poses a negative challenge to the stability of ELM. Wang et al. (2011) proposed an Efficient Extreme Learning Machine (EELM) as a high-quality feature mapping algorithm. EELM makes a proper selection of input weights and biases before it calculates the output weight. The focus was to ensure a full column rank hidden neuron output H. Their work improved learning rates and ensured a robust network structure. Over-parameterized (a large number of hidden neurons) design of ELM usually results in an ill-conditioning problem (Dash and Patel, 2015; Janakiraman, Nguyen, and Assanis, 2016; Mohapatra, Chakravarty, and Dash, 2015b; Zhang, Yang, Cao, et al., 2018) Janakiraman et al (2016) attempt to have a bounded parameter via stochastic gradient descent. The results of the work were evaluated using the Lyapunov approach for error measure and the boundedness of the learning rates. Their work avoided the bad regularization of online learning for the identification of non-linear dynamic systems. Although these methods improve the performance of the ELM algorithm, it is still subject to over-fitting, it tends to fall into a local minimum, and the improvement on the accuracy of the algorithm is possible.

Optimization of the input weights and biases can improve the stability and accuracy of ELM (Ling et al., 2019; Maimaitiyiming et al., 2019; Tian, et al., 2019). Meta-heuristic algorithms are being used in recent times by researchers to optimize the parameter settings of ELM (Mohapatra et al., 2015, Eshtay et al (2018b)). The bio-inspired optimization techniques are better meta-heuristic algorithm options to optimize the parameters because they provide near optima solutions that are more acceptable to researchers (Hegazy, Makhlouf, and El-Tawel, 2019; Li, Shuang, Zhao, et al., 2019; Mirjalili et al., 2017). They are Particle Swarm Optimized ELM (PSO-ELM) (Vidhya and Kamaraj, 2017), Genetic Algorithm ELM (GA-ELM) (Yang, Yi, Zhao, et al., 2013), Cuckoo Search Optimization algorithm (CSO-ELM) (Mohapatra et al., 2015), FireFly algorithm (FA) (Su and Cai, 2016; Zhou and Jiao, 2017), Bat Swarm Optimization (BSO) (Alihodzic, Tuba, and Tuba, 2017), Artificial Bee Colony (ABC-ELM) (Wang, Wang, Ai, et al., 2017b), Artificial Immune System ELM (AIS-ELM) (Tian, Li, Wu, et Lal., 2018), Differential Evolution (DE-ELM) (Saporetti, 2019), Improved Grey Wolf Optimization (IGWO) model (Cai, Gu, Luo, et al., 2019), etc. Despite the relative success of these metaheuristics approaches in terms of flexibility and efficiency towards solution finding, they continue to suffer slow speed of convergence, and they are often trapped in local optimal (Liu, Liu, and Li, 2019) thereby affecting the efficiency of ELM. Yang and Duan (2020) proposed a hybrid model of Artificial Bee Colony (ABC) and Differential Evolution (DE) optimization techniques to improve the parameter selection of ELM. The model improved the generalization performance with less processing time offered by ELM. The deficiency of initial random assignment of input weights and biases was also improved, and the results of the classification were also improved. However, the exploitation of ABC is poor (Li, Liu, Le, et al., 2019) and the DE is computationally intensive (Shehab, Abualigah, Al Hamad, et al., 2019).

Moth Flame Optimization (MFO) was proposed by Mirjalili in 2015. The algorithm offers a competitive result compared with other state-of-the-arts metaheuristic optimization algorithms (Luo, Jie, Chen, et al., 2019; Pelusi, Mascella, Tallini, et al., 2020). However, the mechanism of position update and the convergent constant of MFO lay strong emphasis on exploitation rather than a one-to-one assignment of moth and flame provision for exploration (Khalilpourazari and Khalilpourazary, 2019). Each agent may be far away from the optimum point, this might increase uncertainty. The initial position of the algorithm might influence the properties of MFO to a certain degree. Therefore, it is necessary to strike a balance between the local and global search space for the efficient performance of the meta-heuristic algorithms and ensure the selection of optimal parameters for ELM. Xi et al (2019) proposed Gaussian mutation to improve the exploration of MFO, and chaotic exploitation to enhance the local search. The proposed Chaotic Local Search Gaussian mutation Moth-Flame Optimization (CLSGMFO) was used to optimize the kernel function and penalty coefficient of KELM. The algorithm was tested on 23 benchmark functions, and then applied to 2 financial datasets to prove its competitiveness with some other meta-heuristic algorithms. However, their proposed algorithm did not consider the selection of optimal input weights and biases. More so the evaluation of Gaussian mutation function is relatively expensive (Shehab, 2019).

Another promising meta-heuristic algorithm used to improve the exploration is the cross-entropy (CE) algorithm. It was used as an operator to improve the exploration of the FireFly Algorithm (FA) in (Li, 2019), and in the Bat algorithm (Li and Le, 2019). Thus, CE is promising for global optimization search. The algorithms fully absorb the ergodicity, adaptability, and robustness of the cross-entropy method. Therefore, to strike a balance between the exploration and exploitation in the optimization search space, this research proposes to embed the CE into MFO as an exploration operator. This would bring a new co-evolutionary hybrid algorithm called Extreme Learning Machines based Cross-Entropy Moth-Flame Optimization (CEMFO) scheme. Therefore CEMFO is proposed to balance the exploration and exploitation of the optimization search space to select optimal input weights and biases for ELM and ensure improved stability and accuracy of the ELM classification algorithm. The improvement on the ELM classifier will have a contribution to the classification of medical datasets

1.3 Problem Statement

ELM is a promising classification algorithm for the classification of medical datasets. However, the size of the network architecture and the initial settings of the input weights and biases are key issues that greatly affect the overall performance of ELM classification algorithms (He, Liu, Wu, et al., 2019). ELM architecture requires a high number of hidden neurons to have a good performance. This reduces its response to unknown data. Although the rule based approaches have been used to determine the optimal network size, this could not guarantee optimal network structure. More so, the poor initial setting of the input weights and biases makes ELM ill-conditioned, this affects the stability and accuracy of the output of ELM (Eshtay, 2018a). Also, meta-heuristic techniques are employed to select optimal input parameters for the machine learning classification algorithms. However, the imbalance between exploration and exploitation in the metaheuristics algorithm is another issue that can lead to poor-quality solutions. Moreover, most ELM individual based optimization algorithms suffer slow convergence and are stuck in the local optimum (Li and Le, 2019; Yang and Duan, 2020). The resultant effect is the selection of poor input weights and biases for the ELM classification algorithm. Therefore, this research proposes heuristic, meta-heuristic, and co-evolutionary meta-heuristic optimization techniques to improve the network complexity, stability, and exploration and exploitation of optimization search space for Extreme Learning Machines in solving medical classification problems.

Consequently, the hypothesis of this research can be stated as follows:

The performance of extreme learning machines (ELM) in classification could be enhanced with heuristic, meta-heuristic, and co-evolution of meta-heuristic optimization schemes to ensure optimal network structure, stable algorithm, and better accuracy.

1.4 Research Questions

To address the problems of classification specified above, the following research questions are presented:

i. How a heuristic scheme can determine an optimum number of hidden neurons to ensure a compact network structure of the Extreme Learning Machines classification algorithm?

ii. How Moth-Flame optimization algorithm can improve the selection of the input weights and biases of Extreme Learning Machines?

iii. How a hybrid of Cross-Entropy and Moth-Flame optimization techniques could balance the exploration and exploitation of the optimization search space for optimal parameter selection for Extreme Learning Machines?

1.5 Aim of the Research

This research aims to improve the generalization performance of Extreme Learning Machines in terms of a compact network structure, stability, and accuracy using a hybrid Cross-Entropy Moth-Flame-Extreme Learning Machine (CEMFO-ELM).

1.6 Research Objectives

Based on the problem statement and the aim of this research, the research objectives are set as follows:

i. To determine a compact and efficient network size of hidden neurons with a heuristic scheme to improve the performance of Extreme Learning Machines classification.

ii. To enhance the selection of optimum input weights and biases of Extreme Learning Machines classification algorithm with Moth-Flame Optimization algorithm to improve its stability and accuracy.

iii. To propose a hybrid Cross-Entropy Moth-Flame Optimization (CE-MFO) algorithm to balance the exploration and exploitation of the search space for the selection of optimal parameters for Extreme Learning Machines classification.

1.7 Research Scope

This research is proposed within the following scopes:

i. Five (5) medical datasets are used in this research. They are Blood, Breast cancer, Pima Indian diabetes, Bupa liver, and Phoneme datasets. The key feature of these datasets is that they have linearly non-separable separable distribution. Only binary class datasets are considered in this study.

ii. The research focuses on building a compact network size, improved stability, and accuracy of ELM for selected datasets.

iii. Although ELM can be trained with any continuous piece-wise activation function, only sigmoid is used in all the simulations to ensure consistency.

iv. The study considers meta-heuristics algorithms for enhancing ELM classification tasks. It proposes a hybrid of two meta-heuristic algorithms as a co-evolutionary algorithm to balance the exploration and exploitation of the search space of the meta-heuristic algorithms.

v. The proposed algorithm is implemented on MATLAB running on Windows 10 – 64-bit operating systems install on core i7 CPU @ 1.90GHz.

1.8 The Organization of the Thesis

The rest of the thesis is organized as outlined below:

Chapter 2: Literature related to this research is reviewed in order to formulate the research problem. The concepts of classification and classification models, the trend in ELM enhancement leading to the direction of this research are presented. Chapter 3 presents the methodology of this research is presented. The problems are defined, the proposed solutions are designed, datasets are described and the evaluation metrics used for the results are presented. In this chapter 4, a heuristic approach is used to determine the optimal learning structure of ELM. The optimal structures are validated by rule based network structures. The performances are compared with three (3) other machine learning algorithms namely KNN, SVM, and CART on five (5) machine learning medical classification datasets – Blood, Breast cancer, Diabetes, Bupa Liver, and Phoneme. The results are analysed on three evaluation criteria: accuracy, computational time and standard deviation. Chapter 5 proposed Moth-Flame Optimization to enhance the selection of parameters for ELM. The performance of the algorithm is measured using standard classification performance metrics. Chapter 6: This chapter proposed a hybrid of Cross-Entropy and Moth-Flame Optimization to enhance Extreme Learning Machines (CEMFO-ELM). This ensures a balance between exploration and exploitation of the search algorithms. The proposed hybrid algorithm is described and the results are discussed. Chapter 7 conclusions the research.

REFERENCES

Akusok, A. (2016). *Extreme Learning Machines : novel extensions and application to Big Data*.

Alam, M. Z., Rahman, M. S., and Rahman, M. S. (2019). A Random Forest based predictor for medical data classification using feature ranking. *Informatics in Medicine Unlocked*, 15(April).

Albadr, M. A. A., and Tiun, S. (2017). Extreme learning machine: A review. *International Journal of Applied Engineering Research*, 12(14), pp. 4610–4623.

Alihodzic, A., Tuba, E., and Tuba, M. (2017). *An upgraded bat algorithm for tuning extreme learning machines for data classification*. pp. 125–126.

Anuar, M. S. (2016). *An enhance Artifical Bee Colony for Optimiztaion Problems*. Universiti Teknologi Malaysia.

Arasomwan, M. A., and Adewumi, A. O. (2013). On the performance of linear decreasing inertia weight particle swarm optimization for global optimization. *The Scientific World Journal*, 2013, pp. 1–12.

Avci, D., and Dogantekin, A. (2016). An Expert Diagnosis System for Parkinson Disease Based on Genetic Algorithm-Wavelet Kernel-Extreme Learning Machine. *Parkinson's Disease*, 2016, pp. 1–9.

Bai, Z., Huang, G. Bin, Wang, D., Wang, H., and Westover, M. B. (2014). Sparse Extreme Learning Machine for Classification. *IEEE Transactions on Cybernetics*, 44(10), pp. 1.

Baron, C. M. C., and Zhang, J. (2020). Reliable on-line re-optimization control of a fed-batch fermentation process using bootstrap aggregated extreme learning machine. In O. Gusikhin & K. Madani (Eds.), *Lecture Notes in Electrical Engineering* (Vol. 495, pp. 272–294). Springer International Publishing.

Bataineh, A. Al. (2019). A comparative analysis of nonlinear machine learning algorithms for breast cancer detection. *International Journal of Machine Learning and Computing*, 9(3), pp. 248–254.

Beni, G., and Wang, J. (1993). Swarm Intelligence in Cellular Robotic Systems. *Robots and Biological Systems: Towards a New Bionics?*, *2*, pp. 703–712.

Bonyadi, M. R., and Michalewicz, Z. (2017). Particle Swarm Optimization for Single Objective Continuous Space Problems: A Review. *Evolutionary Computation*, 25(1), pp. 1–54.

Botev, Z. I., Kroese, D. P., Rubinstein, R. Y., and L'Ecuyer, P. (2013). The Cross-Entropy Method for Optimization From Estimation to Optimization. *Handbook of Statistics*, pp. 35–59.

Cai, Z., Gu, J., Luo, J., Zhang, Q., Chen, H., Pan, Z., Li, Y., and Li, C. (2019). Evolving an optimal kernel extreme learning machine by using an enhanced grey wolf optimization strategy. *Expert Systems with Applications*, 138, pp. 112814.

Cao, J., and Lin, Z. (2015). Extreme Learning Machines on High Dimensional and Large Data Applications: A Survey. *Mathematical Problems in Engineering*, 2015, pp. 16–18.

Cao, J., Lin, Z., Huang, G. Bin, and Liu, N. (2012). Voting based extreme learning machine. *Information Sciences*, 185(1), pp. 66–77.

Chen, X., Chen, Y., Zomaya, A. Y., Ranjan, R., and Hu, S. (2016). CEVP: Cross Entropy based Virtual Machine Placement for Energy Optimization in Clouds. *Journal of Supercomputing*, 72(8), pp. 3194–3209.

Chen, Y., Kloft, M., Yang, Y., Li, C., and Li, L. (2018). Mixed kernel based extreme learning machine for electric load forecasting. *Neurocomputing*, 312, pp. 90–106.

Cheng, R., and Jin, Y. (2015). A competitive swarm optimizer for large scale optimization. *IEEE Transactions on Cybernetics*, 45(2), pp. 191–204.

Da Silva, C. A. S., and Krohling, R. A. (2018). Semi-Supervised Online Elastic Extreme Learning Machine for Data Classification. *Proceedings of the International Joint Conference on Neural Networks*, 2018-July, pp. 1–8.

Da Silva, C. A. S., and Krohling, R. A. (2019). Semi-Supervised Online Elastic Extreme Learning Machine with Forgetting Parameter to deal with concept drift in data streams. *Proceedings of the International Joint Conference on Neural Networks*, 2019-July(July), pp. 1–8.

Dai, H., Cao, J., Wang, T., Deng, M., and Yang, Z. (2019). Multilayer one-class extreme learning machine. *Neural Networks*, 115, pp. 11–22.

Das, S. R., Kuhoo, Mishra, D., and Rout, M. (2019). An optimized feature reduction based currency forecasting model exploring the online sequential extreme learning machine and krill herd strategies. *Physica A: Statistical Mechanics* *and Its Applications*, 513, pp. 339–370.

Das, S. R., Mishra, D., and Rout, M. (2017). A hybridized ELM-Jaya forecasting model for currency exchange prediction. *Journal of King Saud University - Computer and Information Sciences*.

Dash, S. K., and Patel, D. (2015). Short-term electric load forecasting using Extreme Learning Machine - a case study of Indian power market. *2015 IEEE Power, Communication and Information Technology Conference (PCITC)*, pp. 961–966.

Decherchi, S., Gastaldo, P., Zunino, R., Cambria, E., and Redi, J. (2013). Circular-ELM for the reduced-reference assessment of perceived image quality. *Neurocomputing*, 102, pp. 78–89.

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

Deng, W. Y., Bai, Z., Huang, G. Bin, and Zheng, Q. H. (2016). A Fast SVD-Hidden-nodes based Extreme Learning Machine for Large-Scale Data Analytics. *Neural Networks*, 77, pp. 14–28.

Deng, W., Zheng, Q., and Chen, L. (2009). Regularized extreme learning machine. *2009 IEEE Symposium on Computational Intelligence and Data Mining, CIDM 2009 - Proceedings*, *60825202*, pp. 389–395.

Duan, M., Li, K., Yang, C., and Li, K. (2018). A hybrid deep learning CNN–ELM for age and gender classification. *Neurocomputing*, 275, pp. 448–461.

Eftimov, T., and Korošec, P. (2019). Understanding exploration and exploitation powers of meta-heuristic stochastic optimization algorithms through statistical analysis. *GECCO 2019 Companion - Proceedings of the 2019 Genetic and Evolutionary Computation Conference Companion*, pp. 21–22.

Eshtay, M., Faris, H., Heidari, A. A., Al-Zoubi, A. M., and Aljarah, I. (2020). AutoRWN: automatic construction and training of random weight networks using competitive swarm of agents. *Neural Computing and Applications*, 0.

Eshtay, M., Faris, H., and Obeid, N. (2018a). Improving Extreme Learning Machine by Competitive Swarm Optimization and its application for medical diagnosis problems. *Expert Systems with Applications*, 104, pp. 134–152.

Eshtay, M., Faris, H., and Obeid, N. (2018b). Metaheuristic-based extreme learning machines: a review of design formulations and applications. *International*

Journal of Machine Learning and Cybernetics, 0(0), pp. 0.

Eshtay, M., Faris, H., and Obeid, N. (2020). A competitive swarm optimizer with hybrid encoding for simultaneously optimizing the weights and structure of Extreme Learning Machines for classification problems. *International Journal of Machine Learning and Cybernetics, 0123456789*.

Falcone, R., Lima, C., and Martinelli, E. (2020). Soft computing techniques in structural and earthquake engineering: a literature review. *Engineering Structures*, 207(June 2019), pp. 110269.

Faris, H., Mirjalili, S., Aljarah, I., Mafarja, M., and Heidari, A. A. (2020a). *Nature-Inspired Optimizers* (S. Mirjalili, J. Song Dong, & A. Lewis (eds.); Vol. 811). Springer International Publishing.

Faris, H., Mirjalili, S., Aljarah, I., Mafarja, M., and Heidari, A. A. (2020b). *Salp Swarm Algorithm: Theory,Literature Review, and Applicationin Extreme Learning Machines* (Vol. 811). Springer International Publishing.

Fathurachman, M., and Kalsum, U. (2014). Heart disease diagnosis using extreme learning based neural networks. *International Conference of Advanced Informatics: Concept, Theory and Application (ICAICTA)*, pp. 23–27. http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=7005909

Foster, K. R., Koprowski, R., Skufca, J. D., Broadhurst, D., Kell, D., Duda, R., Hart, P., Stork, D., Breiman, L., Friedman, J., Olshen, R., Stone, C., Tadeusiewicz, R., Ogiela, M., Kunchewa, L., Mitchel, T., Schapire, R., Freund, Y., Schapire, R., … Zermeno, A. (2014). Machine learning, medical diagnosis, and biomedical engineering research - commentary. *BioMedical Engineering OnLine*, 13(1), pp. 94.

Frénay, B., and Verleysen, M. (2016). *Reinforced Extreme Learning Machines for Fast Robust Regression in the Presence of Outliers*. 46(12), pp. 3351–3363.

Gabi, D., Ismail, A. S., Zainal, A., Zakaria, Z., and Abraham, A. (2018). Orthogonal Taguchi-based cat algorithm for solving task scheduling problem in cloud computing. *Neural Computing and Applications*, 30(6), pp. 1845–1863.

García-Flores, R., Sparks, R., Munro, D., and McCubbin, A. (2015). Determining the optimal number of beds in the subacute section of a large hospital. *Proceedings - 21st International Congress on Modelling and Simulation, MODSIM 2015*, pp. 1647–1653.

García, S., Luengo, J., Herrera, F., García, S., Luengo, J., and Herrera, F. (2015). Data

Preprocessing in Data Mining. In *Intelligent Systems Reference Library* (Vol. 72).

Geem, Z. W., and Kim, J. H. (2001). A New Heuristic Optimization Algorithm: Harmony Search. In *Optimization*.

Glover, F. (1989). Tabu Search—Part I. *ORSA Journal on Computing*, 1(3), pp. 190–206.

Gou, J., Ma, H., Ou, W., Zeng, S., Rao, Y., and Yang, H. (2019). A generalized mean distance-based k-nearest neighbor classifier. *Expert Systems with Applications*, 115, pp. 356–372.

Hassanien, A. E., Gaber, T., Mokhtar, U., and Hefny, H. (2017). An improved moth flame optimization algorithm based on rough sets for tomato diseases detection. *Computers and Electronics in Agriculture*, 136, pp. 86–96.

He, C., Liu, C., Wu, T., Xu, Y., Wu, Y., and Chen, T. (2019). Medical rolling bearing fault prognostics based on improved extreme learning machine. *Journal of Combinatorial Optimization*.

Hecht-Nielsen, R. (1968). *Kolmogorov's mapping Neurral Network Exisitence Theorem*. 4(September), pp. 95–98.

Hegazy, A. E., Makhlouf, M. A., and El-Tawel, G. S. (2019). Feature Selection Using Chaotic Salp Swarm Algorithm for Data Classification. *Arabian Journal for Science and Engineering*, 44(4), pp. 3801–3816.

Huang, Gao, Song, Shiji, Gupta, Jatinder N D, and Wu, C. (2014). Semi-supervised and unsupervised extreme learning machines. *IEEE Transactions on Cybernetics*, 44(12), pp. 2405–2417.

Huang, G. Bin, and Chen, L. (2008). Enhanced random search based incremental extreme learning machine. *Neurocomputing*, 71(16–18), pp. 3460–3468.

Huang, G. Bin, Li, M. Bin, Chen, L., and Siew, C. K. (2008). Incremental extreme learning machine with fully complex hidden nodes. *Neurocomputing*, 71(4–6), pp. 576–583.

Huang, G. Bin, and Siew, C. (2004). Extreme Learning Machine: RBF Network Case. *International Conference on Control, Automation, Robotics and Vision, December*, pp. 6–9.

Huang, G. Bin, Zhou, H., Ding, X., and Zhang, R. (2012). Extreme learning machine for regression and multiclass classification. *IEEE Transactions on Systems, Man, and Cybernetics. Part B, Cybernetics*, 42(2), pp. 513–529.

Huang, G. Bin, Zhu, Q., and Siew, C. (2004). Extreme Learning Machine : A New Learning Scheme of Feedforward Neural Networks. *IEEE International Joint Conference on Neural Networks*, 2, pp. 985–990.

Huang, G. Bin, Zhu, Q., Siew, C., Huang, G. Bin, Zhu, Q., Siew, C., Zhu, Q., and Siew, C. (2006). Extreme learning machine: Theory and applications. *Neurocomputing*, 70(1–3), pp. 489–501.

Huang, G., Huang, G. Bin, Song, S., and You, K. (2015). Trends in extreme learning machines: A review. *Neural Networks*, 61, pp. 32–48.

Inaba, F. K., Salles, T. E. O., Perron, S., and Caporossi, G. (2018). DGR-ELM–Distributed Generalized Regularized ELM for classification. *Neurocomputing*, 275, pp. 1522–1530.

Jain, D., and Singh, V. (2018). Feature selection and classification systems for chronic disease prediction: A review. *Egyptian Informatics Journal*.

Janakiraman, V. M., Nguyen, X. L., and Assanis, D. (2016). Stochastic gradient based extreme learning machines for stable online learning of advanced combustion engines. *Neurocomputing*, 177, pp. 304–316.

Jayaweera, C. D., Othman, M. R., and Aziz, N. (2019). Improved predictive capability of coagulation process by extreme learning machine with radial basis function. *Journal of Water Process Engineering*, 32(July), pp. 100977.

Kandhway, P., and Bhandari, A. K. (2019). Spatial context cross entropy function based multilevel image segmentation using multi-verse optimizer. *Multimedia Tools and Applications*, 78(16), pp. 22613–22641.

Karaboga, D., Gorkemli, B., and Ozturk, C. (2014). *A comprehensive survey : artificial bee colony (ABC) algorithm and applications*. pp. 21–57.

Kennedy, J., and Eberhart, R. (1995). *Particle Swarm Optimization*. pp. 1942–1948.

Khalilpourazari, S., and Khalilpourazary, S. (2019). An efficient hybrid algorithm based on Water Cycle and Moth-Flame Optimization algorithms for solving numerical and constrained engineering optimization problems. *Soft Computing*, 23(5), pp. 1699–1722.

Kotsavasiloglou, C., Kostikis, N., Hristu-Varsakelis, D., and Arnaoutoglou, M. (2017). Machine learning-based classification of simple drawing movements in Parkinson's disease. *Biomedical Signal Processing and Control*, 31, pp. 174–180.

Lan, Y., Soh, Y. C., and Huang, G. Bin. (2009). Ensemble of online sequential extreme learning machine. *Neurocomputing*, 72(13–15), pp. 3391–3395.

Lekamalage, L., Kasun, C., Yang, Y., Huang, G. Bin, and Zhang, Z. (2016). Dimension Reduction With Extreme Learning Machine. *Ieee Transactions on Image Processing*, 25(8), pp. 3906–3918.

Li, G., and Le, C. (2019). Hybrid Binary Bat Algorithm with Cross-Entropy Method for Feature Selection. *4th International Conference on Control and Robotics Engineering, ICCRE 2019*, pp. 165–169.

Li, G., Liu, P., Le, C., and Zhou, B. (2019). A novel hybrid meta-heuristic algorithm based on the cross-entropy method and firefly algorithm for global optimization. *Entropy*, 21(5).

Li, G., Shuang, F., Zhao, P., and Le, C. (2019). An improved butterfly optimization algorithm for engineering design problems using the cross-entropy method. *Symmetry*, 11(8).

Li, S., You, Z. H., Guo, H., Luo, X., and Zhao, Z. Q. (2016). Inverse-free extreme learning machine with optimal information updating. *IEEE Transactions on Cybernetics*, 46(5), pp. 1229–1241.

Liang, J. H., and Lee, C. H. (2015). A modification artificial bee colony algorithm for optimization problems. *Mathematical Problems in Engineering*, 2015.

Lichman, M. (2013). *UCI Machine Learning Repository*. http://archive.ics.uci.edu/ml

Linag, N.-Y., Huang, G. Bin, Saratchandran, P., and Sundararajan, N. (2006). A Fast and Accurate Online Sequential Learning Algorithm for Feedforward Networks. *IEEE Transactions on Neural Networks*, 17(6), pp. 1411–1423.

Ling, Q. H., Song, Y. Q., Han, F., Zhou, C. H., and Lu, H. (2019). An improved learning algorithm for random neural networks based on particle swarm optimization and input-to-output sensitivity. *Cognitive Systems Research*, 53, pp. 51–60.

Liu, J., Liu, L., and Li, Y. (2019). A differential evolution flower pollination algorithm with dynamic switch probability. *Chinese Journal of Electronics*, 28(4), pp. 737–747.

López-Úbeda, P., Díaz-Galiano, M. C., Martín-Noguerol, T., Luna, A., Ureña-López, L. A., and Martín-Valdivia, M. T. (2021). Automatic medical protocol classification using machine learning approaches. *Computer Methods and Programs in Biomedicine*, 200.

Luo, J., Chen, H., Hu, Z., Huang, H., Wang, P., Wang, X., Lv, X. E., and Wen, C.

(2019). A new kernel extreme learning machine framework for somatization disorder diagnosis. *IEEE Access*, 7, pp. 45512–45525.

Luo, J., Vong, C.-M., and Wong, P.-K. (2014). Sparse Bayesian extreme learning machine for multi-classification. *IEEE Transactions on Neural Networks and Learning Systems*, 25(4), pp. 836–843.

Ma, C. (2017). An efficient optimization method for extreme learning machine using artificial bee colony. *Journal of Digital Information Management*, 15(3), pp. 135–147.

Ma, J., Yang, L., Wen, Y., and Sun, Q. (2020). Twin minimax probability extreme learning machine for pattern recognition. *Knowledge-Based Systems*, 187, pp. 104806.

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

Maimaitiyiming, M., Sagan, V., Sidike, P., and Kwasniewski, M. T. (2019). Dual activation function-based Extreme Learning Machine (ELM) for estimating grapevine berry yield and quality. *Remote Sensing*, 11(7).

Mangesh Metkari, M. A. P. (2014). Comparative Study of Soft Computing Techniques on Medical Datasets. *International Journal of Science and Research (IJSR)*, 3(12), pp. 761–765.

Martínez-Martínez, J. M., Escandell-Montero, P., Soria-Olivas, E., Martín-Guerrero, J. D., Magdalena-Benedito, R., and Gómez-Sanchis, J. (2011). Regularized extreme learning machine for regression problems. *Neurocomputing*, 74(17), pp. 3716–3721.

Masters, T. (1993). Practical neural network recipes in C++. *Morgan Kaufmann, Burlington*.

Mirjalili, S. (2015). Moth-flame optimization algorithm: A novel nature-inspired heuristic paradigm. *Knowledge-Based Systems*, 89, pp. 228–249.

Mirjalili, S., Gandomi, A. H., Mirjalili, S. Z., Saremi, S., Faris, H., and Mirjalili, S. M. (2017). Salp Swarm Algorithm: A bio-inspired optimizer for engineering design problems. *Advances in Engineering Software*, 114, pp. 163–191.

Mohapatra, P., Chakravarty, S., and Dash, P. K. (2015a). An improved cuckoo search based extreme learning machine for medical data classification. *Swarm and*

Evolutionary Computation, 24, pp. 25–49.

Mohapatra, P., Chakravarty, S., and Dash, P. K. (2015b). An improved cuckoo search based extreme learning machine for medical data classification. *Swarm and Evolutionary Computation*, 24, pp. 25–49.

Mohapatra, P., Nath Das, K., and Roy, S. (2017). A modified competitive swarm optimizer for large scale optimization problems. *Applied Soft Computing Journal*, 59, pp. 340–362.

Nahato, K. B., Nehemiah, K. H., and Kannan, A. (2016). Hybrid approach using fuzzy sets and extreme learning machine for classifying clinical datasets. *Informatics in Medicine Unlocked*, 2, pp. 1–11.

Nayak, D. R., Dash, R., Majhi, B., and Wang, S. (2018). Combining extreme learning machine with modified sine cosine algorithm for detection of pathological brain. *Computers and Electrical Engineering*, 68(November 2017), pp. 366–380.

Ozkan, I. A., Koklu, M., and Sert, I. U. (2018). Diagnosis of urinary tract infection based on artificial intelligence methods. *Computer Methods and Programs in Biomedicine*, 166, pp. 51–59.

Pekel Özmen, E., and Özcan, T. (2020). Diagnosis of diabetes mellitus using artificial neural network and classification and regression tree optimized with genetic algorithm. *Journal of Forecasting*, pp. 0–1.

Pelusi, D., Mascella, R., Tallini, L., Nayak, J., Naik, B., and Deng, Y. (2020). An Improved Moth-Flame Optimization algorithm with hybrid search phase. *Knowledge-Based Systems*, 191, pp. 105277.

Qing, Y., Zeng, Y., Li, Y., and Huang, G. Bin. (2020). Deep and wide feature based extreme learning machine for image classification. *Neurocomputing*, 412, pp. 426–436.

Raghuwanshi, B. S., and Shukla, S. (2020). SMOTE based class-specific extreme learning machine for imbalanced learning. *Knowledge-Based Systems*, 187, pp. 104814.

Repley, B. D. (1993). Statistical aspects of neural networks. Networks and chaos?statistical and probabilistic aspects 50:40–123. In 50 (Ed.), *Neural Networks* (pp. 40–123). Chapman & Hall.

Rodzin, S., and Rodzina, L. (2014). Theory of bionic optimization and its application to evolutionary synthesis of digital devices. *Proceedings of IEEE East-West*

Design and Test Symposium, EWDTS 2014.

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

Salam, M. A., Zawbaa, H. M., Emary, E., Ghany, K. K. A., and Parv, B. (2016). A hybrid dragonfly algorithm with extreme learning machine for prediction. *Proceedings of the 2016 International Symposium on INnovations in Intelligent SysTems and Applications, INISTA 2016*, pp. 1–6.

Saporetti, C. M., Duarte, G. R., Fonseca, T. L., Da Fonseca, L. G., and Pereira, E. (2019). Extreme Learning Machine combined with a Differential Evolution algorithm for lithology identification. *Revista de Informática Teórica e Aplicada*, 25(4), pp. 43.

Shahid, N., Rappon, T., and Berta, W. (2019). Applications of artificial neural networks in health care organizational decision-making: A scoping review. *PLoS ONE*, 14(2), pp. 1–22.

Sheela, K. G., and Deepa, S. N. (2013). *Review on Methods to Fix Number of Hidden*. 2013.

Shehab, M., Abualigah, L., Al Hamad, H., Alabool, H., Alshinwan, M., and Khasawneh, A. M. (2019). Moth–flame optimization algorithm: variants and applications. *Neural Computing and Applications*, 6.

Smith, J. W., Everhart, J. E., Dickson, W. C., Knowler, W. C., and Johannes, R. S. (1988). Using the ADAP Learning Algorithm to Forecast the Onset of Diabetes Mellitus. *Proceedings of the Annual Symposium on Computer Application in Medical Care*, pp. 261–265.

Song, C., Feng, L., Lin, X., and Li, H. (2014). Artificial Bee Colony (ABC) Optimization Algorithm for Training Extreme Learning Machine. *Communications in Control Science and Engineering*, 2(El M), pp. 64–71. www.as-se.org/ccse

Strack, B., Deshazo, J. P., Gennings, C., Olmo, J. L., Ventura, S., Cios, K. J., and Clore, J. N. (2014). Impact of HbA1c Measurement on Hospital Readmission Rates : Analysis of 70 , 000 Clinical Database Patient Records. *BioMed Research International*, 2014.

Su, H., and Cai, Y. (2016). Firefly algorithm optimized extreme learning machine for hyperspectral image classification. *International Conference on*

Geoinformatics, 2016-Janua(41201341).

Subbulakshmi, C. V., and Deepa, S. N. (2015). Medical dataset classification: A machine learning paradigm integrating particle swarm optimization with extreme learning machine classifier. *Scientific World Journal*, 2015.

Tang, R., Fong, S., Dey, N., Wong, R. K., and Mohammed, S. (2017). Cross entropy method based hybridization of dynamic group optimization algorithm. *Entropy*, 19(10).

Tang, R., and Zhang, X. (2020). CART Decision Tree Combined with Boruta Feature Selection for Medical Data Classification. *2020 5th IEEE International Conference on Big Data Analytics, ICBDA 2020*, pp. 80–84.

Tang, Y., Cui, H., and Wang, Q. (2017). Prediction Model of the Power System Frequency Using a Cross-Entropy Ensemble Algorithm. *Entropy*, 19(10), pp. 552.

Tharwat, A. (2020). Classification assessment methods. *Applied Computing and Informatics*.

Tian, H., Li, S., Wu, T., and Yao, M. (2018). An Extreme Learning Machine Based on Artificial Immune System. *Computational Intelligence and Neuroscience*, 2018, pp. 1–10.

Tian, Z., Ren, Y., and Wang, G. (2019). Short-term wind speed prediction based on improved PSO algorithm optimized EM-ELM. *Energy Sources, Part A: Recovery, Utilization and Environmental Effects*, 41(1), pp. 26–46.

Toprak, A. (2018). Extreme learning machine (ELM)-based classification of benign and malignant cells in breast cancer. *Medical Science Monitor*, 24, pp. 6537–6543.

Tsanas, A., Little, M. A., and Mcsharry, P. E. (2013). A methodology for the analysis of medical data. *Handbook of Systems and Complexity in Health*, 1(1), pp. 113–125.

Vidhya, S., and Kamaraj, V. (2017). Particle swarm optimized extreme learning machine for feature classification in power quality data mining. *Automatika*, 58(4), pp. 487–494.

Wang, M., and Chen, H. (2020). Chaotic multi-swarm whale optimizer boosted support vector machine for medical diagnosis. *Applied Soft Computing Journal*, 88, pp. 105946.

Wang, M., Chen, H., Yang, B., Zhao, X., Hu, L., Cai, Z. N., Huang, H., and Tong, C.

(2017). Toward an optimal kernel extreme learning machine using a chaotic moth-flame optimization strategy with applications in medical diagnoses. *Neurocomputing*, 267, pp. 69–84.

Wang, T., Cao, J., Lai, X., and Chen, B. (2018). Deep Weighted Extreme Learning Machine. *Cognitive Computation*, 10(6), pp. 890–907.

Wang, Y., Cao, F., and Yuan, Y. (2011). A study on effectiveness of extreme learning machine. *Neurocomputing*, 74(16), pp. 2483–2490.

Wang, Y., Wang, A. N., Ai, Q., and Sun, H. J. (2017). An adaptive kernel-based weighted extreme learning machine approach for effective detection of Parkinson’s disease. *Biomedical Signal Processing and Control*, 38, pp. 400–410.

Wu, Y., Zhang, Y., Liu, X., Cai, Z., and Cai, Y. (2018). A multiobjective optimization-based sparse extreme learning machine algorithm. *Neurocomputing*, 317, pp. 88–100.

Xu, Y., Chen, H., Heidari, A. A., Luo, J., Zhang, Q., Zhao, X., and Li, C. (2019). An efficient chaotic mutative moth-flame-inspired optimizer for global optimization tasks. *Expert Systems with Applications*, 129(2019), pp. 135–155.

Yang, H., Yi, J., Zhao, J., and Dong, Z. Y. (2013). Extreme learning machine based genetic algorithm and its application in power system economic dispatch. *Neurocomputing*, 102, pp. 154–162.

Yang, Y., and Duan, Z. (2020). An effective co-evolutionary algorithm based on artificial bee colony and differential evolution for time series predicting optimization. *Complex & Intelligent Systems*, 6(2), pp. 299–308.

Yang, Y., and Pedersen, J. O. (1997). A Comparative Study on Feature Selection in Text Categorization. *Proceeding ICML ’97 Proceedings of the Fourteenth International Conference on Machine Learning*, 53(9), pp. 412–420.

Yang, Y., Wang, Y., and Yuan, X. (2012). *Bidirectional extreme learning machine for regression problem and its learning effectiveness*. 23(9), pp. 1498–1505.

Yassin, N. I. R., Omran, S., El Houby, E. M. F., and Allam, H. (2018). Machine learning techniques for breast cancer computer aided diagnosis using different image modalities: A systematic review. *Computer Methods and Programs in Biomedicine*, 156, pp. 25–45.

Yayık, A., Kutlu, Y., and Altan, G. (2019). Regularized HessELM and Inclined Entropy Measurement for Congestive Heart Failure Prediction. *2014*

International Conference on Electronics and Communication Systems, ICECS 2014, arXiv prep(Jul 2019).

Zhai, J., Shao, Q., and Wang, X. (2016). Architecture Selection of ELM Networks Based on Sensitivity of Hidden Nodes. *Neural Processing Letters*, 44(2), pp. 471–489.

Zhang, R., Lan, Y., Huang, G. Bin, and Xu, Z.-B. (2012). *Universal approximation of extreme learning machine with adaptive growth of hidden nodes*. 23(2), pp. 3937–3942.

Zhang, X., Yang, Z., Cao, F., Cao, J., Wang, M., and Cai, N. (2018). Conditioning Optimization of Extreme Learning Machine by Multitask Beetle Antennae Swarm Algorithm. *Neural and Evolutionary Computing*.

Zhao, J., Wang, Z., and Park, D. S. (2012). Online sequential extreme learning machine with forgetting mechanism. *Neurocomputing*, 87, pp. 79–89.

Zhao, X., Zhang, X., Cai, Z., Tian, X., Wang, X., Huang, Y., Chen, H., and Hu, L. (2018). Chaos enhanced grey wolf optimization wrapped ELM for diagnosis of paraquat-poisoned patients. *Computational Biology and Chemistry*, *November*, pp. 0–1.

Zhou, Z., Chen, J., and Zhu, Z. (2018). Regularization incremental extreme learning machine with random reduced kernel for regression. *Neurocomputing*, 321, pp. 72–81.

ZHOU, Z., and JIAO, B. (2017). Extreme Learning Machine Optimized by Improved Firefly Algorithm. *DEStech Transactions on Computer Science and Engineering*, *aita*, pp. 210–214.

LIST OF PUBLICATIONS

Non-index Journal

1. Alade, O.A., Selamat, A. and Sallehuddin, R., (2020). The Effects of Missing Data Characteristics on the Choice of Imputation Techniques. *Vietnam Journal of Computer Science*, *7*(02), pp.161-177. DOI: 10.1142/S2196888820500098.

Indexed Conference Proceedings

2. Alade, O.A., Sallehuddin, R., Radzi, N.H.M. and Selamat, A., (2020). Missing Data Characteristics and the Choice of Imputation Technique: An Empirical Study. Advances in Intelligent Systems and Computing, *1073*(88-97), DOI: 10.1007/978-3-030-33582-3_9 (**SCOPUS-Index)**.

3. Alade O.A., Sallehuddin R, and Selamat. (2019). Empirical Performance Evaluation of Imputation Techniques using Medical Dataset - IOP Conference Series: Materials Science and Engineering, Volume 551, *Joint Conference on Green Engineering Technology & Applied Computing* 4–5 February 2019. DOI https://doi.org/10.1088/1757-899X/551/1/012055 (**SCOPUS-index)**.

4. Alade, O. A., Selamat, A., and Sallehuddin, R. (2018). A Review of Advances in Extreme Learning Machines Techniques and Its Applications BT - *Recent Trends in Information and Communication Technology*. In F. Saeed, N. Gazem, S. Patnaik, A. S. Saed Balaid, & F. Mohammed (Eds.) (pp. 885–895). Cham: Springer International Publishing (**ISI-INDEX)**.

5. Alade, O.A., Sallehuddin, R., and Radzi, N.H.M. (2021). Performance Evaluation of Extreme Learning Machines Classification Algorithm for Medical Datasets. *2021 International Symposium on Biomedical Engineering and Computational Biology* (BECB 2021). (ISBN: 978-1-4503-8411-7). August 13-15, 2021 (**SCOPUS-index)**. **(**Accepted for Publication).