

Review of Optimization in Improving Extreme Learning Machine

Nilesh Rathod^{1,*}, Sunil Wankhade²

¹Research Scholar, Department of Computer Engineering, RGIT, Mumbai, India

²Professor, Department of Information Technology, RGIT, Mumbai, India

Abstract

Now a days Extreme Learning Machine has gained a lot of interest because of its noteworthy qualities over single hidden-layer feedforward neural networks and the kernel functions. Even if ELM has many advantages, it has some potential shortcomings such as performance sensitivity to the underlying state of the hidden neurons, input weights and the choice of functions of activation. To overcome the limitations of traditional ELM, analysts have devised numerical methods to optimise specific parts of ELM in order to enhance ELM performance for a variety of complicated difficulties and applications. Hence through this study, we intend to study the different algorithms developed for optimizing the ELM to enhance its performance in the aspects of survey criteria such as datasets, algorithm, objectives, training time, accuracy, error rate and the hidden neurons. This study will help other researchers to find out the research issues that lowering the performance of the ELM.

Keywords: Extreme learning machine (ELM), Single-feedforward neural networks; Kernel functions, Sensitivity, Input weights and Activation bias

Received on 24 July 2021, accepted on 09 September 2021, published on 17 September 2021

Copyright © 2021 Nilesh Rathod *et al.*, licensed to EAI. This is an open access article distributed under the terms of the [Creative Commons Attribution license](#), which permits unlimited use, distribution and reproduction in any medium so long as the original work is properly cited.

doi: 10.4108/eai.17-9-2021.170960

1. Introduction

This Extreme learning machine (ELM) [1] is considered as the fundamental and efficient for training the single hidden layer feed-forward network. The learning parameters are picked at random from the hidden layer's network design in the ELM, so that the total number of neurons in the hidden layer network can be placed. The least square approach is employed to generate the output weights in the hidden layer in such a way that iterations are not needed for the network offset and weight in this progression. ELM algorithm has the strong generalization ability and effective training speed when compared with the several traditional algorithms based on neural networks [2]. This distinctive feature of effective training generally results from two aspects: i) the weights of the output can be analytically generated instead of frequently utilizing the chain rule for the learning parameters based on

the partial derivatives. ii) the parameters of the latent layers are randomly generated without fine-tuning before generating the training data [3]. In [4], the output weights resolved from the least squares problem can prolong the widespread approximation capability of connectionist model and the weights of the stochastic latent layer. ELM achieved good generalization capability with respect to the faster training speed compared with the support vector machine (SVM) and back-propagation-based neural networks [5]. ELM is an effectual learning algorithm in the single-hidden layer feed forward neural networks. Traditional ELM and SVM provides equivalent significance for all the samples, thereby outcoming in a biased outcome regarding the majority class. The larger class imbalance problems can be effectively resolved by developing variants of Boosting Weighted ELM (BWELM), ELM-like Weighted ELM (WELM) etc. [6].

Several real-life application issues can be considered as an inconvenience based on the imbalanced classifications where the total sample in a particular class is completely exceeded

*Corresponding author. Email: Nilesh.rathod@mctrgit.ac.in

than the total samples in several other classes. The samples with the lesser and the higher-class proportions are considered as the minority and the majority classes. The weights are initialized by the ELM between the hidden layer and the input in a random fashion, and then the randomly assigned weights are utilized for mapping the data with the input based on the feature space. Consequently, these initialized weights remain unaffected during the process of training resulting in some misclassifications in the samples [7]. The algorithmic level techniques, cost sensitive approaches, and data level strategies can all be categorised based on imbalanced issues. The data level techniques, like under sampling and the oversampling [8] [9] vary the data space for reducing the class imbalance effects. The under-sampling approach selects the fraction of data from the majority class examples in a random fashion, and then the loss of information regarding the cost in the data distributions is balanced. For example, Balance Cascade [9] and the Easy Ensemble algorithms utilize the under-sampling methods for the dataset balancing. The method of oversampling randomly makes an exact copy of the minority class for maximizing the minority class cardinality, which may result in over fitting issues. On the contrary, the oversampling method obtains synthetic minority class example for balancing the class distribution, like technique of Synthetic Minority Over-Sampling (SMOTE) [10]. The techniques based on the algorithmic level [11,12] altered the classifier model to solve the imbalanced problems and techniques based on cost-sensitive functions allocates higher penalty for the misclassifications generated in the minority class in comparison with the examples based on majority class. In addition, cost is expensive as the minority class examples produce misclassifications [13,14].

The ELM approach proposed in [15] was considered as the straightforward and effectual single hidden layer feed-forward neural network algorithm, and no iterative adjustments are needed during the training of algorithm with respect to the network weight and bias. In comparison with more traditional neural network algorithm, faster training speed was achieved by the ELM. In [16], a novel approach called the ELM-AE classification algorithm regenerated the auto encoder and the input signals. Based on this approach, a kernel extreme learning machine auto encoder (KELM-AE) was proposed as a multi-label learning algorithm. In this algorithm, KELM module with two-layer was employed as the support model in such a way that the first module was considered as the auto encoder block, and then the information regarding the label nodes were added to the layer of input. In addition, the output layer achieved some output features, which consists of the label's relationships and feature. The second module was considered as the classification module and was employed in the process of classification whereas the non-equilibrium-based label completion matrix algorithm was utilized by the label space. However, the trivial design of OC-ELMs may limit their potential with respect to the high and complex dimensional learning of the datasets, thereby resulting in a sub-optimal performance of generalization. On the other hand, many developments have been noticed and witnessed regarding the

deep networks in the ancient times. The multilayer neural network-based ELMs (ML-ELM) examined the representative achievements, like deep weighted ELM [17], sparse representation-based hierarchical ML-ELM (H-ELM) [18] stacked AEs based ML-ELM [16] and the kernel-based MK-ELM [19] etc., which may result in further enhancement to OC-ELMs [20].

This paper intends to review a total perspective of ELM for different application. This review introduced qualities and problem with the goal of finding research gap. Additionally, this review provide how optimisation will affect extreme learning machine. The following is how the paper is structured: Section 2 discusses several optimization strategies that have been used to extreme learning machines, as well as how these algorithms affect the performance of the ELM methodology on extreme learning machines and as a research tool. Section 3 Comparative analysis develops a comparative table based on the procedures outlined before with respect to the algorithm applied, the parameter being optimised, application. In Section 4, Conclusion, future scope and Summary of the paper.

2. Literature Review

This section evaluates the different optimization techniques employed to extreme learning machines and how these algorithms are affecting the overall performance of ELM.

2.1. Genetic Algorithms

Author Xue et al., introduced a unique algorithm for learning referred to as genetic ensemble of ELM [21]. This approach adopted GA for producing the set of ELM candidates with idea; parameters for decrementing the negative effect of parameters that are not ideal. Then a portion of the ELM candidates are chosen for assembling a new system based on the strategy of sorting. The GE-ELM's execution has been enough surveyed by nine benchmark datasets for both classification and regression issues. GE-ELM created better performance of speculation when differentiated with other ELM strategies with respect to robustness. From the results in both classification and regression issues, it very well may be assumed that GE-ELM can make the networks progressively strong and produce better performance of generalization. Matias et al., also proposed an algorithm for learning referred to as optimized ELM (O-ELM) [22]. The SLFN parameters were determined using an optimization process in this framework. The least square methodology was employed to estimate the output weights; however, Tikhonov's regularisation was applied to improve the SLFN's performance when noisy data was available. The proposed system has been endeavoured with three methodologies of improvement (differential evolution, simulated annealing, genetic algorithm) using 16 benchmark issues accessible in public repositories.

Alexandre et al., specifically discussed the plausibility of a novel search strategy; hybridized with an ELM, allowing features concentrated on the special GA class of minimal searching [23]. This method GA-ELM was considered to be the most appropriate for the present application after various examinations and connection with the presentation of linear regressions of fast learning. The proposed GA-ELM highlight option assessment provides fantastic outcomes as to the probability of proper portrayal. The proposed methodology lets the classifier based on ELM improve its order efficiency from 74.83 percent (without selecting highlights) to 93.74 percent (while using the ideal setting of highlights selected). Alencar et al., proposed a novel procedure to prune neurons in the hidden layers depending on GA [24].

The GAP-ELM performance was assessed on 7 genuine datasets and contrasted with OP-ELM, MLP, RBF and ELM. All techniques were contrasted agreeing with three criteria: time required for training, number neurons in the hidden layer and accuracy [25]. Based on the results the authors concluded that GAP-ELM is a substantial option for the tasks of classification. Avci, & Dogantekin, proposed a diagnosis FRAMEWORK for Parkinson disease by applying GA, wavelet kernel- (WK) and ELM [26]. A SLNN (single layer neural network) classifier was utilized and prepared by the learning technique of ELM. The datasets for the disease are obtained from the machine learning database of UCI. The structure of WK-ELM includes of three parameters of WK that can be adjusted. These parameters along with the hidden neuron number played a significant role in the ELM's performance. GA was helpful in determining the hidden neuron count and optimising parameter values [27-28]. The developed GA-WK-ELM was assessed based on the performance parameters of specificity, sensitivity, accuracy, and ROC curves. The developed model gave the highest classification accuracy of 96.81%.

In the above reference papers, in the experimentation inputs have been normalized in the range [0,1] and output normalization [1,1], they set the population size as per the requirement and have taken enough iteration to search the optimal solution. The validation is randomly selected from the training dataset.

Table 1. Accuracy of the suggested algorithm according to datasets [21-28]

California housing	GE- ELM	8	194.3135	0.1348	Not Applicable
Vehicle	GE-ELM	18	63.7221	85.262	Not Applicable
Diabetes	GE-ELM	8	26.4125	84.574	Not Applicable
Image Seg	GE-ELM	19	243.4382	95.869	NA
Satellite image	GE-ELM	36	769.6452	88.256	NA
Letter	GE-ELM	16	1906.7254	81.125	NA
Automobile MPG	GO- ELM	6	4.3751	0.2608	21.75
Cancer	GO- ELM	32	4.0442	0.5782	18.25
Servo	GO- ELM	4	3.6212	0.2320	21.75
Price	GO- ELM	15	3.7553	0.1896	18.70
CPU	GO- ELM	6	3.7444	0.1733	17.20
Concrete Comp. Strength	GO- ELM	8	5.4051	0.2738	22.05
Ailerons	GO- ELM	40	27.4410	0.0912	23.55
Pumadyn	GO- ELM	32	17.9558	0.1442	22.95
Elevators	GO- ELM	18	30.3114	0.0774	24.05
Pole Telecommunication	GO- ELM	48	22.0324	0.4901	25.25
Bank	GO- ELM	32	16.2897	0.2051	23.40
Computer Activity	GO- ELM	21	17.9445	0.0848	23.75
Stock	GO- ELM	9	9.5224	0.6767	23.85
Communities and Crime	GO- ELM	127	9.6412	0.2679	22.10
Pyrimidines	GO- ELM	27	3.8585	0.2372	20.60
Adult	OP-ELM	NA	NA	81.8	31.7
Banana	OP-ELM	NA	NA	99.7	36.7
Breast cancer Wisconsin	OP-ELM	NA	NA	96.4	12.8
Haberman	OP-ELM	NA	NA	75.3	8.5
HIV	OP-ELM	NA	NA	86.6	11.5
Pima Indian diabetes	OP-ELM	NA	NA	76.2	14.3
Ripley	OP-ELM	NA	NA	91.1	27.5
Parkinson dataset	GA-WK-ELM	22	0.21 μ s	96.81	260
Vehicle	GA-ELM	13	NA	93.74	NA

Datasets	Algorithm	No. of attributes	Training time	Testing accuracy
Sinc	GE- ELM	1.	90.3558	0.0051
Abalone	GE- ELM	7	62.1462	0.0718
Boston housing	GE- ELM, GO-ELM	13	24.8512, 4.5463	0.0743, 0.3503

Table 1 presents the algorithm accuracy of the suggested techniques in the aspect of datasets. It discusses the role of genetic algorithm in the ELM used in two stages viz, feature selection and classification. The common parameters investigated are the testing accuracy, training time, the hidden neurons and error rate. The observations prepared from above table are:

- Lack of data generalization
- Least emphasis on multi-objective formulation
- Application of hidden neurons.

These three criteria shall be addressed, for better performance measures like RMSE, error rate and accuracy of the ELM via Genetic algorithm.

2.2. Particle Swarm Optimization (PSO)

PSO has been utilized progressively as a viable procedure for searching global minima. PSO has no complex transformative operators and smaller number of parameters needed for modification. Like this, PSO hybrids and ELM ought to guarantee the training FF neural networks [29]. To acquire a compact network architecture with better performance in terms of generalization Zhao et al., developed an enhanced ELM with the growth of hidden nodes being adaptive (AG-ELM) along with PSO [30]. The PSO is utilized in selecting the optimal biases and weights for overcoming the lack of the standard AG-ELM. One particle in the PSO represents the parameters of a single network and the particle dimension is increased during the process of training. The results obtained from simulations on different test issues confirm that the developed algorithm accomplishes more compact architecture of the network than conventional AG-ELM. Zhang, & Yuan improved the accuracy of diagnosing the fault in power transformers by developing a kernel-based ELM with PSO to optimize the KELM parameters [31].

After the optimization of the parameters the optimized KELM is actualized for classifying the faults in the power transformers. Nine benchmark datasets to test the developed technique. The developed fault diagnosis system was compared with the two more ELMs the BPNN and SVM. The results proved that the developed method is more stable, had a faster speed of learning and achieved an enhanced generalization performance.

ELM-PSO methodology was proposed by Yadav et al., for designing an in-situ bioremediation model for the groundwater that is debased with the compounds of BTEX [32]. The outcomes of this investigation are compared with existing studies, indicating the better effectiveness of the created model. The ELM-PSO is a potential apparatus to structure an ideal in-situ bioremediation framework as it lessens the computational time as well as the framework cost while corrupting the concentration of BTEX to a predefined level of clean-up.

Here in these experiments, the inputs and outputs range have been normalized in the range of $[-1, 1]$ and $[0, 1]$, respectively. For the trial of simulations, the data set was divided into training and testing data sets. The sigmoid function is selected as the activation function. The hidden node parameters are randomly chosen from the range of $[-1, 1]$. In the experiments, some parameters including the population size, maximum iteration number,

Table 2. Algorithm accuracy of the suggested techniques in the aspect of datasets [29-32]

Datasets	Algorithm	No. of attributes	Training time	Testing accuracy
Diabetes	IPSO-ELM,	8	12.1810;	76.40;
	KELM		732.05	83.85
Satellite image	IPSO-ELM	36	3.105e + 3 s	87.30
Image segmentation	IPSO-ELM	19	1.2731e + 3 s	94.62
Abalone	AG-PSO- ELM	8	NA	0.1152
Boston housing	AG-PSO- ELM	13	NA	0.1947
California housing	AG-PSO- ELM	8	NA	0.2189
Census	AG-PSO- ELM	8	NA	0.1041
Delta ailerons	AG-PSO- ELM	5	NA	0.0772
Delta elevators	AG-PSO- ELM	6	NA	0.0909
Machine CPU	AG-PSO- ELM	6	NA	0.0782
Balance Scale	KELM	4	297.19	91.97
Glass	KELM	9	105.53	73.33
Iris	KELM	4	65.75	100
Wine	KELM	13	103.5	100
Liver	KELM	6	132.81	75.44
Vowels	KELM	10	766.94	96.97
Waveform	KELM	21	31123	86.49
Breast cancer	KELM	30	541.52	97.01

The table 2 presents the classification accuracy of the PSO in ELM in the aspect of datasets. As we could observe different forms of PSO-ELM, which is deployed to overcome the issues of classification such as class imbalance, adjusting the weighting parameters, and so on. [33,34,35] Quality improvement of the classifiers is the main objectives of the optimization models. The observations made from the above table are:

- Improper use of hidden layers.
- Optimization in improving the feature selection stage is not focussed.
- Training time is the most important factor, which is least concentrated

2.3. Evolutionary Algorithms

SAE (self-adaptive evolutionary) agents incorporating logic are utilized for managing the elements contributing to the ELM (extreme machine learning) processing [36]. For better performance of the ELM's by using SAE algorithm, the regular soil temperature evaluation is done at different depths of 5, 20, 10, 50, 30, and 100 cm. Optimization of hidden parameters of network nodes is done by employing the SAE algorithm in the designed SAE-ELM model. Different datasets like the air temperature measurements which include minimum, maximum, and average values, solar radiation from the Kerman and Abbas stations in Iran, and atmospheric pressure are employed in the study.

To deliver the developed plant products to clients through storage depots, the two-stage capacitated facility location problem is discussed [37]. The restricted depot storage capacity and plant production are brought into notice relating to the client demands. The aim of this study is to reduce the cost of finding plant and depot locations. The cost includes

both, fixed and transportation costs. In a suitable time -period of computation, more optimal solutions are built with the hybrid evolutionary algorithm framework by approximating the machine learning fitness. Genetic parameters will be used to carry the searching process out and local search strategy will be utilized to refine the best solution in the population [38,39].

Mixing optimization techniques of differential evolution feature selection with extreme learning machine results in the new algorithm for subset selection containing features. The developed feature selection model along with feature subset selection method shows good performance in producing feature sets of optimal number while having higher rates of classification when compared to other techniques. The simulation results point to the same [40,41].

For assessing the performance through experiment, a well-defined set of instances developed the benchmark instances consist of five classes of instances by varying seven parameters. For each instance, the values of the seven parameters are randomly generated.

Table 3. Algorithm accuracy of the suggested techniques in the aspect of datasets [36-41]

Datasets	Algorithm	No. of attributes	Training time	Testing accuracy
Bandar Abbas	SaE-ELM	11	NA	0.8660–1.5338 °C
Kerman	SaE-ELM	11	NA	1.5818 to 2.3422 °C
AVIRIS - Hyperspectral images	DE - ELM	3	14-18s	98.07
KSC – Hyperspectral images	DE - ELM	3	13-15 s	89.79
Bike sharing dataset	Differential Evolution- Metaheuristic Coral reef Optimization (DE-MCRO)	16	0.525	78.595
Concrete compressive strength data set	Differential Evolution- Metaheuristic Coral reef Optimization (DE-MCRO)	9	0.895	27.586
Housing data set	Differential Evolution- Metaheuristic Coral reef Optimization	14	0.250	43.56

Plant dataset	on (DE-MCRO) Hybrid Evolutionary algorithms	7	492.78 s	0.94
River ice images	ELM- Differential Evolution Feature Selection (ELM-DEFS)	7	NA	97.78

The table 3 presents the algorithm accuracy of the suggested techniques in the aspect of datasets. The observations made from the above table are:

- Randomization of hidden layers is not permitted that helps to enhance the reliability of the classifiers.
- Decision variable is dependent on the set of control parameters.
- Some control parameters are tuned with different kernel function, so to decrease the searching complexity.
- Study in optimizing feature selection and the hidden layers are not explored.

2.4. Artificial Bee Colony Optimization

To optimize the ELM parameters which are based on ABC (ant bee colony) optimization method, a new hybrid methodology is proposed. The methodology uses ABC method to specify optimal input weights with biases and MP (Moore-Penrose) inverse to determine the output which are the weights. Original ELM with other evolutionary ELM methods are analysed with the build ABC-ELM algorithm to compare the differences in various datasets of classification [42].

Three significant differences between the improved technique based on the ABC optimization, PS-ABCII [39], and existing techniques include mechanism modification where employee bees become scouts, greedy selection mechanism being ignored, and the population initialization followed by the application of opposition-based learning. To discuss the optimal performance of the PS-ABCII in comparison with existing techniques, ten social functions are highly regarded. Besides these, a hybrid model combining PS-ABCII with ELM is proposed by the authors and is called PS-ABCII-ELM model. Application of PS-ABCII to the ELM results in the increased generalized ELM performance for weight tuning of inputs and ELM's biases in PS-ABCII-ELM model [43].

New method is devised which incorporates building projections of ELM on the fed data to a high dimensionality feature space, followed by ABC algorithm executing unsupervised clustering. Despite facilitation of clusters by ELM projections, the dependence issues faced on the clusters' initializing centres are bested by a metaheuristic technique which includes the ABC algorithm and following local minima convergence where losses are faced because of standard algorithms, including the K-means algorithm [44].

For problems related to optimization, Wang et al. proposed a new ABC algorithm which is known as NABC [45]. In this algorithm, the population quality is improved first by the utilization of chaotic opposition-based learning techniques in NAC first. Further, the ability of convergence is improved by the self-adaptive search method. Chaotic local search technique allows the scout bees to escape the local maxima. Hence, both, the ELM biases, and the input weights, are enhanced by NABC and the ELM's performance of generalization is improved overall [46].

The experimental set up of these algorithms were carried out on the platform of MATLAB 7.0. Here the number of the iterations and population size were set. The number of the employed bees is equal to the on lookers, can be 20. The input weights and the biases are obtained in the range [-1, 1].

Table 4. Algorithm accuracy of the suggested techniques in the aspect of datasets [42-46]

Datasets	Algorithm	No. of attributes	Training time	Testing accuracy
Bupa Liver	Artificial Bee	6	NA.	72.23.
	Colony-ELM(ABC-ELM); NABC-ELM		11.184	0.7322
WDBC	Artificial Bee	8	NA	96.42
	Colony-ELM(ABC-ELM)			
Wine	Artificial Bee	13	NA	98.43.
	Colony-ELM(ABC-ELM); NABC-ELM			0.9895
Australian	Artificial Bee	14	NA	86.38
	Colony-ELM(ABC-ELM)			
Breast cancer	Artificial Bee	9	NA	96.51
	Colony-ELM(ABC-ELM)			
Thyroid	Artificial Bee	5	NA.	94.79.
	Colony-ELM(ABC-ELM); Novel ABC-ELM		4.411	0.964
Cleveland and Heart	Artificial Bee	13	NA	83.22
	Colony-ELM(ABC-ELM)			
Parkinson	Artificial Bee	22	NA	88.31;
	Colony-		11.221	0.9361

	ELM(ABC-ELM); NABC-ELM			
Boiler thermal datasets	PS-ABCII-ELM	10	919.45	0.99328
Balance	ABC-ELM	4	24.61	68.72
Cancer-Diagnostic	ABC-ELM	30	17.98	94.2
Cancer-Original	ABC-ELM	9	20.35	96.49
Cardiotocography-	ABC-ELM	21	19.62	77.85
Cardiotocography-CNAE	ABC-ELM	21	30.1	49.3
	ABC-ELM	856	0.11	55.78
Dermatology	ABC-ELM	34	19.17	88.04
Glass	ABC-ELM	9	14.3	59.35
Iris	ABC-ELM; NABC-LEM	4	9.57;	97.33.
			2.879	0.9769
LIBRAS	ABC-ELM	90	11.25	49.88
Spam	ABC-ELM	57	11.79	80.9
USPST	ABC-ELM	256	5.28	74.06
Diabetes	NABC-ELM	NA	14.996	0.7619
	NABC-ELM	NA	35.662	0.8773
Image segmentation	NABC-ELM	NA	6.982	0.8564
Hepatitis	NABC-ELM	NA		

The table 4 presents the algorithm accuracy of the suggested techniques in the aspect of datasets. The observations made from the above table are:

- Learning speed rate determines the classification accuracy
- Randomness in determining hidden neurons will tend to lower the classifier accuracy.
- Convergence speed also determines the better searching ability.

2.3. Other Algorithms

MBAS (Multitask Beetle Antennae Swarm Algorithm), a new innovative molecule swarm heuristic calculation algorithm, supported by ABC algorithm along with the BAS (Better Antennae Search) algorithm is proposed. Employing MBAS allows the input weights and ELM biases to be optimized further [47]. This method minimized the regression error and condition number and obtains good generalization performance.

A novel Kernel Extreme Learning Machine (KELM) parameter tuning method employs GWO (Grey Wolf Optimization), a new swarm intelligence algorithm [48]. Owing to its ability of simulating a structure's social levels and the ability of hunting present a grey wolf, GWO is used in developing KELM model to predict bankruptcy and has been found out to be extremely efficient. Further, a comparison is made between GWO and existing KELM methods including support vector machines, ELM, random forests, improved ELM. Genetic KELM, grid search KELM, and particle swarm optimization based KELM and is operated on two different datasets in real-time through cross-validation evaluation.

Improved GWO integrates KELM in order to deal with medical diagnosis and is known as IGWO-KELM. The IGWO-KELM proposes a novel framework for prediction

[49]. IGWO-KELM approach is employed in cases where the optimal dataset consisting of medicinal features is to be identified. This approach works in a way that the GA first adopts the diversity of the first position, which is then followed by the present positions of the population being updated by the GWO to discrete searching in the allotted space which, through the optimal existence of feature sub setting, provides the most optimal classification [50].

An alternative learning path for KELM which has been devised utilized the CMFO (Chaotic Moth-Flame Optimization) [51]. Optimization of parameters is carried out by CMFO along with feature selection simultaneously. A detailed comparison of the model put forth in this research, with the KELM models that already exist, is based moth-flame optimization, genetic algorithms, and particle swarm optimization. Various problems in the medical field like diagnosis and detection of Parkinson's disease along with breast cancer support the comparison. The learning method developed here has found its application in medical diagnoses.

A hybrid DA (Dragonfly Algorithm) is put forth with ELM system for predicting the problems. In data regression with classification problems, the model with ELM shows practical results [52,53]. Even after the very fast training efficiency, the model's hidden layers require many nodes. The evaluation time of ELM, thus, leverages the large number of hidden layer nodes in its utilization. Furthermore, with respect to the understanding of optimal weights and settings for the hidden layer biases, no guarantee can be undertaken. The acceleratory motion of the moths is captured in its essence by the DA. Exploitation of the algorithm's hidden layer characteristics helps in choosing the small number of nodes. This results in the optimal selection of weights of the suitable hidden layer and the most efficient biases. A set of assessment indicators supports the analysis of the proposed method, and a comparison on the repository's ten regression datasets is carried out.

A roadblock with customary learning measurement is the high-dimensional knowledge investigation. It appears that the development of improved ELM algorithms will be intriguing. One such fascinating example is the familiarization of ELM with inadequate strategies of coding to manage higher dimensional information effectively. From a hypothetical point of view, the defending of ELM arbitrary system and contemplation of relationships between different algorithms and ELM. In ELM, the main consideration is the irregular aspect of preparation, which makes the preparation extremely competent and guarantees an all-inclusive ability of estimation. Often, over-fitting is eliminated, and better predictions are exhibited due to the irregular aspect preparation. This irregular component and its viability require additional study, analysis, and instructions. Examining the relationship between other algorithms that also receive arbitrary systems and ELM would be intriguing. Some examples of relevant algorithms for counting arbitrary tests are Arbitrary Woodland and Ad boost.

From an all-inclusive perspective, the exploration of the dispersion structure's influence on the hidden boundary layer development shows that any continuous diffusion without the

general estimation, power being given up, can generate the layer's hidden boundary. In genuine applications, the highly speculative exhibitions may be affected by the widely undertaken work on the boundary creation. While it supports further study about the issue and also about the information subordinate speculation mistakes destined for ELM. Even though the calculation learning productivity and speculation execution of ELM have been understood, the application is not exceptionally helpful and seems to be influencing the precision of the outcomes due to the high number of neurons in the shrouded layer. This suggests that continual development of ELM model and the structure and execution speculation of the associated algorithms is necessary. Additionally, combination of competent methods, such as RBF component, along with ABC algorithms is required and comparison between the ABC algorithms in ELM and Piece Highlight Space also has scope. Lately, several scientists have undertaken the task of consolidating ELM with various control algorithms, solving problems with ELM, and introduction of better preparation models.

Table 5. Algorithm accuracy of the suggested techniques in the aspect of datasets [47-53]

Datasets	Algorithm	No. of attributes	Training time	Testing accuracy
Wie slow dataset	Grey Wolf Optimization with ELM	30	Not applicable	84.58
Japanese	Grey Wolf Optimization with ELM	30	Not applicable	86.57
Parkinson	Enhanced Grey Wolf Optimization with ELM; CMFOFS-KELM	22	Not applicable	97.45; 96.87
Breast cancer	Chaotic Moth Flame Optimization Feature Subset – Kernel ELM (CMFOFS-KELM)	9	Not applicable	96.93
CASP	Dragonfly Optimization (DFO-ELM)	9	117.361	0.102
CBM	Dragonfly Optimization (DFO-ELM)	17	71.238	0.000
CCPP	Dragonfly Optimization (DFO-ELM)	4	69.183	0.968
ENB2012_Y1	Dragonfly Optimization (DFO-ELM)	8	34.604	0.556
ENB2012_Y2	Dragonfly Optimization (DFO-ELM)	8	35.692	0.274
Forest Fire	Dragonfly Optimization (DFO-ELM)	12	32.903	18.095

Housing	Dragonfly Optimization (DFO-ELM)	13	34.417	0.436
Relation Network	Dragonfly Optimization (DFO-ELM)	22	337.764	0.009
Slump test	Dragonfly Optimization (DFO-ELM)	10	33.744	1.173
Yacht hydrodynamics	Dragonfly Optimization (DFO-ELM)	6	61.284	0.943

- a) grey wolf optimization, Dragonfly & Coral reef optimization are still in developmental stages.
- b) Lack of studies in the hidden layers.
- c) Higher dimensionalities and mixed quality of features are difficult task

3. Comparative Analysis

Finally, a comparative table is developed based on the techniques discussed in this paper with respect to the algorithm applied, the parameter being optimised, application etc.

From the above table 5, the observations are:

Table 6. Comparative Analysis

<i>Sl No</i>	<i>Reference No</i>	<i>Types of algorithms</i>	<i>Parameters considered</i>	<i>Applications</i>	<i>Limitations/ Research Gap</i>
1	18	A kernel based extreme learning machine (KELM) with particle swarm optimization (PSO)	The vector of the output weights between the hidden layer of L nodes $\beta = [\beta_1, \dots, \beta_L]^T$ The feature mapping= h(x)	It provides better performance and takes less time along with providing better and stable performance in classifying faults to support power transformers.	No practical application has been designed so far.
2	21	Genetic ensemble of extreme learning machine	Population size = 200. No of hidden nodes= [5,200] Benchmark datasets=5	It can be used to make robust and efficient networks.	The testing accuracy of the GE-ELM is lesser than that of E-ELM.
3	22	The grouping genetic algorithm (GGA)	Predicting variables=92	It can be used to perform exact prediction for solar radiation.	The prediction performance is higher only when previous hour data are considered with the given input.
4	23	Genetic Algorithm hybridized with a Extreme Learning Machine	Population=100 individuals Selection function=Stochastic uniform Crossover function=Multi-point stochastic Probability of crossover=0.8 Probability of mutation=0.05 Maximum number of generations=50 Maximum number of stalled generations=20	It can be used for roads having several directions in every lane.	Electrical vehicles have not been considered during the study.
5	24	Optimally Pruned ELM (OP-ELM)	Gene vector=[1 1 1 1] Genetic material= [1 0 1 1]	It can support different kinds of real-time datasets.	The ELM pruning technique should be designed with different criteria of hidden layers.
6	25	Genetic algorithm- (GA-) wavelet kernel- (WK-) Extreme Learning Machines (ELM)	Population=40 UCI dataset=128	Support all types of Parkinson disease prediction process.	This technique is required to be tested for real time data.
7	26	A novel hybrid approach based on artificial bee colony (ABC) optimization	Iteration size=250 Population size=40 No of onlookers=20 Initial inertia weight=0.9 Final inertia weight=0.4	It can be used for solving various domain problems including the classification	Simultaneously determination of the optimal number of hidden neurons and bias and weighted

			Step size range=5 to 40	problems.	parameters.
8	27	Markovian jump Hopfield neural networks with discrete and distributed delays.	The neuron activation function vector= R^n A given finite set $S = \{1, 2, \dots, N\}$	The proposed method is less conservative as well as less computationally demanding.	The delay-dependent analysis along with synthesis carried out on discrete stochastic neural networks consisting of delays.
9	29	The improved extreme learning machine (IPSO-ELM)	Input range= [-1,1] Population size=200 Maximum optimization time period=20 Initial inertia weight=1.2 Final inertia weight=0.4	It can find application in data mining and gene data expression.	This algorithm is yet to be tested for resolving complicated problems of data mining.
10	30	An upgraded ELM with adaptive growth of hidden nodes (AG-ELM)	Input range= [-1,1] Output range= [0,1] Population size=30 Iteration size=50 Initial inertia weight=0.9 Final inertia weight=0.4 Maximum step size=20 Pop size=30	It provides high rate of accuracy for complex networks.	The developed PSO-AG-ELM spent more time than the other ELMs which is a problem to be resolved.
11	31	APSO-KELM	Input range= [-1,1]	It is applied in cancer classification.	The APSO-KELM is complex and hence requires optimization and GPU also can be incorporated.
12	32	IGWO-KELM	Number of iterations=8 Size of the population=100	It provides better quality of solutions.	Minimum number of selected features are obtained.
13	33	Error minimized extreme learning machine (EM-ELM)	Input range= [-1,2] Output range= [0,1] Population size=30 Maximum iteration number=10 Tolerance rate=0.02 Initial inertia weight=0.9 Final inertia weight=0.4	It solves complex classifications.	The developed technique took more time in comparison with the ELM.
14	34	SDPSO-ELM	the hidden nodes $L = 30$, for the other three algorithms, the hidden nodes $L = 10$, the particle number $N = 20$, the tolerance rate for fitness function $\varepsilon = 0.1$, the maximum iteration number is set to 200	It improves the forecasting accuracy as compared to RBFNN algorithm.	No work has been carried out on the latest control strategies to enhance the ELM algorithm's performance along with some enhanced intelligent methods.
15	38	DE-ELM	Parameter interval= [0.4,1]	It can be used to classification hyper spectral images.	Other evolutionary computation techniques can be investigated as an alternative to the DE-ELM method.
16	39	Differential equation coral reef optimization (DECRO)	Training set size=12,911 Test set size=3478 Number of attributes=16	It predicts flawlessly the precision.	Implementation of simultaneous parameter optimization and feature selection
17	41	A hybrid evolutionary algorithm (HEA) with genetic operations and local search	The n-dimensional space= R^n Input vector= x .	It provides better results for TSCFLP.	No work is done in complex situations to solve combinatorial problems of

					optimization.
18	43	An ELM LOO bound, MOCLPSO	Population size=200 Crossover function=0.8	It provides high learning speeds for applications.	The performance of the algorithm for huge datasets is yet to be evaluated.
19	44	Multitask Beetle Antennae Swarm Algorithm	Output range= [0,1]	It reduces the computation cost while achieving the best results for the datasets.	The performance of the algorithm for huge datasets is yet to be evaluated.
20	49	Grey Wolf Optimization	No of iterations=250	It helps to predict bankruptcy	Implementation of simultaneous parameter optimization and feature selection for further improvement of KELM performance.
21	51	Chaotic moth-flame optimization strategy (CMFO)	Sigmoid activation function=50	It supports medical decision making in the domain of medical diagnosis.	Implementation of simultaneous parameter optimization and feature selection
22	52	Hybrid dragonfly algorithm	No of dragonflies =7 No of iterations=100 Inertia factor=0.1 Individual-best factor=0.1 Crossover Fraction=0.8 Migration Fraction=0.2 Input nodes=7 Hidden nodes=100 Activation fun=sigmoid Output nodes=1 No. of Iterations for ELM=1000	It can be used to optimize ELM model.	The performance of the algorithm with optimizers other than GA and PSO is yet to be carried out.
23	54	Competitive Swarm Optimizer (CSO)	Noise range=[-2,2]	It reduces the size of the network.	
24	56	Coral Reefs Optimization – Extreme Learning Machine (CRO–ELM) algorithm	Inputs weights = w_i and bias = b_i The number of neurons in the hidden layer=20	It depicts the real test data against the best CRO–ELM prediction.	The regular averaging o the input predictive variables needs an extra variable to adjust the system to model the solar cycle.

From the above analysis, the research issues and findings pertaining to this field are collectively presented as follows:

Kernel Extreme Learning Machine Auto Encoder algorithm is developed for improving the accuracy of classification. However, the challenge lies in using the effective information contained in the output space and combining these algorithms to construct a unified multi-label learning framework for better performance results. Hybrid machine learning model is devised for improving the global and local search capabilities. However, this method failed to solve multi-objective optimization problems and also failed to deploy this method to a high-performance computing platform in order to reduce the computational burden. Under Bagging ensemble method is designed to address the class imbalance problems, but the challenge lies in evaluating other types of ensemble methods to address the large imbalance problems more accurate and effectively. The major problem in ELM is that

it may utilize huge number of hidden neurons as compared to classical gradient descent technique. In addition, the arbitrary initialization of input weights and hidden biases may affect the ELM performance. ELM training is an optimization issue that offers an opportunity to solve various issues and offer improved generalization performance. However, the accuracy, training speed, stability and network complexity are the major issues faced by ELM in classification.

Machine learning techniques are devised for supporting different tasks that are related to make decision. Thus, intelligent classifiers are utilized for improving ELM and classification accuracy of the data. Traditional neural network-based classifiers have been updated, but they are prone to overfitting and local optimum concerns, and they have remained a hot topic for performance enhancement using various ensemble methods. However, recently ELM has gained fame for solving classification problems. Unlike the earlier classic gradient-based learning algorithms,

which only work for differentiable activation functions and are prone to issues like local optimum, inappropriate learning rate, overfitting, and so on, as a result, ELM can handle non-differentiable activation functions and tends to arrive at a solution quickly and without such modest difficulties. Thus, the motive is to devise novel method for classifying data, and to improve the ELM. The goal is to devise improved learning-based model that are capable to reveal its reasoning to deal with input case and offer reasoning for prediction. ELM needs improvement, it can be done through optimization [57]. Through this study we can say that need of real time optimization [58] to improve ELM.

4. Conclusion

During this study, ELM applications and implementation has been reviewed. There is a lot of research work done about the research topic through various interesting algorithms and the observatory results has proved the efficiency, accuracy, and easy implementation in various fields through ELM and its variants. ELM has shown great advantages when compared with the other state-of-the art learning algorithms such as SVMs and algorithms of deep learning.

Despite many advantages of Extreme learning machine to its credit, there exist some problems with state to researchers. it's been ascertained that the classification boundary for learning parameters of hidden layer. ELM cannot handle massive high dimensional information, it needs additional hidden nodes than conventionally standardisation algorithms, and cannot be parallelized because of the presence of pseudo-inverse circulation. Although, a number of these difficult problems had been tried in some variants of hybridizations and optimizations of ELM.

From the review, we observe that the existing optimization techniques have certain drawbacks. Some of them being separate implementation of parameter optimization and feature selection (or) some of them are complex. Hence, in future, we aim to develop an effective ELM with multi-objective optimization in order to overcome the drawbacks of the existing ELM techniques. KELM with PSO is recognized as the best one to support ELM. KELM out of the considerable number of algorithms studied, and PSO has been implemented for the conclusion of strength transformers problem. KELM's parameters are further advanced by using PSO to boost KELM's performance. Deficiency identification of KELM-dependent strength transformers with PSO was contrasted with the other two ELMs, Back-engendering Neural Network (BPNN) and Bolster Vector Machines (SVM) samples from Disintegrated Gas Analysis (DGA). The estimation is yet to be completed for analysers other than GA and PSO. The measurement display for big data sets is yet to be tested.

Future Scope:

To research further, we can investigate the following problems for future Scope - To enhance the ELM, handling of high dimension data with ELM will be a curious topic to research about. To investigate further we can research about How to clarify the effectiveness of random mechanism. The connection between ELM and other related algorithms which accepts random mechanism that will be some highly interesting topic for study. From a theoretical perspective we can justify the optimization is used to improve ELM and this can be studied between ELM and other algorithms like cuckoo search, invasive weed optimization etc will be some highly interesting topic for study.

Acknowledgement

The paper "Improving Extreme Learning Machine through Optimization A Review." has been accepted in part for presentation at International Conference on Advanced Computing and Communication Systems (ICACCS) 2021.

References

- [1] Huang G-B, Zhu Q-Y, Siew C-K., "Extreme learning machine: theory and applications", *Neurocomputing*, vol.70, no.1-3, pp.489-501, 2006.
- [2] Cheng Y, Zhao D, Wang Y, Pei G., "Multi-label learning with kernel extreme learning machine autoencoder", *Knowledge-Based Systems*, vol.178, pp.1-0, August 2019.
- [3] Werbos PJ., "Generalization of backpropagation with application to a recurrent gas market model", *Neural networks*, vol.1, no.4, pp.339-56, January 1988.
- [4] Huang GB, Chen L, Siew CK., "Universal approximation using incremental constructive feedforward networks with random hidden nodes", *Neural Networks*, vol.17, no.4, pp.879-92, July 2006.
- [5] Huang GB., "An insight into extreme learning machines: random neurons, random features and kernels", *Cognitive Computation*, vol.6, no.3, pp.376-90, September 2014.
- [6] Raghuvanshi BS, Shukla S., "Class imbalance learning using UnderBagging based kernelized extreme learning machine", *Neurocomputing*, vol.329, pp.172-87, February 2019.
- [7] Galar, M., Fernandez, A., Barrenechea, E., Bustince, H., Herrera, F., "A review on ensembles for the class imbalance problem: Bagging, boosting and hybrid-based approaches", *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol.42, pp.63-484, 2012.
- [8] He, H., Garcia, E. A., "Learning from imbalanced data", *IEEE Transactions on Knowledge and Data Engineering*, vol.21, pp.263-1284, 2009.
- [9] Liu, X. Y., Wu, J., Zhou, Z. H., "Exploratory undersampling for class imbalance learning", *IEEE Transactions on Systems, Man and Cybernetics, Part B (Cybernetics)*, vol.39, pp.539-550, 2009.
- [10] Chawla, N. V., Bowyer, K. W., Hall, L. O., Kegelmeyer, W. P., "SMOTE: Synthetic minority over-sampling technique", *Journal of Artificial Intelligence Research*, vol.16, pp.321-357, 2002.
- [11] Cieslak, D. A., Hoens, T. R., Chawla, N. V., Kegelmeyer, W. P., "Hellinger distance decision trees are robust and skew-insensitive", *Data Mining and Knowledge Discovery*, vol.24, pp.136-158, 2012.

- [12] Tang, Y., Zhang, Y., Chawla, N. V., Krasser, S., "SVMs modeling for highly imbalanced classification", *IEEE Transactions on Systems, Man and Cybernetics, Part B (Cybernetics)*, vol.39, pp.281–288, 2009.
- [13] Mathew, J., Pang, C. K., Luo, M., Leong, W. H., "Classification of imbalanced data by oversampling in kernel space of support vector machines", *IEEE Transactions on Neural Networks Learning Systems*, pp.1–12, 2018.
- [14] Zong, W., Huang, G. -B., Chen, Y., "Weighted extreme learning machine for imbalance learning", *Neurocomputing*, vol.101, pp.229–242, 2013.
- [15] Shukla S, Raghuwanshi BS., "Online sequential class-specific extreme learning machine for binary imbalanced learning", *Neural Networks*, vol.119, pp.235-48, November 2019.
- [16] Kasun LL, Zhou H, Huang GB, Vong CM., "Representational learning with extreme learning machine for big data", *IEEE intelligent systems*, vol.28, no.6, pp.31-4, November 2013.
- [17] Wang, T., Cao, J., Lai, X., Chen, B, "Deep weighted extreme learning machine", *Cognitive Computation*, vol.10, no.6, pp.890–907, 2018.
- [18] Tang, J., Deng, C., Huang, G, "Extreme learning machine for multilayer perceptron", *IEEE Transactions on Neural Networks and Learning Systems*, vol.27, no.4, pp.809–821, 2016.
- [19] Wong CM, Vong CM, Wong PK, Cao J., "Kernel-based multilayer extreme learning machines for representation learning", *IEEE transactions on neural networks and learning systems*, vol.29, no.3, pp.757-62, December 2016.
- [20] Dai H, Cao J, Wang T, Deng M, Yang Z., "Multilayer one-class extreme learning machine", *Neural Networks*, vol.115, pp.11-22, July 2019.
- [21] Xue, X., Yao, M., Wu, Z., & Yang, J. "Genetic ensemble of extreme learning machine," *Neurocomputing*, 129, pp.175-184, 2014.
- [22] Matias, T., Souza, F., Araújo, R., & Antunes, C. H. "Learning of a single-hidden layer feedforward neural network using an optimized extreme learning machine," *Neurocomputing*, 129, pp.428-436, 2014.
- [23] Alexandre, E., Cuadra, L., Salcedo-Sanz, S., Pastor-Sánchez, A., & Casanova-Mateo, C. "Hybridizing extreme learning machines and genetic algorithms to select acoustic features in vehicle classification applications," *Neurocomputing*, 152, pp.58-68, 2015
- [24] Alencar, A. S., Neto, A. R. R., & Gomes, J. P. P. "A new pruning method for extreme learning machines via genetic algorithms," *Applied Soft Computing*, 44, pp.101-107, 2016.
- [25] Avci, D., & Dogantekin, A. "An expert diagnosis system for parkinson disease based on genetic algorithm-wavelet kernel-extreme learning machine," *Parkinson's disease*, 5264743. 2016.
- [26] Ertuğrul, Ö.F. and Kaya, Y, "A detailed analysis on extreme learning machine and novel approaches based on ELM," *American Journal of computer science and engineering*, vol.1, no.5, pp.43-50, 2014.
- [27] Chen, Y., & Zheng, W. X. "Stochastic state estimation for neural networks with distributed delays and Markovian jump. *Neural Networks*," 25, pp.14-20, January 2012
- [28] Albadra, M. A. A., & Tiuna, S. "Extreme learning machine: a review." *International Journal of Applied Engineering Research*, 12(14), pp.4610-4623, 2017.
- [29] Han, F., Yao, H. F., & Ling, Q. H. "An improved evolutionary extreme learning machine based on particle swarm optimization," *Neurocomputing*, 116, pp.87-93, 2013.
- [30] Zhao, M. R., Zhang, J. M., & Han, F. "An improved extreme learning machine with adaptive growth of hidden nodes based on particle swarm optimization," In 2014 International Joint Conference on Neural Networks (IJCNN) (pp. 886-890). IEEE, 2014, July.
- [31] Zhang, L., & Yuan, J. "Fault diagnosis of power transformers using kernel based extreme learning machine with particle swarm optimization," *Applied Mathematics & Information Sciences*, 9(2), pp.1003-1010, 2015.
- [32] Yadav, B., Ch, S., Mathur, S., & Adamowski, J. "Estimation of in-situ bioremediation system cost using a hybrid Extreme Learning Machine (ELM)-particle swarm optimization approach," *Journal of hydrology*, 543, pp.373-385, 2016.
- [33] Han, F., Zhao, M. R., Zhang, J. M., & Ling, Q. H. "An improved incremental constructive single-hidden-layer feedforward networks for extreme learning machine based on particle swarm optimization," *Neurocomputing*, 228, pp.133-142, 2017.
- [34] Zeng, N., Zhang, H., Liu, W., Liang, J., & Alsaadi, F. E. "A switching delayed PSO optimized extreme learning machine for short-term load forecasting," *Neurocomputing*, 240, pp.175-182, 2017.
- [35] Sun, W., Wang, C., & Zhang, C. "Factor analysis and forecasting of CO2 emissions in Hebei, using extreme learning machine based on particle swarm optimization," *Journal of cleaner production*, 162, 1095-1101, 2017.
- [36] Li H, Yang X, Li Y, Hao LY, Zhang TL., "Evolutionary extreme learning machine with sparse cost matrix for imbalanced learning", *ISA transactions*, vol.100, pp.198-209, May 2020.
- [37] Nahvi, B., Habibi, J., Mohammadi, K., Shamshirband, S., & Al Razgan, O. S. "Using self-adaptive evolutionary algorithm to improve the performance of an extreme learning machine for estimating soil temperature," *Computers and Electronics in Agriculture*, 124, pp.150-160, 2016.
- [38] Bazi, Y., Alajlan, N., Melgani, F., AlHichri, H., Malek, S., & Yager, R. R. "Differential evolution extreme learning machine for the classification of hyperspectral images," *IEEE Geoscience and Remote Sensing Letters*, 11(6), pp.1066-1070, 2013
- [39] Yang, Z., Zhang, T., & Zhang, D. "A novel algorithm with differential evolution and coral reef optimization for extreme learning machine training," *Cognitive neurodynamics*, 10(1), pp.73-83, 2016
- [40] Bharathi, P. T., & Subashini, P. "Optimal feature subset selection using differential evolution and extreme learning machine," *Int. J. Sci. Res*, 3(7), pp.898-1905, 2017
- [41] Guo, P., Cheng, W., & Wang, Y. "Hybrid evolutionary algorithm with extreme machine learning fitness function evaluation for two-stage capacitated facility location problems," *Expert Systems with Applications*, 71, pp.57-68, 2017
- [42] Wang, Y., Wang, A., Ai, Q., & Sun, H. "A novel artificial bee colony optimization strategy-based extreme learning machine algorithm," *Progress in Artificial Intelligence*, 6(1), pp.41-52, 2016
- [43] Mao, W., Tian, M., Cao, X., & Xu, J. "Model selection of extreme learning machine based on multi-objective optimization," *Neural Computing and Applications*, 22(3-4), pp.521-529, 2013

- [44] Ma, C. "An Efficient Optimization Method for Extreme Learning Machine Using Artificial Bee Colony," *Journal of Digital Information Management*, 15(3), 2017
- [45] Li, G., Niu, P., Ma, Y., Wang, H., & Zhang, W. "Tuning extreme learning machine by an improved artificial bee colony to model and optimize the boiler efficiency," *Knowledge-Based Systems*, 67, pp.278-289, 2014.
- [46] Alshamiri, A. K., Singh, A., & Surampudi, B. R. "Artificial bee colony algorithm for clustering: an extreme learning approach," *Soft Computing*, 20(8), pp.3163-3176, 2016.
- [47] Zhang, X., Yang, Z., Cao, F., Cao, J., Wang, M., & Cai, N. "Conditioning Optimization of Extreme Learning Machine by Multitask Beetle Antennae Swarm Algorithm," arXiv preprint arXiv:1811.09100, 2018.
- [48] Li, Q., Chen, H., Huang, H., Zhao, X., Cai, Z., Tong, C., ... & Tian, X. "An enhanced grey wolf optimization-based feature selection wrapped kernel extreme learning machine for medical diagnosis," *Computational and mathematical methods in medicine*, 9512741. 2017.
- [49] Wang, M., Chen, H., Li, H., Cai, Z., Zhao, X., Tong, C., ... & Xu, X. "Grey wolf optimization evolving kernel extreme learning machine: Application to bankruptcy prediction," *Engineering Applications of Artificial Intelligence*, 63, pp.54-68, 2017.
- [50] Cai Z, Gu J, Luo J, Zhang Q, Chen H, Pan Z, Li Y, Li C., "Evolving an optimal kernel extreme learning machine by using an enhanced grey wolf optimization strategy", *Expert Systems with Applications*, vol.138, pp.112-814, December 2019.
- [51] Wang, M., Chen, H., Yang, B., Zhao, X., Hu, L., Cai, Z., ... & Tong, C. "Toward an optimal kernel extreme learning machine using a chaotic moth-flame optimization strategy with applications in medical diagnoses," *Neurocomputing*, 267, pp.69-84, 2017
- [52] Salam, M. A., Zawbaa, H. M., Emary, E., Ghany, K. K. A., & Parv, B. "A hybrid dragonfly algorithm with extreme learning machine for prediction," In 2016 International symposium on innovations in intelligent systems and applications (INISTA) pp. 1-6. IEEE, 2016, August
- [53] Moorthy, U. and Gandhi, U.D, "A novel optimal feature selection technique for medical data classification using ANOVA based whale optimization," *Journal of Ambient Intelligence and Humanized Computing*, pp.1-12, 2020
- [54] Eshtay, M., Faris, H. and Obeid, N, "Improving extreme learning machine by competitive swarm optimization and its application for medical diagnosis problems," *Expert Systems with Applications*, vol.104, pp.134-152, 2018.
- [55] Galar, M., Fernandez, A., Barrenechea, E., Bustince, H., Herrera, F., "A review on ensembles for the class imbalance problem: Bagging, boosting and hybrid-based approaches", *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol.42, pp.63-484, 2012.
- [56] Salcedo-Sanz, S., Casanova-Mateo, C., Pastor-Sánchez, A., & Sánchez-Girón, M. "Daily global solar radiation prediction based on a hybrid Coral Reefs Optimization-Extreme Learning Machine approach," *Solar Energy*, 105, pp.91-98, 2018.
- [57] Nilesh, Rathod, and Wankhade Sunil. "Improving Extreme Learning Machine through Optimization A Review." In 2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS), vol. 1, pp. 906-912. IEEE, 2021.
- [58] M. Nguyen, L. D. Nguyen, T. Q. Duong and H. D. Tuan, "Real-Time Optimal Resource Allocation for Embedded UAV Communication Systems," in *IEEE Wireless Communications Letters*, vol. 8, no. 1, pp. 225-228, Feb. 2019.