Journal of Space Safety Engineering xxx (xxxx) xxx

Contents lists available at ScienceDirect

Journal of Space Safety Engineering

journal homepage: www.elsevier.com/locate/jsse

Anomaly detection of satellite telemetry based on optimized extreme learning machine

Sara Abdelghafar [a,1,*], Ashraf Darwish [b,1], Aboul Ella Hassanien [c,1], Mohamed Yahia [d], Afaf Zaghrout [e]

[a] Mathematics and Computer Science Division, Faculty of Science, Al Azhar University, Egypt
[b] Faculty of Science, Helwan University, Egypt
[c] Faculty of Computers and Artificial Intelligence, Cairo University, Egypt
[d] Space Division, National Authority for Remote Sensing and Space Sciences, Egypt
[e] Mathematics Department, Faculty of Science, Al Azhar University, Egypt

A R T I C L E I N F O

Keywords:
Anomaly detection
Extreme Learning Machine
Grey Wolf Optimization
Optimized predictive model
Satellite health monitoring
Telemetry mining

A B S T R A C T

In aerospace, anomaly detection based on telemetry data is a critical satellite health monitoring task that is important for identifying unusual or unexpected events and for taking measurements to improve system safety and avoid serious problems. This paper introduces a novel optimized predictive model for detecting anomalies using the Grey Wolf Optimization (GWO) algorithm and an Extreme Learning Machine (ELM), called GWO-ELM. The proposed GWO-ELM is used to find anomalous events by comparing the actual observed values with the predicted intervals of telemetry data; the GWO is applied to optimize the ELM's input weights and the bias parameters of hidden neurons to improve its prediction accuracy and ability to detect anomalies. A performance evaluation of GWO-ELM is conducted on the NASA shuttle valve benchmark dataset, which contains samples of Labeled anomalies and various metrics are collected. The experimental results for GWO-ELM show that it makes predictions with high efficiency, is stable when detecting anomalies, and requires little computational time. In addition, the results of GWO-ELM compared with those of the basic ELM algorithm with randomized parameter selection and a support vector machine (SVM), demonstrate the effectiveness and superiority of the proposed model.

1. Introduction

Health monitoring of artificial satellites involves analyzing telemetry data. Such data are considered the only basis on which a ground station can estimate and monitor the health state of an in-orbit satellite. Telemetry is non-stationary time series data that involves thousands of sensed values representing a state of the health, and mode of each subsystem as well as constantly varying environmental measurements [1,2]. Satellite health monitoring targets anomaly detection, which is critical because the early anomaly detection is important for avoiding disastrous situations such as loss of satellite control or serious faults. Most of the anomalies in the data translate to important and critical information, for example, anomalies in telemetry may indicate sensor faults, equipment or sensor failure, or degraded system performance. Recently, machine learning (ML) techniques have been successfully applied in monitoring the health of in-orbit satellites because the learning algorithms can learn to understand a system's behavior and updated real conditions to evaluate control function capabilities and detect abnormal behaviors such as anomalies or potential failures. Several techniques have been proposed to address detecting telemetry anomalies in space systems; however, predictive techniques are the most effective technique because they have strong learning ability, high detection efficiency, and are independent of expert prior knowledge [3]. Combined, these factors make prediction the most suitable choice for conducting anomaly detection in space systems.

In this paper, the prediction phase of the proposed model is based on an ELM, which has become increasingly widely applied in various learning domains for prediction and control applications due to its rapid learning speed, simple structure, and high generalizability. The ELM also avoids many challenges that affect other gradient-based learning methods, such as learning rate, local minima and epoch selection, and termination criteria [4,5]. The other phase of the proposed model involves optimization, which is applied to improve the ELM's performance by finding the optimal input weights and hidden bias parameters, because

* Corresponding author.
E-mail address: sara.abdelghafar@yahoo.com (S. Abdelghafar).
[1] Scientific Research Group in Egypt, (SRGE) www.egyptscience.net

https://doi.org/10.1016/j.jsse.2019.10.005
Received 31 August 2019; Received in revised form 2 October 2019; Accepted 17 October 2019
Available online xxx

Please cite this article as: S. Abdelghafar, A. Darwish and A.E. Hassanien et al., Anomaly detection of satellite telemetry based on optimized extreme learning machine, Journal of Space Safety Engineering, https://doi.org/10.1016/j.jsse.2019.10.005

randomly selecting these parameters requires more hidden neurons than does the classical gradient learning method, which delays the prediction response and reduces the chances of finding the optimal output weight. GWO was selected to optimize the ELM parameters because it has a stronger search capability compared to other swarm optimization methods, requires few control parameters and no information derivation for its initial search, and achieves favorable convergence due to its good balance between exploitation and exploration during the search iterations [6,7].

Based on the reasons above, this paper proposes a novel optimized predictive model, called GWO-ELM, for detecting anomalies from satellite telemetry data. GWO-ELM first obtains the predicted values and then adds an error margin to the predicted values based on the confidence interval method to introduce an estimated normal range of new observed values, which are used to detect anomalies when values exceed this range. Different metrics regarding prediction, anomaly detection, stability, and time complexity have been applied to evaluate the performance of the proposed GWO-ELM model. In addition, a comparative analysis of GWO-ELM's prediction and detection accuracy is applied to demonstrate its effectiveness and superiority compared to the basic ELM algorithm with randomized parameter selection and an SVM. The evaluation uses the NASA shuttle valve benchmark dataset, which includes various anomaly samples labeled by NASA experts [8]. The remainder of this paper is organized as follows. Section 2 introduces briefly the related work. Section 3 introduces the basic operations and background of the GWO and ELM methods. In Section 4, the proposed model is presented in detail. Section 5 presents the experimental datasets and the evaluation metrics and reports the experimental results. Section 6 concludes this work and suggests some future research directions.

2. Related work

The effectiveness of prediction has been demonstrated by a variety of models proposed by many works to detect telemetry anomalies based on a variety of ML techniques, such as relevance vector learning [9], which was experimented using spacecraft telemetry of attitude determination and control subsystem (ADCS) and propulsion subsystem, the proposed approach of this study is to detect anomaly of telemetry by the predictive model that is based on adopted relevance vector regression (RVR) hybrid with extended autoregressive (AR), the model showed its performance and superiority over-limit checking method in efficiency of detecting anomalies, also in avoiding false alarms that occur in the case of applying traditional AR and RVR. Another approach of using prediction for detection anomalies was proposed in [10], based on least square support vector machine (LS-SVM) as an optimized predictive model for detecting anomalies of telemetry, which proved its detection efficiency but with some limitations such as most of the parameters that need to be set by experiments and expert experience. Recently, long short term memory (LSTM) technique of recurrent neural networks is addressed in [11] for proposing a predictive model that used with dynamically threshold method for anomaly detection of spacecraft, as LSTM has proven its efficiency of prediction which increases the ability of anomaly detection. LSTM is introduced also with other five neural network algorithms in [12] that have been compared to the linear regression model to prove the efficiency of neural network techniques for prediction and then anomaly detection of long lifetime satellite. All previous examples show how the ability of anomaly detection in satellites telemetry data depends mainly on the effectiveness of predictive models.

The approach of optimizing an ELM using an evolutionary algorithm has been introduced in many works, as it was first introduced in [13], which proposed using differential evolution (DE) to optimize an ELM model. The performance of the resulting DE-ELM model was validated on four real benchmark classification problems. Subsequently, more works have applied different meta-heuristic algorithms to optimize ELM models for different applications. For example, the DE was used again for classification of hyperspatial images [14], and [15] proposed and tested the optimized ELM model on sixteen benchmark datasets using three optimization algorithms; DE, the genetic algorithm (GA) and simulated annealing (SA). Many optimized ELM models have also been proposed with different optimization algorithms. For example, [16] used the bacterial foraging (BF) algorithm, which simulates biological bacterial food-searching behavior, and [17] compared its ability to optimize an ELM with that of the group search optimizer (GSO) which is based on a generic animal social search behavior, and particle swarm optimization (PSO). The PSO has been applied for ELM optimization in multiple works that target different applications, including different prediction problems [18], medical classification applications [19], and as introduced in [20], a hybrid evolutionary approach based on PSO with a clustering method to obtain a local best topology. This hybrid approach was validated with classification problem using four benchmark classification datasets. In recent works, two other methods have been used to optimize an ELM: artificial bee colony (ABC) optimization was used in [21] for different classification problems and tested using eight benchmark datasets, and the chemical reaction optimization (CRO) was used in [22] and tested on a classification problem.

Recently, other models of GWO and ELM have been introduced in different work, as for wind power predication problem that has been studied in [23], as multi-objective grey wolf optimizer (MOGWO) is utilized with AdaBoost method, wavelet packet decomposition (WPD) and the outlier-robust extreme learning machine (ORELM) to propose an ensemble model with different hyper-parameters, the proposed ensemble model shows good performance in terms of convergence and accuracy prediction. Also in [24] for the same problem, MOGWO has been utilized again to optimize ELM initial weights and thresholds, prediction accuracy and stability are used as the objective functions in the proposed experiment of this study that proves the superiority of MOGWO over other two multi-objective algorithms in terms of both prediction accuracy and stability. Another work for the wind power prediction is [25], the proposed model is a hybrid model based on empirical wavelet transform (EWT) for decomposition and the inverse empirical wavelet transform (IEWT) for reconstruction, with optimized predictive model that is based on GWO and regularized extreme learning machine (RELM) to promote the efficiency of the prediction as that has been demonstrated in the experimental results.

As well as, GWO has been successfully applied also to a variety of model problems similar to those of our proposed model. Such as, GWO was used to optimize Q-Gaussian Radial Basis Functional-link neural networks; the results showed that the proposed approach using GWO achieves highly competitive performance compared with other evolutionary algorithms [26]. The work proposed in [27] used GWO to optimize multi-layer perceptrons (MLPs) by finding the optimal weights and bias parameters during the training phase. The experimental results showed that the GWO achieved a high level of accuracy for finding appropriate approximations of the optimal weights and bias values for MLPs compared to five other meta-heuristic optimization algorithms when tested on five standard classification datasets and three function-approximation datasets. Additionally, the GWO outperformed five competing meta-heuristic algorithms in optimizing the parameter of an SVM, resulting in faster convergence and good stability for a failure recovery problem of satellite telemetry data as was proposed in [28].

3. Basics and background

3.1. Grey Wolf Optimization

GWO is a recent evolutionary algorithm proposed by Mirjalili et al. [29]. The GWO algorithm simulates the hunting behavior of a grey wolf pack. In a wolf pack, the group leader is called the alpha, which leads the group in all decisions such as hunting, movement, sleeping and waking up. The secondary boss of the grey wolf pack is called the beta, who acts as the main consultant to the Alpha for decision making, reinforces the alpha's orders, and gives the alpha feedback. Additionally, the beta is responsible for group discipline. The third level of leaders is called deltas, who lead the lowest ranking wolves, called omegas, by transmitting orders from the leaders to the other group members. The GWO algorithm achieves an optimum solution by simulating the grey wolf pack's encircling mechanism when hunting prey in the wild. The obtained first three best solutions are mathematically presented as the leaders of the grey wolves: alpha represents the optimum solution, beta is the second-best, and delta is the third best. The worst solution is defined as an omega. Therefore, the optimization in the GWO algorithm mimics the hunting wolf pack led by alpha, beta and delta. The omega wolves follow these leaders. Eqs. (1) and (2) present the grey wolves' encircling mechanism for hunting prey:

$$\overrightarrow{D} = |\overrightarrow{C}.\overrightarrow{X}_p(t) - \overrightarrow{X}(t)| \tag{1}$$

$$\overrightarrow{X}(t+1) = \overrightarrow{X}_p(t) - \overrightarrow{A}.\overrightarrow{D} \tag{2}$$

where t is the number of iterations, $\overrightarrow{X}_p(t)$ is the current position of prey; $\overrightarrow{X}(t)$ is the current position of a wolf; and $\overrightarrow{D}$ is the distance between wolves and prey used to determine the updated grey wolf positions. $\overrightarrow{A}$ and $\overrightarrow{C}$ are coefficient vectors used for exploitation and exploration, respectively and are obtained by Eqs. (3) and (4), respectively:

$$\overrightarrow{A} = 2\overrightarrow{v}.\overrightarrow{r}_1 - \overrightarrow{v} \tag{3}$$

$$\overrightarrow{C} = 2.\overrightarrow{r}_2 \tag{4}$$

where $\overrightarrow{v}$ linearly decreases from 2 to 0 over the course of the iterations, and $\overrightarrow{r}_1$ and $\overrightarrow{r}_2$ are random values in the range [0,1] used to mimic the prey being hunted. Whenever $\overrightarrow{A}$'s value decreases, the wolves approach the prey more closely and can attack it. In contrast, the vector $\overrightarrow{C}$ represents natural obstacles that prevent the wolves from approaching the prey. GWO's search process is initialized with a random population; then, the optimal solution is gradually obtained in the iterations, by updating the positions of the grey wolves based on their distance from the prey, saving the three obtained optimum solutions; alpha α, beta β and delta δ, and then forcing the omega to optimize their positions [6,30] as shown below:

$$\overrightarrow{D}_\alpha = |\overrightarrow{C}_1.\overrightarrow{X}_\alpha - \overrightarrow{X}|; \overrightarrow{D}_\beta = |\overrightarrow{C}_2.\overrightarrow{X}_\alpha - \overrightarrow{X}|;$$
$$\overrightarrow{D}_\delta = |\overrightarrow{C}_3.\overrightarrow{X}_\alpha - \overrightarrow{X}| \tag{5}$$

$$\overrightarrow{X}_1 = \overrightarrow{X}_\alpha - A_1.(\overrightarrow{D}_\alpha); \overrightarrow{X}_2 = \overrightarrow{X}_\beta - A_2.(\overrightarrow{D}_\beta);$$
$$\overrightarrow{X}_3 = \overrightarrow{X}_\delta - A_3.(\overrightarrow{D}_\delta) \tag{6}$$

3.2. Extreme learning machine

An ELM is a learning algorithm introduced by Huang et al. [31] for training single-layer feedforward neural networks (SLFNs). The ELM consists of three layers: input, hidden and output. The training process is initialized by choosing random values for the input layer weights and the biases of the hidden layer, and it calculates the output weights using the least-squares method to achieve the smallest training error. For N arbitrary samples $(x_i, y_i), i = 1, \ldots, N$, where $x_i = [x_{i1}, x_{i2}, \ldots, x_{id}]^T \in R_d$ and $y_i = [t_{i1}, t_{i2}, \ldots, t_{im}]^T \in R_m$ are the i^{th} input and target output the output of a generalized activation function with L hidden nodes can be defined mathematically by:

$$\sum_{j=1}^{L} W_j G(a_j, b_j, x) = o_i \tag{7}$$

where W_j is the output weight vector connecting the jth hidden neurons and the output neurons, a_j refers to the weight vector connecting the jth hidden neurons and input neurons, and b_j is the bias of the jth hidden neuron [21,32]. The generalized structure of an ELM is shown in Fig. 1.

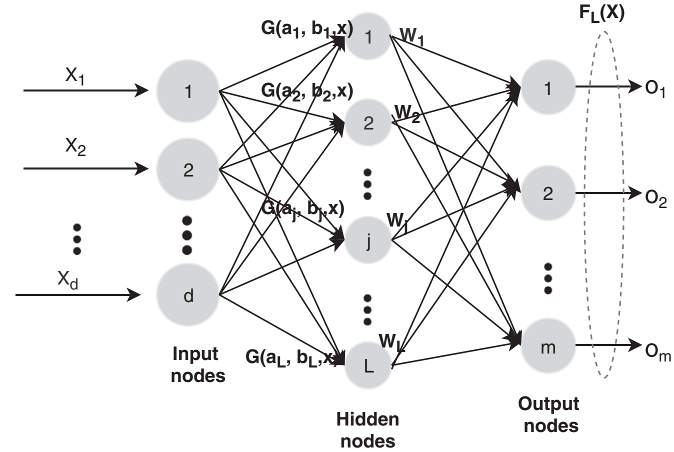

Fig. 1. Generalized ELM structure.

A standard SLFN can approximate the output of N samples with zero error [31], $\sum_{i=1}^{N} \|o_i - t_i\| = 0$. Therefore, Eq. (7) can be represented as follow:

$$\sum_{j=1}^{L} W_j G(a_j, b_j, x) = t_i \tag{8}$$

Hence it can be also represented by following:

$$HW = T \tag{9}$$

where $T = [t_1, t_2 \ldots, t_m]^T$ is a matrix of the target output, and H is the hidden layer output function mapping the learning feature space, $H(x) = [G(a_1, b_1, x), \ldots, G(a_L, b_L, x)]$, when the input parameters (a_j, b_j) are randomly generated over the training samples. The output weight W is obtained by Eq. (10) be finding the minimum norm least-square solution that minimizes the training error and achieves the best generalization performance [33].

$$W = H^{+}T \tag{10}$$

where H^{+} is the Moore–Penrose (MP) generalized inverse [34].

Due to the random choices of input weights and hidden bias parameters, an ELM requires large numbers of hidden neurons, which may lead to a non-optimal parameter set causing the model to be unable to reach the global optimum solution in some cases. As showed in the previous section, many works have been proposed that use meta-heuristic methods to find the optimal parameters for ELM models, these algorithms can maximize the network's generalizability by choosing optimal input weights and hidden bias values during the training process, which repeats until a termination criterion is achieved.

4. The proposed anomaly detection model

This section introduces the proposed optimized predictive model GWO-ELM for detecting anomalies in satellite telemetry data. In general, anomaly detection prediction is generally conducted in two phases: the first phase involves computing the predicted values, and the second phase obtains the uncertainty interval of these predicted values (Fig. 2 and Algorithm 1).

In the first phase of the proposed anomaly detection model, GWO-ELM is used to build an optimized predictive model for obtaining the predicted values of telemetry data. To obtain the optimized predictive model GWO is applied to find the optimal ELM parameters. GWO searches to find the best set of input weights (a) and biases (b) over a set of iterations; then, the best prediction performance is achieved by finding the optimal output weights (W) and hidden layer output matrix (H) values to obtain the optimal target output (T). The flowchart for optimizing the ELM starts with the initial best solution of GWO, which

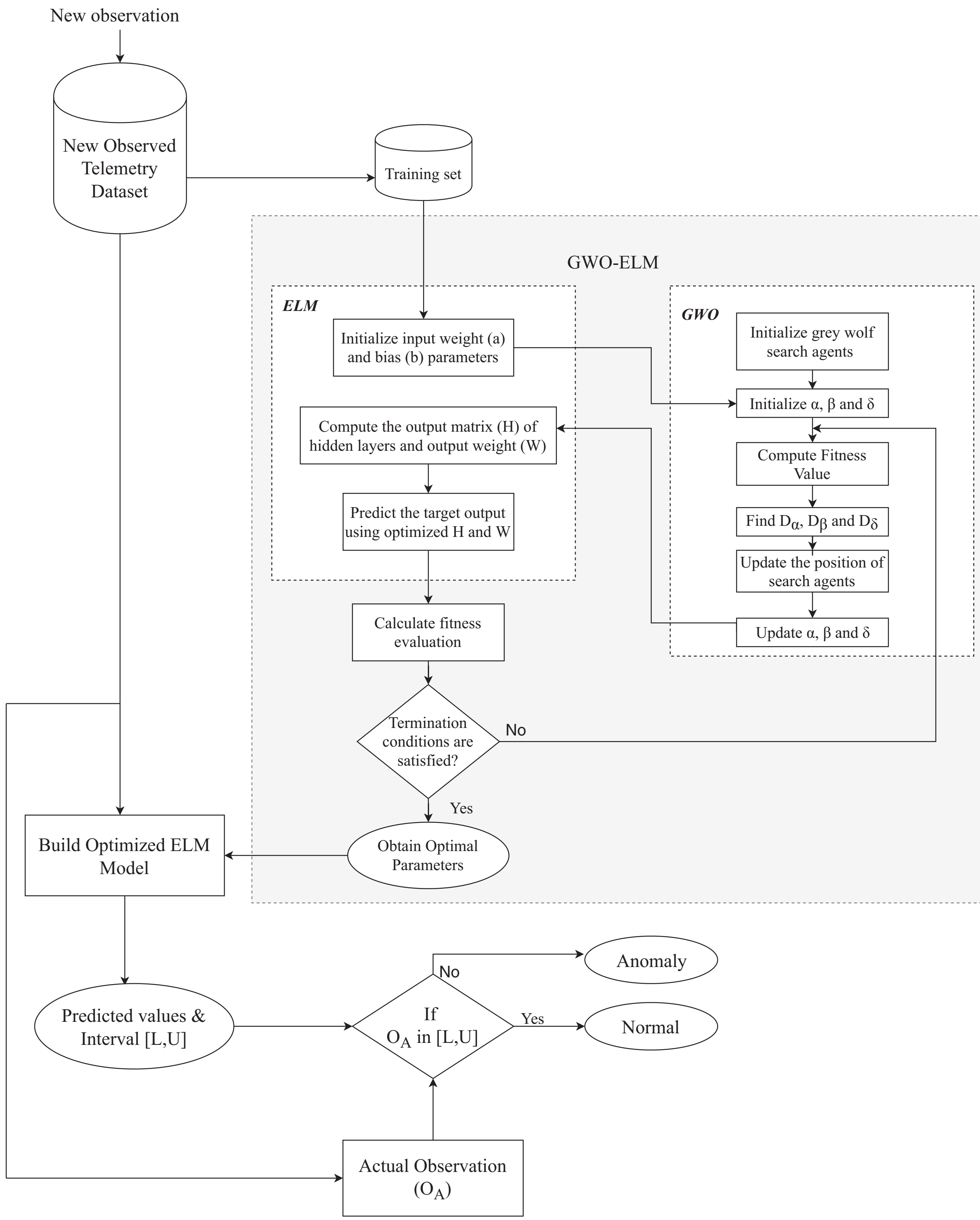

Fig. 2. The proposed anomaly detection model based on GWO-ELM.

is generated based on randomly initialized input weights and bias parameters of ELM. Then, the fitness of each search agent is evaluated and the positions of α, β and δ are computed and used to update the positions of the wolves in each search. The goal is to find the probable estimated positions of each search agent through the iterations to achieve the optimal GWO solution. Then, control returns to ELM, which applies the optimized a and b to the ELM training model. Because the sigmoid function is selected as the activation function, H is calculated based on Eq. (11), and W is calculated from Eq. (10).

$$H = \sum_{j=1}^{L} g(a_j.x_i + b_j) \tag{11}$$

The prediction RMSE (Root Mean Square Error) of the training set is used as the fitness function. It determines the best fitness value by minimizing

Algorithm 1 Anomaly detection based on optimized predictive model GWO-ELM.

Input: $D = (x_i, y_i)|x_i \in R_d, y_i \in R_m, i = 1, \ldots, N$: Training dataset with x_i is input y_i is target output.
and $T = (x_i, y_i)|x_i \in R_d, y_i \in R_m, i = 1, \ldots, N$: Testing dataset with x_i is input and y_i is target output.
L: Hidden Nodes, $g(x)$: Activation Function, $F(x) = \frac{1}{1+RMSE}$: Fitness Function, P: Population Size,
$MaxIter$: Maximum number of iterations, $MaxFit$: 1, A and C: coefficient vectors.
Output: H: Hidden Layer Matrix, W: Output Weight, D_{Ano}: Anomaly Dataset and D_{Nor}: Normal
Dataset.
Initialize Input Weights (a_i) and biases (b_i)randomly [-1,1];
Initialize t← 0;
Initialize grey wolf search agents;
Initialize α β and δ with initial a_i and b_i;
While t < MaxIter and fitness value < MaxFit **Do**
Evaluate the fitness of each search agent;
Compute D_α, D_β and D_δ through Eq. (5);
Update the position of each search agent;
Update α, β and δ;
Insert updated best solution in the activation function $g(x)$ to calculate H from Eq. 11;
Find W using MP generalized inverse from Eq. (10);
Calculate the predicted target output using extracted H and W;
Calculate fitness evaluation using RMSE from Eq. (12);
t← t+1;
End while
return H and W
Initialize: set $D_{Nor} \leftarrow \Phi$, $D_{Ano} \leftarrow \Phi$.
Calculate the target predicted value Y_p of testing set using Eq. (9);
Calculate U and L using Eq. (13);
For all instance y_i in actual observation of target output Y_A **Do**
If $y_i \geq L$ and $\leq U$
$S_{Nor} \leftarrow y_i$
else
$S_{Ano} \leftarrow y_i$
End if
Update $D_{Nor} \leftarrow D_{Nor} \cup S_{Nor}$
Update $D_{Ano} \leftarrow D_{Ano} \cup S_{Ano}$
End for
return D_{Ano} and D_{Nor}

the prediction error as shown in Eq. (12) [35]:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_{P_i} - y_{A_i})^2} \tag{12}$$

Where $Y_P = [y_{P_1}, y_{P_2}, \ldots, y_{P_n}]$ is the prediction value of n samples and $Y_A = [y_{A_1}, y_{A_2}, \ldots, y_{A_n}]$ is the actual value.

This evaluation continues for the output of each population until it reaches the optimal fitness value or a maximum number of iterations. When generation terminates, the obtained H and W values are associated with the optimal fitness value of the training set that will be used to predict the target output of new observed data.

After the predicted values have been obtained, the second phase of the proposed model computes the uncertainty interval of the predicted values by adding the margin of error based on the confidence interval method. This estimated uncertainty interval is used to anomalies in the observed data by comparing them with the previously observed values, if each value is located within the interval, it is labeled as normal; conversely, if the value lies outside the interval it is labeled as an anomaly. This process is defined in the following Eq. (13) over Y_P at value t [36]:

Table 1
Samples of NASA shuttle valve dataset with anomalies.

Sample No.	Sample Size	Anomaly Size(%)
S1	2000	5%
S2	3000	7.5%
S3	4000	10%
S4	3000	12.5%
S5	5000	15%

$$\left(\bar{y} - t_{\frac{\alpha}{2}}(n-1)\frac{S}{\sqrt{n}}, \bar{y} + t_{\frac{\alpha}{2}}(n-1)\frac{S}{\sqrt{n}}\right) \tag{13}$$

where $\bar{y}$ is the mean of $Y_P i$, and S is the mean square deviation. Therefore, the upper limit of output normal range (U) is defined as $\bar{y} + t_{\frac{\alpha}{2}}(n-1)\frac{S}{\sqrt{n}}$, and the lower limit ($L$) is $\bar{y} - t_{\frac{\alpha}{2}}(n-1)\frac{S}{\sqrt{n}}$. So The rules of anomaly detection are as following:

if $L \leq Y_P \geq U$ the observed measure of telemetry is normal.
if $Y_P \geq U or Y_P \leq L$ the observed measure of telemetry is an anomaly.

5. Experimental results and discussion

The experiments were implemented in MATLAB R2016a on a computer with an Intel (R) Core (TM) i5-6200U CPU@2.40 GHz with 8 GB of RAM. Based on the results of our preliminary experiment to obtain the optimum values of the initial parameters, the number of GWO search agents was set to 30, the maximum number of iterations was 100, the number of hidden neurons in ELM was 30, and the sigmoid function was selected as the activation function. For the SVM, it is a supervised learning algorithm that first introduced by Boser in 1992 [37] for classification and nonlinear estimation function, and then SVM was extended to the regression by Drucker et al. [38], which has been proven its excellent performance in real-value function estimation. SVMs objective for regression is to get function *f*(*x*), which has at most ε deviation from the target value for each training point *x*. SVM regression is considered a nonparametric technique as it relies on kernel functions. In the proposed experiment, the polynomial kernel is selected for the kernel function which is commonly used with SVM. The experiment was conducted after scaling and normalizing the data, then dividing it into 70% and 30% ratio for training and testing dataset respectively to evaluate the performance of the proposed model against ELM and SVM.

5.1. Data description

The experiment uses the NASA shuttle valve dataset to evaluate the efficiency of the proposed model for anomaly detection from space application telemetry data. The shuttle valve dataset was compiled from health monitoring measurements of the current for electromagnetic valves in the space shuttle under various conditions; voltage, temperature, and impedance. Based on the high temperature, outlier voltage or poppet impedance, the dataset includes abnormal ranges marked by NASA experts. The dataset is divided into training and test sets, where the training set includes the normal samples that have been recorded at the normal range of voltage, temperature, and poppet impedance, then test on some normal samples merged with abnormal samples which are recorded at outlier voltage, high temperature (69–71°C) or poppet impedance values at 9 or 45 mils. The normal range of voltage is 18, 20, 22, 24, 26, 28 or 30 V, while the outlier is considered at 14, 16 and 32 V. In this work, different sizes of labeled anomalies were added randomly to normal ranges and used in the evaluation experiment to test the detection ability of the proposed model, as shown in Table 1.

5.2. Performance evaluation measures

Different evaluation criteria are applied to evaluate the performance of the proposed model. RMSE and mean absolute error (MAE) are used

Table 2
Confusion matrix of anomaly detection evaluation measures.

Actual	Detected	
	Normal	Anomaly
Normal	T_N	F_P
Anomaly	F_N	T_P

to measure the prediction accuracy, which must be evaluated first because the accuracy of the predicted values directly impact the model's ability to perform anomaly detection. The RMSE and MAE assess the prediction performance and are calculated as shown in Eq. (12) and Eq. (14), respectively [39].

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|Y_{Pi} - Y_{Ai}| \quad (14)$$

For anomaly detection, the evaluation measures are reported in the form of a confusion matrix as shown in Table 2 [40], where T_N represents a normal sample range predicted correctly, and F_P represents a normal sample range predicted incorrectly. For anomalous samples, T_P and F_N respectively represent a correctly predicted and incorrectly predicted range.

The evaluation standard measures that based on the confusion matrix can be defined and calculated to evaluate the performance of the detection approach as following [41]:

1. Sensitivity: Represents anomaly detection rate by finding the proportion of anomaly that was correctly predicted to the total number of actual anomaly samples.

$$Sensitivity = \frac{T_P}{T_P + F_N} \quad (15)$$

2. PPV: Is the proportion of anomaly that was correctly predicted to the total number of samples detected as anomaly.

$$PPV = \frac{T_P}{T_P + F_P} \quad (16)$$

3. F-Score: Is a measure of the balance between sensitivity and PPV through the following equation:

$$F - Score = \frac{2 * T_P}{2 * T_P + F_P + F_N} \quad (17)$$

4. False Positive Rate (FPR): Is the proportion of the actual normal samples that were incorrectly predicted as an anomaly to the total number of actual normal samples.

$$FPR = \frac{F_P}{F_P + T_N} \quad (18)$$

5. False Negative Rate (FNR): Is the proportion of the actual anomaly samples that were incorrectly predicted as normal to the total number of actual anomaly samples.

$$FNR = \frac{F_N}{T_P + F_N} \quad (19)$$

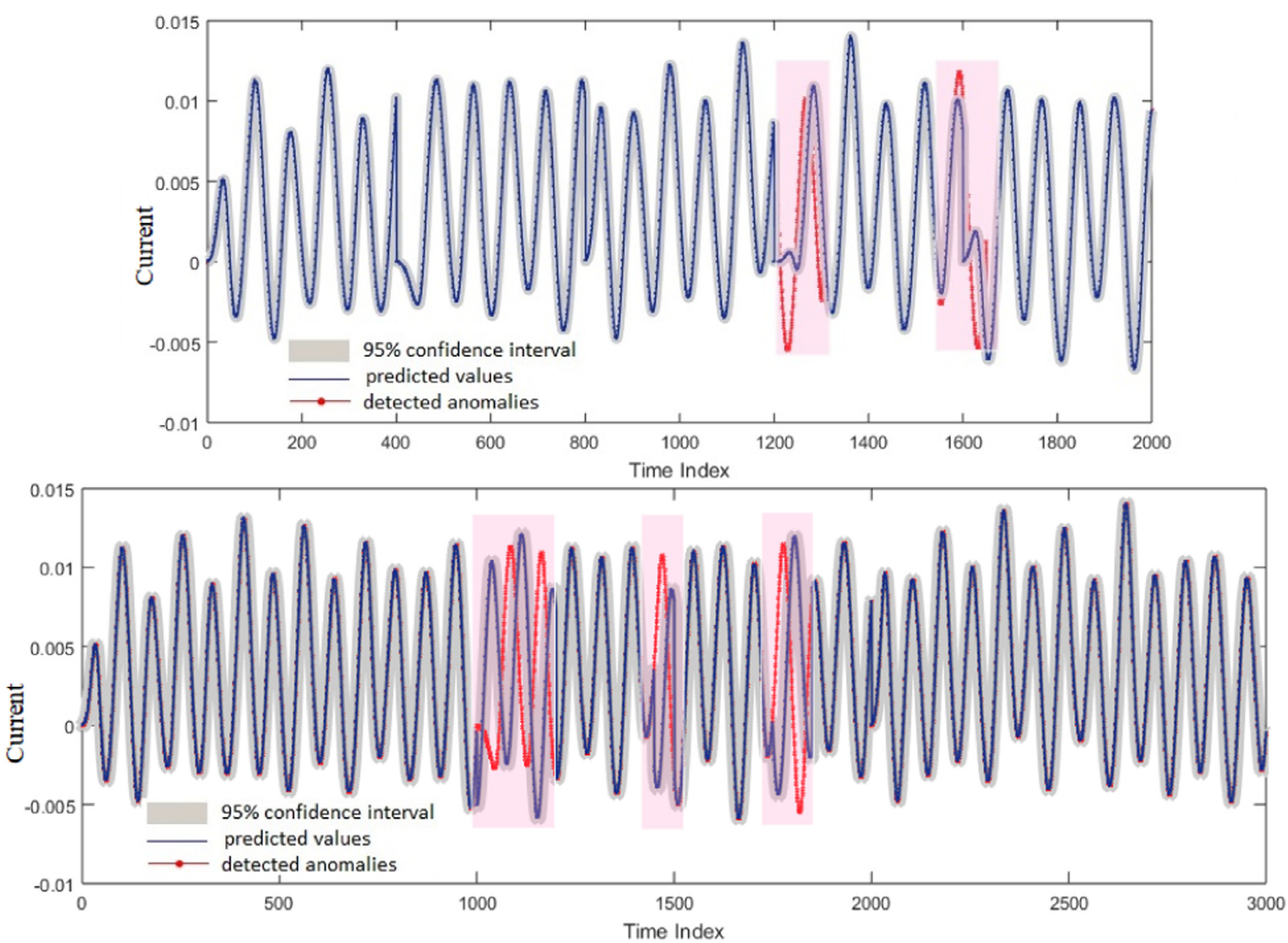

Fig. 3. Detection results for anomalies of samples *S*1 and *S*2. (shaded red region determines anomaly ranges). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 3
Prediction performance evaluation results for GWO-ELM.

Sample	RMSE	MAE	Prediction Acc%
S1	0.0045	0.00395	99.55%
S2	0.0049	0.0043	99.51%
S3	0.006	0.0053	99.4%
S4	0.013	0.0114	98.7%
S5	0.015	0.0132	98.5%

Table 4
Anomaly detection performance evaluation results for GWO-ELM.

Sample	Sensitivity	PPV	F-Score	FPR	FNR	MCC	Detection Acc%
S1	1	0.9852	0.9926	0.0063	0	0.9895	99.56%
S2	1	0.96	0.9796	0.0058	0	0.9773	99.55%
S3	1	0.9474	0.973	0.0056	0	0.9691	99.25%
S4	0.9831	0.9063	0.9431	0.0148	0.017	0.9376	98.83%
S5	0.913	0.9545	0.9333	0.0059	0.087	0.9249	98.46%

Table 5
Time complexity and stability evaluation results for GWO-ELM.

Sample	Mean	Std.	Execution Time (Sec)
S1	0.0312	0.009402	26.3
S2	0.04	0.0016	28.5
S3	0.0051	0.000316	37.2
S4	0.0125	0.00934	32.4
S5	0.0183	0.0076	58.8

6. MCC: Is used to evaluate the performance of the detection method by measuring the correlation between actual and detected results.

$$MCC = \frac{T_P * T_N - F_P * F_N}{\sqrt{(T_P + F_N)(T_P + F_P)(T_N + F_N)(T_N + F_P)}} \tag{20}$$

MCC range is between −1 and 1, where 1 means that all results were detected correctly, on the contrary, −1 means that all results were detected wrongly.

7. Overall detection accuracy is the proportion of the total number of correctly predicted to the total detected samples.

$$ACC = \frac{T_P + T_N}{T_P + T_N + F_P + F_N} \tag{21}$$

Regarding the optimization process, time complexity and stability are the most important characteristics of the algorithm and should be included in the evaluation metrics. The CPU time required for training is used to evaluate the computational complexity of the algorithm, and stability is measured through the mean and standard deviation of the fitness evaluation. These metrics are calculated as defined in Eq. (22) and Eq. (23) respectively [42].

$$\mu = \frac{\sum_{i=1}^{n_r} Q_i}{n_r} \tag{22}$$

$$Std = \sqrt{\frac{\sum_{i=1}^{n_r} (Q_i - \mu)^2}{n_r}} \tag{23}$$

where n_r is the number of iterations and Q is the optimal solution obtained each iteration.

5.3. Results, discussion and analysis

According to the prediction evaluation measures, Table 3 shows the obtained results of GWO-ELM for RMSE, MAE and prediction accuracy on the five samples of the applied dataset. As shown by the results, neither RMSE nor MAE exceeded 0.015 for all the samples, which indicates that the estimation precision of GWO-ELM in the prediction process is high. The accuracy rates range from 98.5% to 99.55%.

Table 6
Comparison between GWO-ELM, ELM and SVM with five samples of NASA shuttle dataset.

Sample	Method	Prediction Acc%	Detection Acc%
S1	GWO-ELM	99.55%	99.56%
	ELM	99%	98.97%
	SVM	98.97%	98.1%
S2	GWO-ELM	99.51%	99.55%
	ELM	98.67%	98.64%
	SVM	98.9%	98%
S3	GWO-ELM	99.4%	99.25%
	ELM	97.75%	96.5%
	SVM	96.3%	95.75%
S4	GWO-ELM	98.7%	98.83%
	ELM	97.67%	95.6%
	SVM	96.1%	95.45%
S5	GWO-ELM	98.5%	98.46%
	ELM	96.57%	94.4%
	SVM	95.1%	94.51%

For the anomaly detection evaluation measures, Table 4 reveals that GWO-ELM obtained high performances on all the evaluation measures for the five samples. The detection accuracy results range from 98.5% to 99.6% for 15% and 5% anomaly ranges in the samples, respectively. The MCC never fell below 0.925 even when the anomaly increased to 15% of the sample size, and it reached approximately 0.99 when the anomaly was decreased to 5% of the sample size. The strength of the detection rate is indicated by the sensitivity results, which reached 1 on the first three samples, and in the F-Score, which represents the harmonic mean between sensitivity and PPV, and whose results range from 0.993 to 0.933 on the five samples. On the other side, the efficiency of anomaly detection is also indicated by avoiding FPR which represents the false alarms for anomalies that were incorrectly predicted, whose results never higher than 0.01 and in most samples decreased to 0.005. Also, the results of FNR never higher than 0.08 even when the anomaly increased to 15% of the sample size and decreased to 0 on the first three samples. Fig. 3 shows examples of the detection results for two samples, $S1$ and $S2$.

Regarding the optimization evaluation aspects, mean and standard deviation results indicate that the GWO-ELM model is stable and that it achieved good convergence with 100 iterations for all the applied samples, as shown in Table 5. In addition, execution time results indicate that the experiments with all the samples reached an optimal fitness evaluation with low computational time (which never exceeded 59 sec for a sample size of 5,000). The comparison results in Table 6, show the superiority of the proposed model GWO-ELM in both prediction and detection accuracy compare to the ELM and SVM algorithms. These results reflect the suitability and effectiveness of the GWO-ELM for detecting anomalies from space application telemetry data.

6. Conclusion and future work

This paper contributes to health monitoring in space applications that require anomaly detection to identify unusual or unexpected events and measurements to improve system safety and avoid serious faults or disastrous situations. The contributions of this paper are summarized as follows. A novel predictive model to detect anomalies in aerospace applications based on GWO and ELM is proposed. The ELM was selected due to its rapid learning capability, simple structure, and high generaliz ability, which has been demonstrated in various domains for prediction and control applications. As well as, GWO is known for its strong search capability and efficient balance between exploration and exploitation during the search process. The experiments demonstrated the effectiveness of the proposed optimized GWO-ELM model for detecting anomalies from a NASA public shuttle valve benchmark dataset. The evaluation was conducted using different prediction metrics that measure a model's ability to perform anomaly detection. Regarding the optimization aspects, stability and time complexity were measured. The experimental results illustrated that the proposed model has high prediction efficiency and can stably detect anomalies with low computational time. In addition, a comparative evaluation demonstrated that GWO-ELM is superior with regard to prediction and detection accuracy compared to over the basic ELM algorithm with randomized parameter selection and an SVM algorithm.

However, the proposed model has some limitations caused by applying the static confidence interval with the standard error method to produce the estimated normal range. This aspect requires further research in the future to ensure that the adaptive predictive interval is applicable to real-time detection problems. Additionally, anomaly detection for multivariate samples of real space applications with multiple modes and different conditions deserves extensive study in the future.

Acknowledgments

This work is supported by Egypt Knowledge and Technology Alliance (E-KTA) for Space Science, which is funded by The Academy of Scientific Research & Technology (ASRT), and coordinated by National Authority for Remote Sensing & Space Sciences (NARSS).

References

[1] T. Yairi, N. Takeishi, T. Oda, Y. Nakajima, N. Nishimura, N. Takata, A data-driven health monitoring method for satellite housekeeping data based on probabilistic clustering and dimensionality reduction, IEEE Trans. Aerosp. Electron.Syst. 53 (3) (2017) 1384–1401.
[2] D.R. Azevedo, A.M. Ambrosio, M. Vieira, Applying data mining for detecting anomalies in satellites, in: Proc. EDCC, Sibiu, Romania, 2012. 212–217
[3] T. Yairi, Y. Kawahara, R. Fujimaki, Y. Sato, K. Machida, Telemetry mining: A machine learning approach to anomaly detection and fault diagnosis for space systems, in: Proc. SMC-IT, Pasadena, CA, USA, 2006, pp. 468–476.
[4] S. Ding, X. Xinzheng, R. Nie, Extreme learning machine and its applications, Neural Comput. Appl. 25 (2014) 549–556.
[5] H. Yang, P. Thomas, O. Finka, Fault detection based on signal reconstruction with auto-associative extreme learning machines, Eng. Appl. Artif.Intell. 57 (2017) 105–117.
[6] H. Faris, I. Aljarah, M.A. Azmi, S. Mirjalili, Grey wolf optimizer: a review of recent variants and applications, Neural Comput. Appl. 30 (2018) 413–435.
[7] W. Mingjing, H. Chen, L. Huaizhong, C. Zhennao, Z. Xuehua, T. Changfei, L. Jun, X. Xin, Grey wolf optimization evolving kernel extreme learning machine: application to bankruptcy prediction, Eng. Appl. Artif. Intell. 63 (2017) 54–68.
[8] B. Farrell, S. Santuro, NASA shuttle valve data, 2019Online. Available: http://www.cs.t.edu/pkc/nasa/data.
[9] R. Fujimaki, T. Yairi, K. Machida, An anomaly detection method for spacecraft using relevance vector learning, in: Proc. Pacific-Asia Conference on Knowledge Discovery and Data Mining, Hanoi, Vietnam, 2005, pp. 785–790.
[10] L. Datong, P. Jingyue, G. Song, W. Xie, Y. Peng, P. Xiyuan, Fragment anomaly detection with prediction and statistical analysis for satellite telemetry, in: IEEE Access, volume 5, 2017, pp. 19269–19281.
[11] K. Hundman, V. Constantinou, L. Christopher, I. Colwell, T. Soderstrom, Detecting spacecraft anomalies using LSTMs and nonparametric dynamic thresholding, in: Proc. 24th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, London, United Kingdom, 2018, pp. 387–395.
[12] S.K. Ibrahim, A. Ahmed, M.A. Zeidan, I.E. Ziedan, Machine learning methods for spacecraft telemetry mining, IEEE Trans. Aerosp. Electron.Syst. 55 (2019) 1816–1827.
[13] Y.Q. Zhu, A.K. Qin, P.N. Suganthan, G.B. Huang, Evolutionary extreme learning machine, Pattern Recognit. 38 (2005) 1759–1763.
[14] Y. Bazi, N. Alajlan, F. Melgani, H. AlHichri, S. Malek, R.R. Yager, Differential evolution extreme learning machine for the classification of hyperspectral images, IEEE Geosci. Remote Sens. Lett. 11 (6) (2014) 1066–1070.
[15] T. Matiasa, F. Souzaa, A. Rui, C.H. Antunes, Learning of a single-hidden layer feedforward neural network using an optimized extreme learning machine, Neurocomputing 129 (2014) 428–436.
[16] J.H. Cho, D.J. Lee, Parameter optimization of extreme learning machine using bacterial foraging algorithm, J. Fuzzy Logic Intell. Syst. 17 (2007) 807–812.
[17] D.N.G. Silva, L.D.S. Pacifico, T.B. Ludermir, An evolutionary extreme learning machine based on group search optimization, in: Proc. IEEE Congress of Evolutionary Computation (CEC), New Orleans, LA, USA, 2011, pp. 574–580.
[18] Y. Xu, S. Yang, Evolutionary extreme learning machine - based on particle swarm optimization, in: Proc. Third International Symposium on Neural Networks, Chengdu, China, 2006, pp. 644–652.
[19] S. Saraswathi, S. Sundaram, N. Sundararajan, M. Zimmermann, M.N. Hamilton, IC-GA-PSO-ELM Approach for accurate multiclass cancer classification resulting in reduced gene sets in which genes encoding secreted proteins are highly represented, IEEE/ACM Trans. Comput. Biol.Bioinform. 8 (2) (2011) 452–463.
[20] D.S.L. Pacifico, T.B. Ludermir, Evolutionary extreme learning machine based on particle swarm optimization and clustering strategies, in: Proc. The 2013 International Joint Conference on Neural Networks (IJCNN), Dallas, TX, USA, 2013.
[21] M. Chao, An efficient optimization method for extreme learning machine using artificial bee colony, J. Digital Inf. Manage. 15 (3) (2017) 135–147.
[22] T.V. Tran, Y.N. Wang, An evolutionary extreme learning machine based on chemical reaction optimization, J. Inf. Optim. Sci. 38 (8) (2017) 1265–1290.
[23] H. Liu, Z. Duan, Y. Li, H. Lu, A novel ensemble model of different mother wavelets for wind speed multistep forecasting, Appl. Energy 228 (2018) 1783–1800.
[24] Y. Hao, C. Tian, A novel two-stage forecasting model based on error factor and ensemble method for multi-step wind power forecasting, Appl. Energy 238 (2019) 368–383.
[25] H. Liu, H. Wu, Y. Li, Smart wind speed forecasting using EWT decomposition, GWO evolutionary optimization, RELM learning and IEWT reconstruction, Energy Convers. Manage. 161 (2018) 266–283.
[26] N. Muangkote, K. Sunat, S. Chiewchanwattana, An improved grey wolf optimizer for training q-gaussian radial basis functional-link nets, in: Proc. the 2014 International Conferenceon Computer Science and Engineering (ICSEC), Khon Kaen, Thailand, 2014.
[27] S. Mirjalili, How effective is the grey wolf optimizer in training multi-layer perceptrons, Appl. Intell. 43 (1) (2015) 150–161.
[28] S. Abdelghafar, A. Darwish, A.E. Hassanien, Cube satellite failure detection and recovery using optimized support vector machine, in: Proc. the International Conference on Advanced Intelligent Systems and Informatics, Cairo, Egypt, 2018, pp. 664–674.
[29] S. Mirjalili, A. Lewis, Grey wolf optimizer, Adv. Eng. Softw. 69 (2014) 46–61.
[30] E. Elhariri, N. El-Bendary, A.E. Hassanien, A. Abraham, Grey wolf optimization for one against one multi-class support vector machines, in: Proc. IEEE Soft Computing and Pattern Recognition (SoCPaR) conference, Fukuoka, Japan, 2015, pp. 7–12.
[31] G.B. Huang, Q.Y. Zhu, C.S. Kheong, Extreme learning machine: theory and applications, Neurocomputing 70 (2006) 489–501.
[32] B. Han, B. Hea, R. Nian, M. Mengmeng, S. Zhang, L. Minghui, A. Lendasse, LARSEN-ELM: selective ensemble of extreme learning machines using LARS for blended data, Neurocomputing 149 (2015) 285–294.
[33] M.B. Li, G.B. Huang, P. Saratchandran, N. Sundararajan, Fully complex extreme learning machine, Neurocomputing 68 (2005) 306–314.
[34] Y. Miche, A. Sorjamaa, P. Bas, O. Simula, C. Jutten, A. Lendasse, OP-ELM: Optimally pruned extreme learning machine, IEEE Trans. Neural Netw. 21 (2010) 158–162.
[35] B.H. Markus, S.L. Lancianese, M.B. Nagarajan, I.Z. Ikpot, A.L. Lerner, A. Wism, Prediction of biomechanical properties of trabecular bone in MR images with geometric features and support vector regression, IEEE Trans. Biomed. Eng. 58 (6) (2011) 1820–1826.
[36] H. Wang, Anomaly detection of network traffic based on prediction and self-adaptive threshold, Int. J. Future Gener.Commun. Netw. 8 (6) (2015) 205–214.
[37] B.E. Boser, I.M. Guyon, V.N. Vapnik, A training algorithm for optimal margin classifiers, in: Proc. the fifth annual workshop on Computational Learning Theory, Pittsburgh, Pennsylvania, USA, 1992, pp. 144–152.
[38] H. Drucker, C.J. Burges, L. Kaufman, A.J. Smola, V.N. Vapnik, Support vector regression machines, Adv. Neural Inf. Process. Syst. 9 (1997) 155–161.
[39] L. Ren, L. Zhao, S. Hong, S. Zhao, H. Wang, L. Zhang, Remaining useful life prediction for lithium-ion battery: a deep learning approach, IEEE Access 6 (2018) 50587–50598.
[40] K.V. Shihabudheen, M. Mahesh, G.N. Pillai, Particle swarm optimization based extreme learning neuro-fuzzy system for regression and classification, Expert Syst. Appl. 92 (2018) 474–484.
[41] R. Shailendra, J.P. Hyuk, Semi-supervised learning based distributed attack detection framework for iot, Appl. Soft Comput. 72 (2018) 79–89.
[42] M.R. Bonyadi, Z. Michalewicz, Analysis of stability, local convergence, and transformation sensitivity of a variant of the particle swarm optimization algorithm, IEEE Trans. Evol. Comput. 20 (3) (2016) 370–385.