Article

Using enhanced sparrow search algorithm-deep extreme learning machine model to forecast end-point phosphorus content of BOF

Lingxiang Quan[1], Ailian Li[1]*, Guimei Cui[1], Shaofeng Xie[2]

1 School of Information Engineering, Inner Mongolia University of Science and Technology, Baotou 014010, China; qlx301103@163.com (L.Q.); cguimei1@163.com (G.C.)
2 Department of Infrastructure, Inner Mongolia University of Science and Technology, Baotou 014010, China; 759124819@qq.com
* Correspondence: E-mail: xsfls854598@126.com

Abstract : An effective technology for predicting the end-point phosphorous content of basic oxygen furnace (BOF) can provide theoretical instruction to improve the quality of steel via controlling the hardness and toughness. Given the slightly inadequate prediction accuracy in the existing prediction model, a novel hybrid method was suggested to more accurately predict the end-point phosphorus content by integrating an enhanced sparrow search algorithm (ESSA) and a multi-strategy with a deep extreme learning machine (DELM) as ESSA-DELM in this study. To begin with, the input weights and hidden biases of DELM were randomly selected, resulting in that DELM inevitably had a set of non-optimal or unnecessary weights and biases. Therefore, the ESSA was used to optimize the DELM in this work. For the ESSA, the Trigonometric substitution mechanism and Cauchy mutation were introduced to avoid trapping in local optima and improve the global exploration capacity in SSA. Finally, to evaluate the prediction efficiency of ESSSA-DELM, the proposed model was tested on process data of the converter from the Baogang steel plant. The efficacy of ESSA-DELM was more superior to that of other DELM-based hybrid prediction models and conventional models. The result demonstrated that the hit rate of end-point phosphorus content within ±0.003%, ±0.002%, and ±0.001% was 91.67%, 83.33%, and 63.55%, respectively. The proposed ESSA-DELM model could possess better prediction accuracy compared with other models, which could guide field operations.

Keywords: End-point phosphorus content; Deep extreme learning machine; Sparrow search algorithm; Trigonometric substitution; Cauchy mutation

1. Introduction

In the basic oxygen furnace (BOF) steelmaking process, phosphorus in the steel is a harmful element for steel grades, reducing the toughness and plasticity of the steel and affecting the quality of the steel [1, 2]. Therefore, all-steel grades must remove the phosphorus content during the converter smelting process within the specified range of the steel grade. Hence, it is

 doi:10.20944/preprints202112.0192.v1

significant to accurately forecast and control the end-point phosphorus content of BOF [3].

Recently, some machine learning models have been used to predict the end-point phosphorus content of BOF. Among of them, the BP neural network (BPNN) [4] is the most widely used. Li et al. [5] developed an L-M algorithm based on a BPNN to predict the phosphorus content at the end of the converter steelmaking. He et al. [6] set up a principal component analysis (PCA)-BP model for end-point phosphorus content prediction. Zhu et al. [7] established a Prediction model of end-point phosphorus content for BOF based on monotone-constrained BPNN. The above hybrid BPNN prediction model improves the prediction accuracy to some extent compared with traditional BPNN, but it is still insufficient. The main reason is that the BPNN needs to set a large number of network training parameters, resulting in slow training speed performance and poor generalization ability.

Extreme learning machine (ELM) proposed by Huang et al. is a single-layer feedforward neural networks (SLFNs) [8, 9]. ELM possesses the advantages of fast solution speed, high accuracy, and simple parameter setting compared to BPNN. And ELM has been applied in many fields, such as wind power prediction, bearing fault diagnosis, and pattern clustering [10-12]. Nevertheless, the ELM is randomly generated due to its input weights and thresholds, and has only one hidden layer, resulting in the robustness of the model being poor.

Based on extreme learning machine-autoencoder (ELM-AE), deep extreme learning machine (DELM), named multi-layer extreme learning machine (ML-ELM) is proposed by Kasun et al. [13]. DELM does not need to be fine-tuned [14]. And it need less training time than deep learning. Furthermore, DELM shows excellent generalization performance like deep learning. So DELM is employed to construct a prediction model of the end-point phosphorus content of BOF in this study.

Nevertheless, the input weights and hidden biases of DELM are randomly generated. DELM inevitably acquires a set of non-optimal or unnecessary weights and biases [14]. In addition, DELM may be trapped in an overfitting problem in all training data as well. To tackle these problems of the DELM model for the improvement of the prediction capability, this paper intends to employ a new optimization algorithm in the literature. Recently, many optimization algorithms have been developed in various applications, like power Load Forecasting [15], life prediction of lithium batteries [16], brain tumor diagnosis [17], and polymer electrolyte fuel cell (PEMFC) stack [18].

Based on sparrow predation and anti-predation behavioral traits, sparrow search algorithm (SSA) is a new intelligent optimization method constructed by XUE [19]in 2020, which is established. Compared with the existing optimization algorithms, SSA has a better optimization ability and faster efficiency. However, similar to other algorithms, there are still some shortcomings in SSA, such as loss of population diversity in the later iteration and easily being stuck in local optimum and search stagnation. To mitigate the aforementioned issues and improve

 doi:10.20944/preprints202112.0192.v1

the global optimization ability, various improvement mechanisms have been presented. Liu et al. [20] adopted adaptive weight parameters to balance the search and exploitation capability of the sparrow search algorithm, improved the ability of SSA to get rid of stagnation with the help of the Cauchy-Gaussian mechanism. Yuan et al. [21] introduced a center of gravity inverse learning mechanism to initialize the population, added weight factors to update the follower positions in the sparrow algorithm to enhance the global exploration capability of SSA, and finally introduced a mutation strategy in the follower positions to increase the likelihood of SSA in escaping local extremes. Therefore, to furthermore overcome the problems associated with standard SSA, an Enhanced SSA (ESSA) is put forward based on Trigonometric substitution (TS) strategy and Cauchy mutation.

Considering the above mentioned, based on deep extreme learning machine and enhanced sparrow search algorithm (ESSA-DELM), a prediction model of end-point phosphorus content of BOF was proposed in this paper. The many input weights and biases of the DELM were optimized by ESSA. For ESSA, the Trigonometric substitution strategy and Cauchy mutation were adopted to solve the inadequacy of the SSA global optimization search. Finally, some DELM-based hybrid prediction models and conventional models were applied to validate the performance of ESSA-DELM, and the result proved that the performance of ESSA-DELM was significantly superior to that of the other models. The main contributions of this study are:

(a) TS and Cauchy mutation are applied to enhance the optimization capacity of SSA.

(b) The proposed ESSA is used to optimize the weights and biases randomly generated by DELM.

(c) The established ESSA-DELM will be used to predict the end-point phosphorus content of BOF.

The rest of the study is organized as follows. A brief description of the ELM model, DELM model, and optimization algorithms is given in Section 2. Section 3 evaluates the Enhanced Sparrow Search Algorithm (ESSA) and validates the proposed ESSA-DELM prediction model of end-point phosphorus in terms of performance. The summary and scope for future study are shown in Section 4.

2. Methodology and model development

2.1. Extreme learning machine (ELM)

ELM is a type of single hidden layer feed-forward network (SLFN) [8]. As shown in Figure 1, given training dataset $\{(x_i, y_i) | \ x_i \in R^g, y_i \in R^c, i = 1, 2, \ldots, N\}$, where N signifies the number of samples, g and c signifies the dimension number of input vector x and output vector y respectively, the output of ELM with l hidden neurons can be described as:

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 December 2021 doi:10.20944/preprints202112.0192.v1

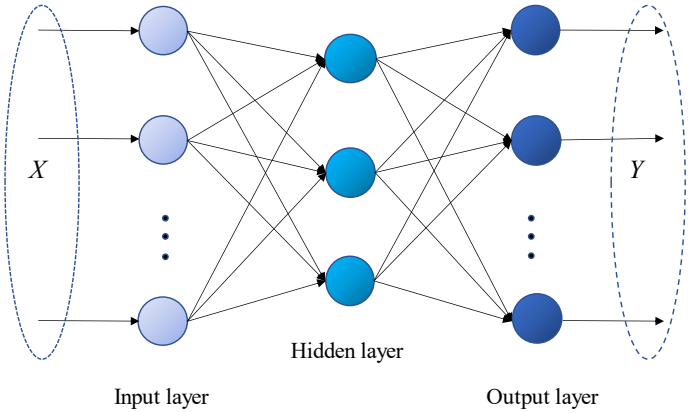

Figure 1. ELM structure diagram

$$y_j = \sum_{i=1}^{l} \beta_i G\left(w_i \cdot x_j + v_i\right), \quad j = 1, 2, \cdots N \tag{1}$$

where v_i is the threshold; $w_i = \left[w_{il}, w_{i2}, \cdots, w_{ig}\right]^T$ describes the input weight vector; $\beta_i = \left[\beta_{il}, \beta_{i2}, \cdots, \beta_{ic}\right]^T$ represents output weight vector; G signifies the activation function; $y_j = \left[y_{jl}, y_{j2}, \cdots, y_{jc}\right]^T$ denotes the output vector. Eq. (1) can be expressed succinctly as

$$H\beta = Y \tag{2}$$

where $H\left(w_i, v_i, x_j\right) = \begin{bmatrix} G\left(w_1 \cdot x_1 + v_1\right) & \cdots & G\left(w_L \cdot x_1 + v_L\right) \\ \vdots & \ddots & \vdots \\ G\left(w_1 \cdot x_N + v_1\right) & \cdots & G\left(w_L \cdot x_N + v_L\right) \end{bmatrix}_{N \times L}$

$$\beta = \begin{bmatrix} \beta_1^T \\ \vdots \\ \beta_L^T \end{bmatrix}_{L \times c} \qquad Y = \begin{bmatrix} y_1^T \\ \vdots \\ y_N^T \end{bmatrix}_{N \times c}$$

where H signifies the hidden layer output matrix. The output weight matrix β of the ELM model can be calculated by a generalized inverse matrix of the hidden layer.

$$\beta = H^{+}Y \tag{3}$$

where H^{+} describes Moore-Penrose generalized inverse matrix.

2.2. Deep Extreme learning machine (DELM)

2.2.1. Extreme learning machine-autoencoder algorithm (ELM-AE)

The ELM-AE [19] is a neural network that can both reproduce the input data and autoencoder, established by Kasun et al. [9], possessing the characteristics of fast computation and high accuracy rate as well as ELM. Furthermore, similar to the ELM, ELM-AE contains an input layer, a single hidden layer, and an output layer, and the major difference is that the output layer of ELM-AE and the input layer are identical [22].

Given there are N distinctive data sets $x_i \in R_g \times R_j, (i = 1, 2, \cdots, N)$, where n defines the number of the

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 December 2021 doi:10.20944/preprints202112.0192.v1

features. The outputs of the hidden layer for ELM-AE can be represented as:

$$h = g(ax + b) \tag{4}$$

where $a^T a = I, b^T b = 1$, the mathematical relationship between the outputs of the hidden layer and the outputs of the output layer can be expressed as:

$$h(x_i)\beta = x_i^T, i = 1, 2, \cdots, N \tag{5}$$

where β represents the output weight of the output layer.

2.2.2. Deep Extreme learning machine (DELM)

DELM introduces ELM-AE to train the parameters of all the hidden layers. Simultaneously the hidden layer activation functions of DELM can be linear or nonlinear piecewise [14]. When the number of nodes in the k th hidden layer is equivalent to the number of the $k-1$ th hidden layer, it could be concluded that the activation function $g(x)$ remains linear, else $g(x)$ should be nonlinear piecewise. So the output of the k th hidden layer may be expressed as follows:

$$H_k = g\left(\left(\beta^k\right)^T H_{k-1}\right) \tag{6}$$

where H_k signifies the output matrix of DELM k th hidden layer (when $k-1=0$, this layer denotes the input layer, and H_k describes the input of DELM)

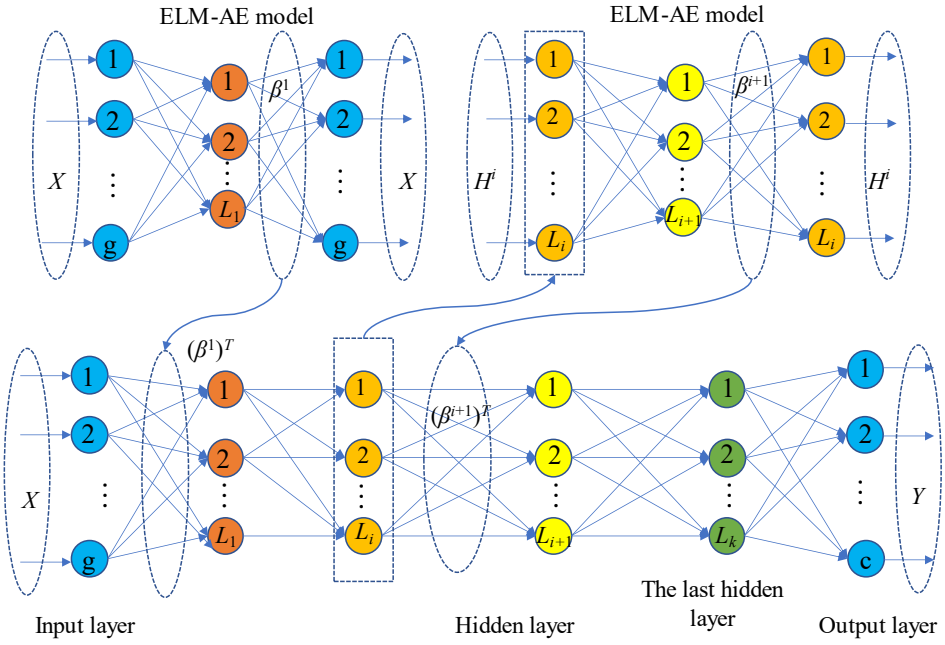

Figure 2. Structure of DELM.

2.3. Sparrow search algorithm (SSA)

Sparrows colony primarily consists of discoverers and followers during the foraging process. The discoverer provides the searching zone and direction for the sparrow colony because of its better fitness value, whereas the follower exploits the

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 December 2021 doi:10.20944/preprints202112.0192.v1

location of the discoverer to obtain food [19]. When a sparrow colony perceives threat and the alarm value is higher than the security value, it will engage in anti-predatory behavior. SSA can be simply abstracted into a discoverer-followers-early warning model.

$$X_{id}^{t+1} = \begin{cases} X_{id}^{t} \cdot \exp\left(\dfrac{-i}{\alpha \cdot T}\right) & R_2 < ST \\ X_{id}^{t} + Q \cdot L & R_2 \geq ST \end{cases} \tag{7}$$

where t means the current iteration; T denotes the maximum number of iterations; x_{id}^{t} shows the current position information of the i th sparrow in the d th dimension; α describes a random number in $[0,1)$; Q is a random number obeying normal distribution; L signifies a $1 \times d$ matrix with all parameters being 1; $R_2 \in [0,1]$ and $ST \in [0.5,1]$ represents the alarm value and security threshold, respectively. When $R_2 < ST$, It denotes a secure hunting environment. If $R_2 \geq ST$, it indicates that some sparrows have detected the predator around them and all sparrows must migrate to safer areas as soon as possible [19].

All sparrows are followers except all discoverers in population. Followers update their position and describe as follows:

$$X_{i,j}^{t+1} = \begin{cases} Q \cdot \exp\left(\dfrac{X_{\text{worst}}^{t} - X_{id}^{t}}{i^2}\right) & i > n/2 \\ X_{p}^{t+1} + \left|X_{id}^{t} - X_{p}^{t+1}\right| \cdot A^{+} \cdot L & \text{other} \end{cases} \tag{8}$$

where X_{worst} signifies the worst global position; X_p denotes the optimal position occupied by the discoverer; A is a $1 \times$ d matrix with random elements of 1 or -1, and $A^{+} = A^{T}(AA^{T})^{-1}$. When $i > n/2$, it means that the i th followers with low fitness need to fly to other locations for food because it is in a state of hungry.

Given the presence of predator, 10% ~ 20% of sparrow colonies are responsible for scouting and warning, with the location updated as follows:

$$X_{i,j}^{t+1} = \begin{cases} X_{best}^{t} + \beta \cdot \left|X_{id}^{t} - X_{best}^{t}\right| & f_i > f_g \\ X_{id}^{t} + k \cdot \left(\dfrac{\left| X_{id}^{t} - X_{wort}^{t} \right|}{(f_i - f_w) + \varepsilon}\right) & f_i = f_g \end{cases} \tag{9}$$

where β is the step correction parameter, which obeys the standard normal distribution; X_{best} signifies the current global best position; $K \in [-1,1]$ describes a uniform random number; e is the smallest constant to avoid a zero denominator; f_i denotes the fitness value of the current sparrow; f_g and f_w are the current global best fitness and worst fitness, respectively. ε denotes a minimal constant, this study takes $\varepsilon = 10\text{e-}50$.

2.4. Enhanced Sparrow Search Algorithm

To overcome the convergence stagnation and being trapped

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 December 2021 doi:10.20944/preprints202112.0192.v1

into local convergence of standard SSA, this paper presents Enhanced Sparrow Search Algorithm (ESSA), incorporating Trigonometric substitution strategy (TS) and Cauchy mutation strategy. On the one hand, TS is introduced to balance the development and exploration ability of SSA, and additionally, a step search factor and a position inertia factor are introduced to further strengthen the seeking ability of algorithm. On the other hand, the sparrow individuals are perturbed by using the Cauchy mutation to enhance the global searching capability of SSA. The flowchart of the proposed ESSA is shown in Figure 3.

2.4.1 Trigonometric substitution strategy (TS)

The TS strategy mainly uses the sine function [23] to update the current position of discoverers, which makes the sparrow position change continuously to improve the exploitation and exploration ability of SSA, thus improving the global searching capability. The update formula of discoverers is as follow:

$$X_{id}^{t+1} = X_{id}^{t} \times \left|\sin R_1\right| + R_3 \times \sin R_1 \times \left|x_1 \times X_p^{t+1} - x_2 \times X_{id}^{t}\right| \quad R_2 < ST \tag{10}$$

where $R_1 \in [0, 2\pi]$ is a random number, which determines the moving distance of i th individual at the d th iteration; $R_3 \in [0, \pi]$ represents a random number, which controls the moving direction of the i th individual at the iteration; $x_1 = -\pi + (1-\tau) \times 2\pi$ and $x_2 = -\pi + \tau \times 2\pi$ are coefficients obtained from the gold partition number $\tau = (\sqrt{5}-1)/2$, which narrow the search space of the algorithm and guide individuals to gradually converge to the optimal value.

To further balance the global exploring and exploiting, the new update of R_3 with the following nonlinear adaptive changes is as follow [24]:

$$R_3 = \frac{1}{(t+1)\hat{}(\xi+\eta)\hat{}3} * \left(\frac{\xi^* t\hat{}\xi}{T\hat{}\xi} + \frac{\eta^* t\hat{}\eta}{T\hat{}\eta} \right) \tag{11}$$

where ξ and η are weight coefficients, and $\xi = 3$, $\eta = 5$. Eq. (11) shows that inertia weights are negatively correlated with the number of iterations. In the later stage of the search, a smaller inertia weight can be used to seek the optimal value within a narrower area, thus accelerating the convergence speed.

Considering that the population individual position update is affected by the current position during the whole search process, a nonlinear position inertia factor w is introduced to further improve the searching ability of sparrows, which shows a positive correlation with the iteration in Eq. (12). A smaller w can lessen the impact of individual position updates on the current solution position and improve global searching ability in the early stages. Besides, the greater w can take advantage of the strong dependence of current location information on individual location updates in the later iterations of the algorithm, speeding up convergence. The position update of discoveries is presented in Eq. (13):

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 December 2021 doi:10.20944/preprints202112.0192.v1

$$w = \frac{e^{\frac{t}{T}} - 1}{e - 1} \tag{12}$$

$$X_{id}^{t+1} = w \times X_{id}^{t} \times \left|\sin R_1\right| + R_3 \times \sin R_1 \times \left|x_1 \times X_p^{t+1} - x_2 \times X_{id}^{t}\right| \quad R_2 < ST \tag{13}$$

2.4.2. Cauchy mutation strategy

The description of Cauchy density function as Eq. (14):

$$f(x) = \frac{1}{\pi} \frac{t}{x^2 + t^2}, \quad t > 0, -\infty < x < \infty \tag{14}$$

where t describes a proportional parameter. Its distribution function is:

$$F(x) = \frac{1}{2} + \frac{1}{\pi} \arctan\left(\frac{x}{t}\right) \tag{15}$$

The Cauchy distribution, like the normal distribution, is a continuous type of probability distribution with a small magnitude at coordinate 0. It can form a large perturbation because the bilateral show a flat and long posture and converge to 0 with a slow speed. Since the Cauchy mutation originates from the Cauchy distribution, introducing the Cauchy mutation into the sparrow individual position update will generate a large perturbation, which will expand the scope of the algorithm to obtain the best solution, and then move away from the local optimum. Eq. (16) presents the position update of followers.

$$X_{i,j}^{t+1} = X_{best}^{t} + gamrnd() \bullet C \bullet X_{best}^{t} \quad \text{other} \tag{16}$$

where C represents a Cauchy distributes random number; $gamrnd()$ is a gamma random number, which can further enhance the abruptness of the Cauchy mutation mainly by the jumpiness of the random number selection.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 December 2021 doi:10.20944/preprints202112.0192.v1

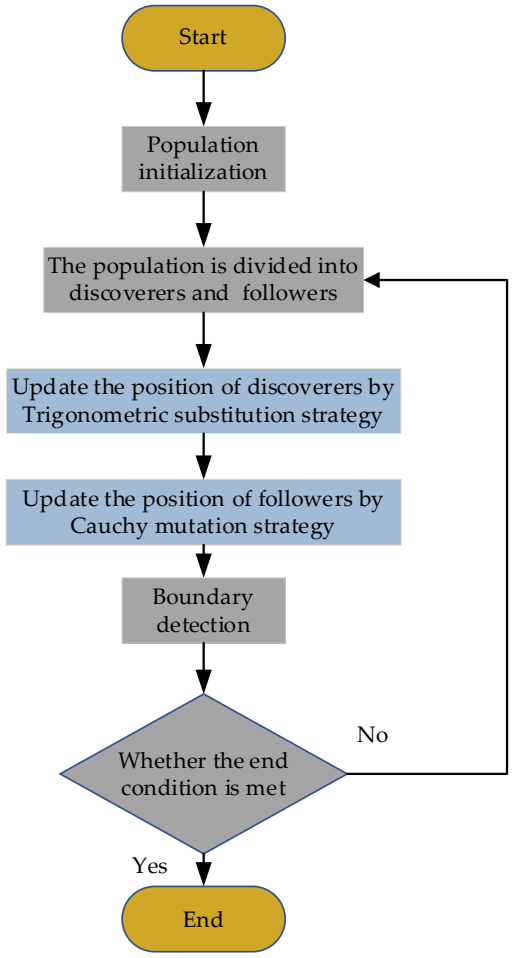

Figure 3. The flowchart of ESSA.

3 Results and discussion

3.1 The evaluation of ESSA

To evaluate the performance of ESSA, the proposed ESSA was compared with four traditional intelligent algorithms, including butterfly optimization algorithm(BOA)[25], sine cosine algorithm(SCA)[26], whale optimization algorithm(WOA)[27], sparrow search algorithm(SSA), and five advanced intelligent algorithms, like improved grey wolf optimizer (IGWO)[28], leader slime mould algorithm (LSMA)[29], leader Harris hawks optimization (LHHO)[30], adaptive opposition slime mould algorithm (AOSMA)[31], hybrid butterfly optimization algorithm with particle swarm optimization (HPSOBOA)[32]. Table 1 signifies the parameters settings for 10 algorithms. In addition, 9 classic test functions including 4 unimodal functions F1~F4 and 5 multimodal functions F5~F9 were employed to evaluate the properties of several algorithms (Table 2) [19,33]. To ensure the fairness of the experiment, the initial population size (uniformly randomly generated) was set to 50, the dimension of the solution space to be 30, and the maximum evaluation times to be 500. And each algorithm still runs 30 times independently in each classical test function.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 December 2021 doi:10.20944/preprints202112.0192.v1

Table 1. Parameter setting for compared algorithms.

Algorithm	Parameters
BOA	a = 0.1, $c(0)$ = 0.01, p = 0.6
SCA	a=2
WOA	a_1=[2 0]; a_2=[-2 -1]; b = 1
SSA	ST=0.8; Proportion of discoverers PD=0.2; proportion of scouter SD=0.2
IGWO	a was linearly decreased from 2 to 0
LSMA	Z=0.03
LHHO	β=1.5
AOSMA	δ=0:03
HPSOBOA	a_{first} = 0.1, a_{final} = 0.3, $c(0)$ = 0.01, p = 0.6, $x(0)$ = 0.315, ρ = 0.295, c_1 = c_2 = 0.5
ESSA	ST=0.8; Proportion of discoverers PD=0.2; proportion of scouter SD=0.2; $\xi = 3$; $\eta = 5$

Table 2. Description of nine benchmark functions.

Function	Search range	F_{min}
$F_1(x)=\sum_{i=1}^{n}x_i^2$	[-100,100]	0
$F_2(x)=\sum_{i=1}^{n}\lvert x_i\rvert+\prod_{i=1}^{n}\lvert x_i\rvert$	[-10,10]	0
$F_3(x)=\sum_{i=1}^{n}\left(\sum_{j-1}^{i}x_j\right)^2$	[-100,100]	0
$F_4(x)=\max_i\left\{\lvert x_i\rvert,1\le i\le n\right\}$	[-100,100]	0
$F_5(x)=\sum_{i=1}^{n}\left[x_i^2-10\cos(2\pi x_i)+10\right]$	[-5.12,5.12]	0
$F_6(x)=\sum_{i=1}^{n}\left(y_i^2-10\cos(2\pi y_i)+10\right)$ $y_i=\begin{cases}x_i & \lvert x_i\rvert<\frac{1}{2}\\ \frac{\text{round}\,(2x_i)}{2} & \lvert x_i\rvert\ge\frac{1}{2}\end{cases}$ for $i=1,2,\ldots,D$	[-5.12,5.12]	0
$F_7(x)=-20\exp\left(-0.2\sqrt{1/n\sum_{i=1}^{n}x_i^2}\right)-\exp\left(1/n\sum_{i=1}^{n}\cos(2\pi x_i)\right)+20+e$	[-32,32]	0
$F_8(x)=1/4000\sum_{i=1}^{n}x_i^2-\prod_{i=1}^{n}\cos\left(x_i/\sqrt{i}\right)+1$	[-600,600]	0
$F_9(x)=\sum_{i=1}^{n}\lvert x_i\sin(x_i)+0.1x_i\rvert$	[-10,10]	0

With the purpose of more visually illustrating the convergence of ESSA on different types of test functions, nine convergence trend graphs of SSA were presented in Figure 4. For better observation, the ordinate was the logarithm of base 10. When the curve is no longer shown with the increase of iteration number, it means that the algorithm has obtained the theoretical

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 December 2021 doi:10.20944/preprints202112.0192.v1

optimal solution 0.

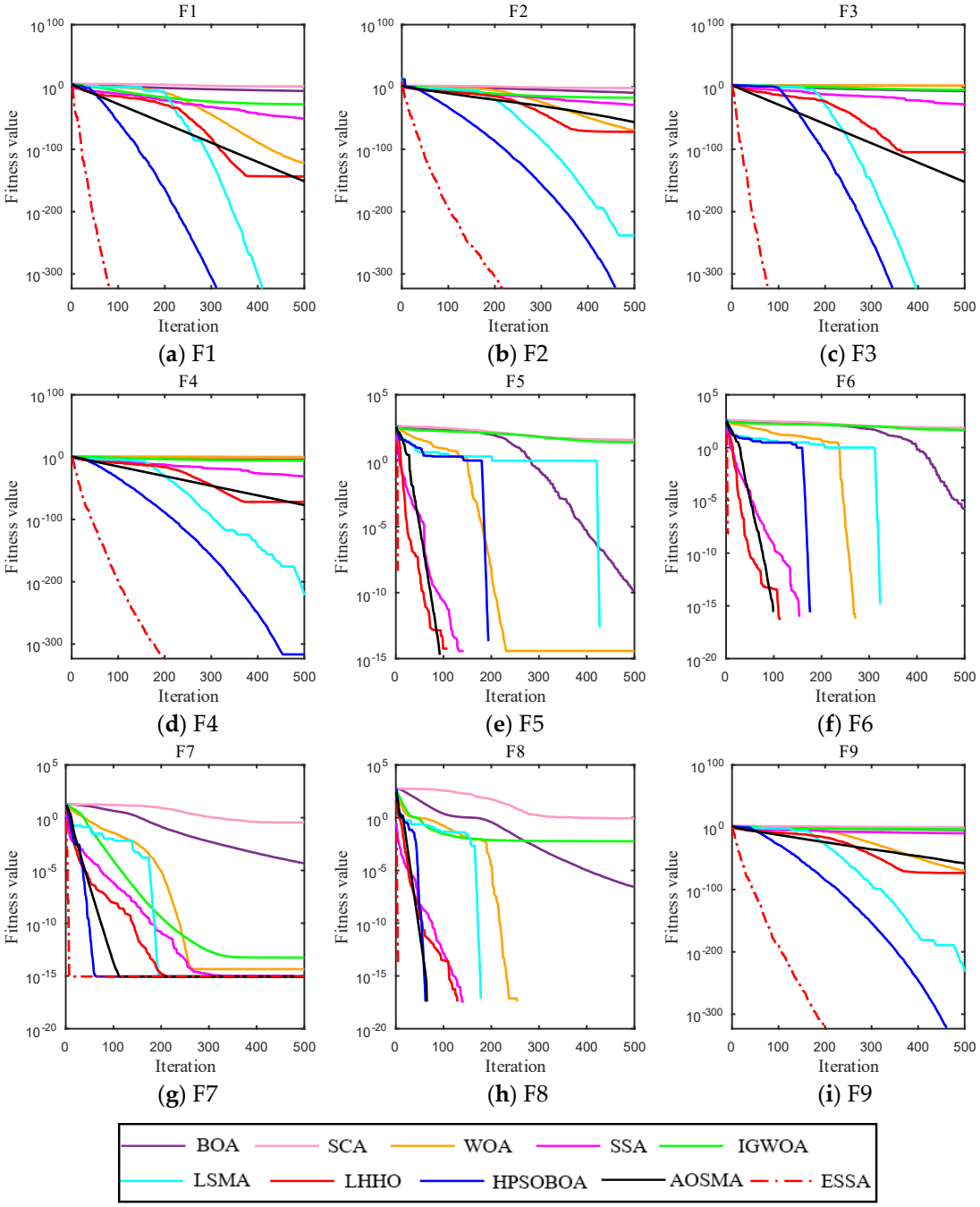

Figure 4. Convergence curves of nine algorithms (**a**) F1; (**b**) F2; (**c**) F3; (**d**) F4; (**e**) F5; (**f**) F6; (**g**) F7; (**h**) F8; (**i**) F9.

Figure 4 shows the convergence curves of the ESSA and other algorithms on the 9 classic test functions F1-F9. In dealing with unimodal function F1-F4, it is clear that ESSA has better convergence in the early stages of the iteration. On the other hand, some traditional algorithms fall into the local optimum solution too early at the beginning of the iteration like SSA, BOA, SCA, WOA. In addition, the ESSA has high fluctuations in the initial iterations when processing multimodal functions F5, F6, F8, F9, and all obtain a theoretical optimum (0). For F7, although the

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 December 2021 doi:10.20944/preprints202112.0192.v1

results of the four algorithms are similar, ESSA shows a fast convergence speed compared to SSA, LSMA, LHHO, AOSMA, HPSOBOA. To summarize, the ESSA algorithm proposed in this paper outperforms other algorithms in terms of convergence speed and accuracy.

To further illustrate the effectiveness of ESSA, two criteria were taken for comparison: Average value (Ave) and Standard deviation (Std). Table 3 gives the comparative results between the ESSA algorithm and other algorithms in three dimensions.

Table 3. Comparisons of ESSA and other algorithms on 9 test functions in three dimensions.

		Dim=30		Dim=50		Dim=100	
Fun	Algorithm	Ave	Std	Ave	Std	Ave	Std
	BOA	3.29E-07	7.93E-08	4.28E-07	1.26E-07	4.36E-07	1.12E-07
	SCA	7.87E+00	1.48E+01	8.41E+02	9.70E+02	1.12E+04	7.74E+03
	WOA	1.07E-121	5.88E-121	1.45E-121	7.78E-121	3.71E-119	2.02E-118
	SSA	1.37E-60	7.45E-60	9.25E-57	5.07E-56	1.73E-50	9.46E-50
F1	IGWO	2.86E-28	9.79E-28	2.52E-20	2.13E-20	2.24E-12	1.32E-12
	LSMA	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
	LHHO	2.98E-147	1.16E-146	2.63E-142	1.39E-141	5.49E-137	3.01E-136
	AOSMA	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
	HPSOBOA	3.82E-152	9.02E-153	3.75E-152	1.00E-152	3.57E-152	1.57E-152
	ESSA	**0.00E+00**	**0.00E+00**	**0.00E+00**	**0.00E+00**	**0.00E+00**	**0.00E+00**
	BOA	2.68E-10	4.27E-10	3.86E+23	1.59E+24	1.93E+49	6.36E+49
	SCA	2.25E-02	2.39E-02	5.42E-01	5.66E-01	7.98E+00	7.06E+00
	WOA	2.55E-69	1.14E-68	1.12E-66	5.63E-66	4.78E-68	1.65E-67
	SSA	2.56E-31	1.15E-30	2.23E-29	1.21E-28	8.10E-29	4.19E-28
F2	IGWO	6.06E-18	4.74E-18	4.95E-13	2.37E-13	1.42E-08	5.62E-09
	LSMA	1.52E-261	0.00E+00	4.23E-287	0.00E+00	3.39E-241	0.00E+00
	LHHO	7.03E-77	1.58E-76	5.60E-73	3.07E-72	4.01E-75	1.77E-74
	AOSMA	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
	HPSOBOA	2.46E-59	1.19E-58	5.06E-52	1.79E-51	1.54E+36	6.00E+20
	FESSA	**0.00E+00**	**0.00E+00**	**0.00E+00**	**0.00E+00**	**0.00E+00**	**0.00E+00**
	BOA	1.95E-07	5.08E-08	2.18E-07	5.09E-08	2.72E-07	7.92E-08
	SCA	7.94E+01	4.26E+01	5.04E+02	1.64E+02	2.56E+03	6.82E+02
	WOA	1.28E+02	9.89E+01	5.08E+02	2.92E+02	2.44E+03	1.63E+03
	SSA	3.71E-24	2.03E-23	2.85E-23	1.42E-22	7.96E-27	4.11E-26
F3	IGWO	6.55E-06	1.21E-05	1.25E-01	1.36E-01	5.21E+01	2.88E+01
	LSMA	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
	LHHO	1.28E-103	6.98E-103	2.33E-97	1.28E-96	7.17E-86	3.92E-85
	AOSMA	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
	HPSOBOA	4.02E-153	1.19E-153	7.22E-153	6.77E-153	1.54E-152	1.57E-152
	ESSA	**0.00E+00**	**0.00E+00**	**0.00E+00**	**0.00E+00**	**0.00E+00**	**0.00E+00**
	BOA	4.28E-05	6.65E-06	4.60E-05	9.44E-06	5.28E-05	8.93E-06
	SCA	3.77E+00	1.37E+00	6.71E+00	8.54E-01	9.09E+00	1.92E-01
	WOA	3.06E-01	7.19E-01	3.59E-01	6.60E-01	1.33E+00	1.56E+00
	SSA	1.15E-29	4.84E-29	1.46E-29	8.02E-29	1.09E-26	5.97E-26
F4	IGWO	2.33E-06	4.18E-06	2.16E-03	2.85E-03	4.82E-01	2.98E-01
	LSMA	2.98E-212	0.00E+00	9.53E-205	0.00E+00	1.91E-194	0.00E+00
	LHHO	1.26E-70	6.93E-70	4.09E-71	2.24E-70	4.41E-73	2.23E-72
	AOSMA	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
	HPSOBOA	1.04E-77	6.28E-79	1.03E-77	6.24E-79	1.07E-77	6.14E-79
	ESSA	**0.00E+00**	**0.00E+00**	**0.00E+00**	**0.00E+00**	**0.00E+00**	**0.00E+00**
	BOA	3.64E-10	1.11E-09	2.24E-10	5.75E-10	1.04E-08	2.67E-08
	SCA	3.77E+01	3.74E+01	1.18E+02	7.14E+01	2.43E+02	1.29E+02

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 December 2021 doi:10.20944/preprints202112.0192.v1

	WOA	3.79E-15	1.44E-14	0.00E+00	0.00E+00	7.58E-15	2.88E-14
	SSA	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
F5	IGWO	2.25E+01	9.70E+00	5.63E+01	3.49E+01	1.52E+02	1.17E+02
	LSMA	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
	LHHO	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
	AOSMA	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
	HPSOBOA	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
	ESSA	**0.00E+00**	**0.00E+00**	**0.00E+00**	**0.00E+00**	**0.00E+00**	**0.00E+00**
	BOA	8.45E-04	3.35E-03	4.97E-08	2.68E-07	2.30E-10	6.26E-10
	SCA	6.33E+01	3.59E+01	2.12E+02	6.20E+01	4.61E+02	1.37E+02
	WOA	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
	SSA	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
F6	IGWO	5.75E+01	3.85E+01	6.91E+01	4.55E+01	2.60E+02	7.96E+01
	LSMA	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
	LHHO	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
	AOSMA	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
	HPSOBOA	**0.00E+00**	**0.00E+00**	**0.00E+00**	**0.00E+00**	**0.00E+00**	**0.00E+00**
	ESSA	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
	BOA	4.50E-05	8.19E-06	5.27E-05	1.16E-05	5.11E-05	8.94E-06
	SCA	6.17E-01	1.05E+00	3.38E+00	1.62E+00	7.69E+00	1.91E+00
	WOA	4.20E-15	2.63E-15	4.80E-15	3.00E-15	3.97E-15	2.59E-15
	SSA	8.88E-16	0.00E+00	8.88E-16	0.00E+00	8.88E-16	0.00E+00
F7	IGWO	5.61E-14	8.84E-15	1.78E-11	1.01E-11	1.01E-07	3.47E-08
	LSMA	8.88E-16	0.00E+00	8.88E-16	0.00E+00	8.88E-16	0.00E+00
	LHHO	8.88E-16	0.00E+00	8.88E-16	0.00E+00	8.88E-16	0.00E+00
	AOSMA	8.88E-16	0.00E+00	8.88E-16	0.00E+00	8.88E-16	0.00E+00
	HPSOBOA	8.88E-16	0.00E+00	8.88E-16	0.00E+00	8.88E-16	0.00E+00
	ESSA	**8.88E-16**	**0.00E+00**	**8.88E-16**	**0.00E+00**	**8.88E-16**	**0.00E+00**
	BOA	2.03E-07	1.32E-07	4.81E-07	2.80E-07	7.30E-07	2.84E-07
	SCA	1.09E+00	9.09E-01	1.08E+01	1.04E+01	9.04E+01	4.65E+01
	WOA	0.00E+00	0.00E+00	0.00E+00	0.00E+00	3.70E-18	2.03E-17
	SSA	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
F8	IGWO	2.55E-03	5.08E-03	6.75E-03	8.91E-03	3.58E-03	6.48E-03
	LSMA	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
	LHHO	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
	AOSMA	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
	HPSOBOA	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00	0.00E+00
	ESSA	**0.00E+00**	**0.00E+00**	**0.00E+00**	**0.00E+00**	**0.00E+00**	**0.00E+00**
	BOA	3.15E-05	1.39E-05	4.44E-05	9.45E-06	5.06E-05	8.89E-06
	SCA	1.27E+00	3.17E+00	4.92E+00	3.58E+00	2.86E+01	1.56E+01
	WOA	3.99E-71	8.38E-71	1.66E-69	5.27E-69	1.93E-69	5.10E-69
	SSA	1.57E-08	4.11E-08	1.39E-06	3.76E-06	8.06E-06	3.86E-05
F9	IGWO	3.76E-04	3.04E-04	1.17E-03	1.11E-03	4.22E-03	2.81E-03
	LSMA	1.36E-254	0.00E+00	2.66E-225	0.00E+00	3.57E-252	0.00E+00
	LHHO	3.41E-75	1.76E-74	6.02E-75	3.17E-74	5.78E-75	2.40E-74
	AOSMA	0.00E+00	0.00E+00	1.78E-278	0.00E+00	0.00E+00	0.00E+00
	HPSOBOA	1.31E-58	5.49E-58	3.44E-61	9.10E-61	4.61E-57	2.36E-56
	ESSA	**0.00E+00**	**0.00E+00**	**0.00E+00**	**0.00E+00**	**0.00E+00**	**0.00E+00**

The results on nine classic benchmark functions (F1-F9) in different dimensions are listed in Table 3. It could be observed that in the unimodal test function (F1-F4) and multimodal test function (F5-F6), the proposed ESSA achieves a theoretical optimum (0) except for F8 and minimum standard deviation (0) for all dimensions, suggesting that ESSA has strong ability on search accuracy and robustness compared with BOA, SCA, WOA,

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 December 2021 doi:10.20944/preprints202112.0192.v1

SSA, IGWO, LSMA, LHHO, HPSOBOA. Whereas, AOSMA is equivalent to ESSA in terms of convergence accuracy and stability on nine classic test functions. However, as shown in Figure 4, it could be undoubtedly observed that the convergence speed of ESSA is faster than that of AOSMA, showing excellent competitiveness with advanced algorithms.

From Table 3 and Figure 4, it could be deduced that the ESSA has better convergence capability and robustness compared with the other nine algorithms. The primary reason is that the TS mechanism introduces sine change to equilibrate the exploitation and exploration of the SSA, and further enhances the convergence speed of the algorithm by two nonlinear weights. In addition, the Cauchy mutation strategy helps the sparrow individuals with the current best fitness to further improve the global optimization capability of SSA. According to the above analysis, the performance of the ESSA algorithm performs best, and it was chosen for the following experiments.

3.2 Prediction model of end-point phosphorus content based on ESSA-DELM

With the intention of testing the performance of ESSA-DELM, this study adopted the converter production data sets of Baogang steel plants to conduct experiments. The reactions that occurred in the converter are very complex, and end-point phosphorus content is affected by numerous influential factors. Therefore, the 10 variables as shown in Table 4 were selected as inputs of the model by incorporating scholarly research [7, 34] and SPSS data analysis.

Table 4. Input variables for the model.

Process parameter	Units	Process parameter	Units
Silicon content in hot metal	%	Oxygen supply time	min
Manganese Content in hot metal	%	Oxygen consumption	N m^3
Phosphorus content in hot metal	%	Quicklime addition	kg
Sulfur content in hot metal	%	Dolomite addition	kg
Hot metal temperature	℃	Iron ball	kg

To validate that the prediction model based on ESSA-DELM has a better prediction accuracy and generalization performance, this paper introduced the BPNN model, ELM model, DELM model, BOA-DELM model, WOA-DELM model, SCA-DELM model, and SSA-DELM model to establish a prediction model of end-point phosphorus content, where 200 heat date were collected for the experiment including 140 heat for train data and 60 for predict data. Through the experimental test, the hidden layer node of BPNN and ELM was set to 5 with one hidden layer, the hidden layer node of DELM, BAO-DELM, WOA-DELM, SCA-DELM, SSA-DELM, and ESSA-DELM were all set to 5 with ten hidden layers. The comparison of predicted value and actual value of end-point phosphorus content under 8 models are illustrated in Figure 4.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 December 2021 doi:10.20944/preprints202112.0192.v1

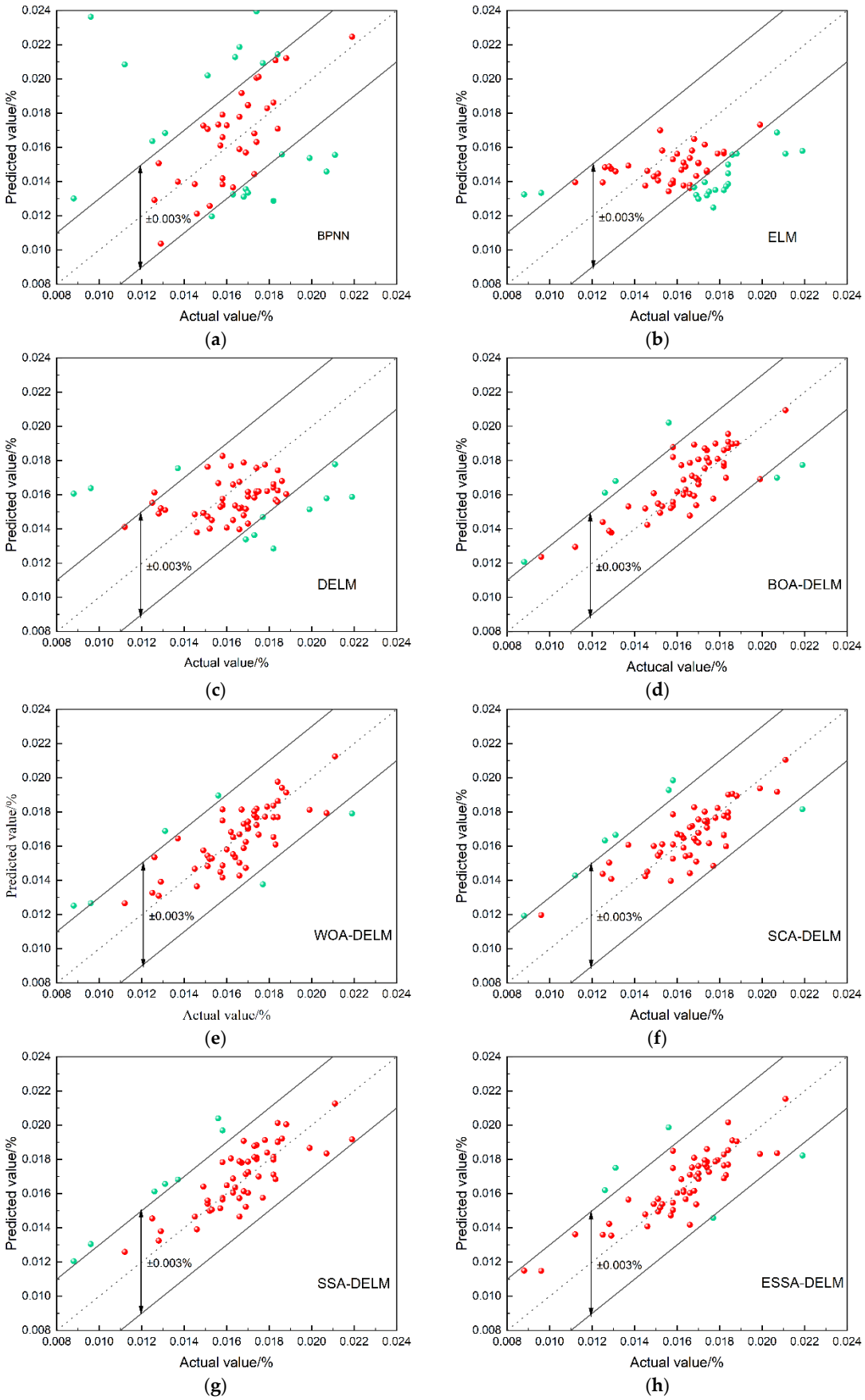

Figure 5. Comparison of end-point phosphorus content between prediction value and actual value of eight models. (**a**) BPNN model; (**b**) ELM model; (**c**) DELM model; (**d**) BOA-DELM model; (**e**) WOA-DELM

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 December 2021 doi:10.20944/preprints202112.0192.v1

model; (**f**) SCA-DELM; (**g**) SSA-DELM; (**h**) ESSA-DELM.

From the scatter plot in Figure 5, the proposed ESSA-DELM is performed closer to the ideal line (y=x) between the actual and predicted values, and with fewer points outside of the error range from -0.003% to 0.003%. In addition, the prediction curves of the ESSA-DELM models follow the actual value better than the other seven models, as intuitively presented in Figure 6. And it is also observed that the ESSA-DELM model presents the minimum prediction errors as shown in Figure 7. Besides the prediction accuracy, namely the hit rate, is represented by the distribution ratio of the difference between the true value and predicted value in different error ranges. The hit rates of the eight models in error ranges [-0.001%,0.001%], [-0.002%,0.002%], and [–0.003%, 0.003%] are shown in Table 5 and Figure 8.

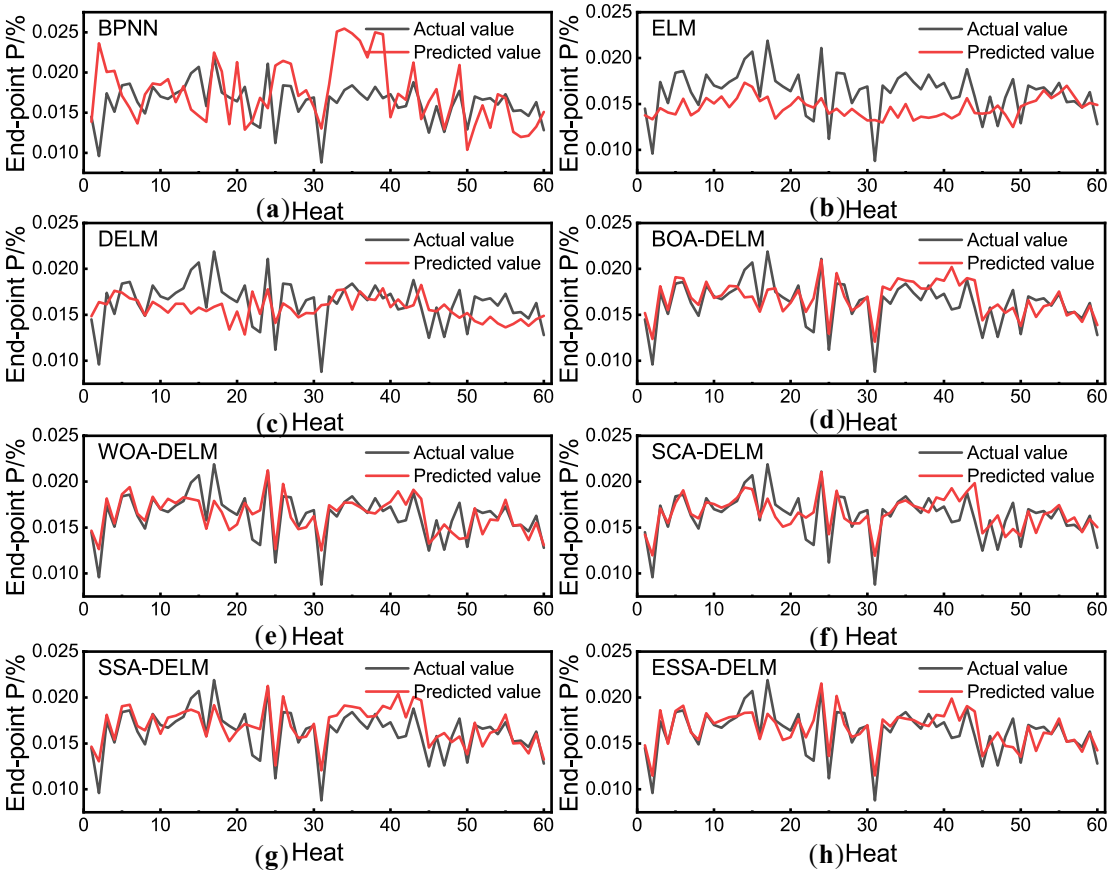

Figure 6. End-point phosphorus content prediction in the testing phase by different models (**a**) BPNN model; (**b**) ELM model; (**c**) DELM model; (**d**) BOA-DELM model; (**e**) WOA-DELM model; (**f**) SCA-DELM model; (**g**) SSA-DELM model; (**h**) ESSA-DELM model.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 December 2021 doi:10.20944/preprints202112.0192.v1

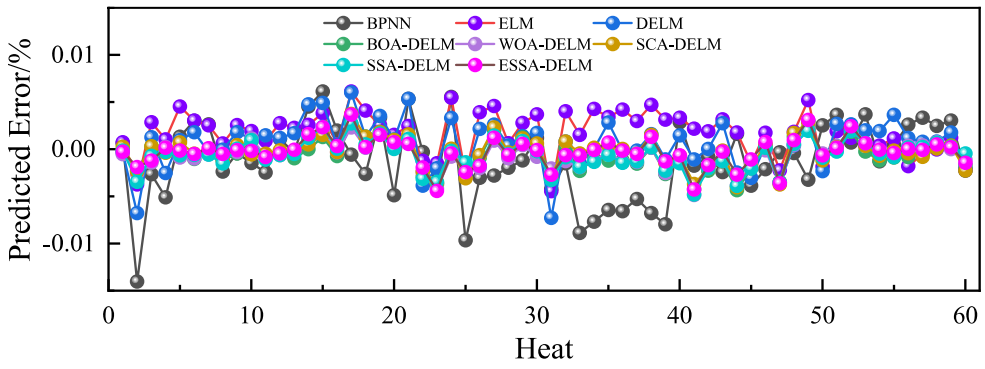

Figure 7. Prediction error comparison of the eight models.

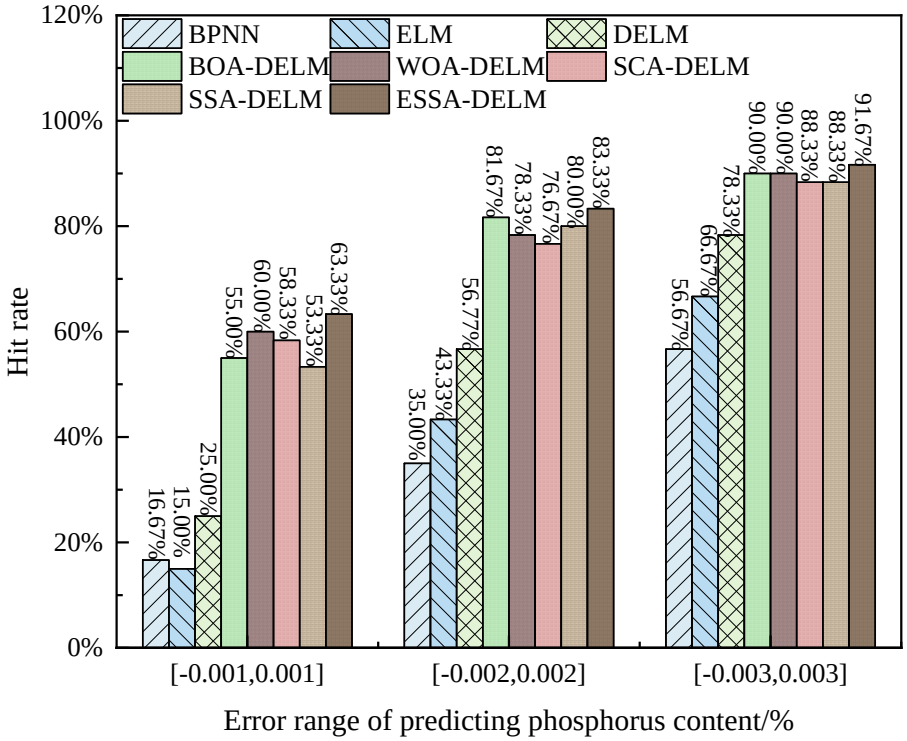

Figure 8. Hit rates comparison of the eight models in three error ranges.

Table 5 Hit rates of the eight models in different error range.

Prediction model	Hit rate		
	[-0.001%,0.001%]	[-0.002%,0.002%]	[-0.003%,0.003%]
PNN	16.67%	35.00%	56.67%
ELM	15.00%	43.33%	66.67%
DELM	25.00%	56.67%	78.33%
BOA-DELM	55.00%	81.67%	90.00%
WOA-DELM	60.00%	78.33%	90.00%
SCA-DELM	58.33%	76.67%	88.33%
SSA-DELM	53.33%	80.00%	88.33%
ESSA-DELM	**63.33%**	**83.33%**	**91.67%**

From Table 5 and Figure 8, the hit rate based on ESSA-DELM model is 78.33% when the prediction errors are within ±0.003%;

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 December 2021 doi:10.20944/preprints202112.0192.v1

the hit rate of the model is 56.67% when the prediction errors are within ±0.002%; the hit rate of the model is 25.00% when the prediction errors are within ±0.001%, which demonstrated the best prediction accuracy. The hit rates of BOA-DELM, WOA-DELM, SCA-DELM, and SSA-DELM in the error range of [-0.001%,0.001%], [-0.002%,0.002%], and [-0.003%, 0.003%] are significantly greater than those of the DELM model. It is visually demonstrated that the forecast precision of the model has been improved because of the integration of the intelligent algorithm with the DELM model. Furthermore, the DELM model displays the hit rate amongst the conventional models with 78.33%, 56.67%, and 25.00% when the prediction errors are within ±0.003%, ±0.002%, and ±0.001%, respectively. It could be observed that the DLEM model shows a higher hit rate compared with the ELM model, which also proves that the DELM model possesses superior prediction precision than the original ELM model. These results indicate that a better hit rate has been obtained by the ESSA-DELM model than another prediction model.

To further verify the performance of the ESSA-DELM model, this study also adopted mean absolute error (MAE), the root mean square error (RMSE) and the mean absolute percentage error (MAPE), determination coefficient (R^2), and Nash–Sutcliffe Efficiency (NSE) [35] as the evaluation criteria. The corresponding results among the eight prediction models are shown in Table 6.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|\left(y_i - \hat{y}_i\right)\right| \tag{17}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2} \tag{18}$$

$$MAPE = 100 \times \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right| \tag{19}$$

$$R^2 = \frac{\left(n\sum_{i=1}^{n}\hat{y}_i y_i - \sum_{i=1}^{n}\hat{y}_i \sum_{i=1}^{n} y_i\right)^2}{\left(n\sum_{i=1}^{n}\hat{y}_i^2 - \left(\sum_{i=1}^{n}\hat{y}_i\right)^2\right)\left(n\sum_{i=1}^{n} y_i^2 - \left(\sum_{i=1}^{n} y_i\right)^2\right)} \tag{20}$$

$$NSE = 1 - \frac{\sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2}{\sum_{i=1}^{n}\left(y_i - y_{\text{mean}}\right)^2} \tag{21}$$

where n indicates the number of samples; y_i and $\hat{y}_i$ represents the actual value and predicted result of model respectively. Additionally, y_{mean} describes the average of input variables. The optimal values of these indices are presented in Table 6. The closer the NSE and R are to 1, the better the model performance.

 doi:10.20944/preprints202112.0192.v1

Table 6. Prediction performance of different models.

Prediction model	RMSE	MAE	MAPE	R^2	NSE
BPNN	0.0042066	0.003285	21.51%	0.056638	-1.9951
ELM	0.0028669	0.0024928	15.35%	0.11124	-0.39114
DELM	0.0026365	0.0020723	13.65%	0.019823	-0.17656
BOA-DELM	0.0016629	0.0012042	7.98%	0.56696	0.53197
WOA-DELM	0.0016192	0.0011818	7.84%	0.55902	0.55625
SCA-DELM	0.0016395	0.0012029	8.14%	0.55604	0.54504
SSA-DELM	0.0016685	0.0012595	8.41%	0.59329	0.52879
ESSA-DELM	0.0015366	0.001074	7.15%	0.61391	0.60034

According to the statistical results derived from RMSE, MAE, MAPE, R^2, and NSE metrics, in Table 6, DELM shows a better predictive performance with RMSE=0.0026365, MAE=0.0020723, MAPE=15.35%, R^2=0.019823, and NSE=-0.17656 compared with the conventional models like ELM and BPNN. Because of the introduction of SSA, SSA-DELM shows a better predictive performance with RMSE=0.0016685, MAE=0.0012595, MAPE=8.41%, R^2=0.59329, and NSE=0.52879, compared with DLEM. However, compared with BOA-DELM, SCA-DELM, and WOA-DELM, SSA-DELM shows no advantage in terms of RMSE, MAE, MAPE, and NSE performance, except for R^2. Moreover, ESSA-DELM shows the best predictive performance with RMSE=0.0015366, MAE=0.001074, MAPE=7.15%, R^2=0.61391, and NSE=0.60034 with the introduction of TS strategy and Cauchy mutation. It is obvious that ESSA-DELM outperforms BOA-DELM, SCA-DELM, and WOA-DELM on RMSE, MAE, and MAPE, and it outperforms SSA-DELM in terms of R^2. In this paper, the proposed ESSA-DELM prediction model obtain high prediction accuracy performance among all prediction models, which can offer a good reference for industrial operation. However, the R^2 and NSE are not particularly high, because the markers determining the endpoint phosphorus content still interact with one another. In conclusion, ESSA-DELM demonstrated promising performance in phosphorus content prediction

4. Conclusion

This paper proposes the ESSA-DELM prediction model for the end-point phosphorus content of BOF. Nevertheless, given the numerous random input weights and biases in DELM model, the prediction precision will be degraded. Thus this research introduces an enhanced sparrow search algorithm to optimize. ESSA is obtained by introducing Trigonometric substitution and Cauchy mutation for enhancing the exploration and exploitation capacity of original SSA. Moreover, the superiority of the algorithm is verified by comparing SSA with other classic intelligent algorithms and advanced algorithms.

Simultaneously, to evaluate the prediction accuracy, the BPNN, ELM, DELM, BOA-DELM, WOA-DELM, SCA-DELM, and SSA-DELM are adopted as a comparative prediction model. The 10 parameters that influence end-point phosphorus content are chosen as inputs, and the end-point phosphorus content is

chosen as output. Some performance evaluation criteria are applied for comparing the model predicted results to actual values. Finally, the experimental result indicates the hit rate of ESSA-DELM within the error range of ±0.003%, ±0.002%, and ±0.001% is 91.67%, 83.33%, and 63.55%, respectively, which has a higher hit rate than the other 7 models. In addition, the performance metrics (RMSE, MAE, MAPE, R^2, NSE) of the ESSA model are superior to other predictive models. Obviously, the proposed ESSA-DELM model could obtain a better prediction performance and guide for controlling the end-point phosphorus content of BOF. In the future, since the values of NSE and R do not show very good performance, the converter data from different steel mills will be collected for further study.

Author Contributions: Conceptualization, L.Q. and A.L.; methodology, L.Q. and A.L.; software, L.Q.; validation, L.Q., A.L. and G.C.; formal analysis, L.Q. and A.L.; investigation, L.Q. and A.L.; resources, L.Q., G.C. and S.X.; data curation, L.Q., G.C. and S.X.; writing—original draft preparation, L.Q.; writing—review and editing, A.L.; visualization, G.C. and S.X. supervision, A.L. All authors have read and agreed to the published version of the manuscript.

Funding: This work was financially supported by the National Natural Science Foundation of China (No. 61763039)

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Bai, X.F.; Sun, Y.; Luo L.; Zhao, C.L. Effect of direct charging of hot recycled slag on hot metal pretreatment dephosphorization in a dephosphorization furnace. J. Iron Steel Res. Int. 2020, 27, 148-159. https://doi.org/10.1007/s42243-019-00296-w.
2. Wang, Z.; Liu, Q.; Liu, H.T.; Wei, S.Z. A review of end-point carbon prediction for BOF steelmaking process. High Temp. Mater. Process. 2020, 39, 653-662. https://doi.org/10.1515/htmp-2020-0098.
3. Jiang, M.F.; Cui, Y.Y.; Wang, D.Y.; Min, Y.; Liu, C.J. Effect of modification treatment for reduction of dephosphorization slag in hot metal bath. J. Iron Steel Res. Int. 2013, 20, 1-6. https://doi.org/10.1016/S1006-706X(13)60036-9.
4. Werbos, P.J. The roots of backpropagation: from ordered derivatives to neural networks and political forecasting, John Wiley & Sons, Hoboken, USA, 1994.
5. Li, C.R.; Zhao, H.W.; Xie, X.; Yin, Q. Prediction of End-Point Phosphorus Content for BOF Based on LM BP Neural Network. Iron. Steel 2011, 46, 23–25. https://doi.org/10.3724/SP.J.1077.2010.10636.
6. He, F.; Zhang, L.Y. Prediction model of end-point phosphorus content in BOF steelmaking process based on PCA and BP neural network. J Process Control 2018, 66, 51-58. https://doi.org/10.1016/j.jprocont.2018.03.005.
7. Zhou, K.X.; Lin, W.H.; Sun, J.K.; Zhang, J.S.; Zhang, D.Z.; Feng, X.M.; Liu, Q. Prediction model of end-point phosphorus content for BOF based on monotone-constrained BP neural network. J. Iron Steel Res. Int. 2021, 1-10. https://doi.org/10.1007/s42243-021-00655-6.
8. Huang, G.B.; Zhu, Q.Y.; Siew, C.K. Extreme learning machine: theory and applications. Neurocomputing. 2006, 70, 489–501. https://doi.org/10.1016/j.neucom.2005.12.126.
9. Huang, G.B.; Wang, D.H.; Lan, Y. Extreme learning machines: a survey. Int J Mach Learn Cybern 2011, 2, 107–122. https://doi.org/10.1007/s13042-011-0019-y.
10. Tan, L.; Han, J.; Zhang, H.T.; Ultra-Short-Term Wind Power Prediction by Salp Swarm Algorithm-Based Optimizing Extreme Learning Machine. IEEE Access. 2020, 8, 44470-44484

https://doi.org/10.1109/ACCESS.2020.2978098.
11. Zheng, L.K.; Wang, Z.J.; Zhao, Z.Y.; Wang, J.Y.; Du, W.H. Research of Bearing Fault Diagnosis Method Based on Multi-Layer Extreme Learning Machine Optimized by Novel Ant Lion Algorithm. IEEE Access. 2019, 7, 89845-89856. https://doi.org/10.1109/ACCESS.2019.2926348.
12. Chen, J.C.; Zeng, Y.J.; Li, Y.; Huang, G.B. Unsupervised feature selection based extreme learning machine for clustering. Neurocomputing. 2020, 386, 198-207. https://doi.org/10.1016/j.neucom.2019.12.065.
13. Su, X.L.; Zhang, S.; Yin, Y.X.; Xiao, W.D. Prediction model of hot metal temperature for blast furnace based on improved multi-layer extreme learning machine. Int. J. Mach. Learn. Cybern. 2019, 10, 2739-2752. https://doi.org/10.1007/s13042-018-0897-3.
14. Tang, J.; Deng, C.; Huang G.B. Extreme Learning Machine for Multilayer Perceptron. IEEE Trans Neural Netw Learn Syst, 2017(4): 809-821. https://doi.org/10.1109/TNNLS.2015.2424995.
15. Wang, X.H.; Gao, X.J.; Wang, Z.X.; Ma, C.R.; Song, Z.X. A Combined Model Based on EOBL-CSSA-LSSVM for Power Load Forecasting. Symmetry 2021, 13, 1579. https://doi.org/10.3390/sym13091579.
16. Zhu, J.; Tan, T.X.; Wu, L.F.; Yuan, H.M. RUL prediction of lithium-ion battery based on improved DGWO-ELM method in a random discharge rates environment. IEEE Access, 2019, 7, 125176-125187. https://doi.org/10.1109/ACCESS.2019.2936822.
17. Liu, T.T.; Yuan, Z.; Wu, L.; Badami, B. An optimal brain tumor detection by convolutional neural network and enhanced sparrow search algorithm. Proc Inst Mech Eng H, 2021, 235, 459-469. https://doi.org/ 10.1177/0954411920987964.
18. Zhu, Y.L.; Yousefi, N. Optimal parameter identification of PEMFC stacks using adaptive sparrow search algorithm. Int. J. Hydrog. Energy. 2021, 46, 9541-9552. https://doi.org/ 10.1016/j.ijhydene.2020.12.107.
19. Xue, J.K.; Shen, B. A novel swarm intelligence optimization approach: sparrow search algorithm. Syst. Sci. Control. Eng. 2020, 8, 22-34. https://doi.org/10.1080/21642583.2019.1708830.
20. Liu, G.Y.; Shu, C.; Liang, Z.W.; Peng, B.H.; Cheng, L.F. A modified sparrow search algorithm with application in 3d route planning for UAV. Sensors,2021, 21, 1224-1224. https://doi.org/ 10.3390/s21041224.
21. Yuan, J.H.; Zhao, Z.W.; Liu, Y.P.; He, B.L.; Wang, L.; Xie, B.B.; Gao, Y.L. DMPPT control of photovoltaic microgrid based on improved sparrow search algorithm. IEEE Access, 2021, 9, 16623-16629. https://doi.org/ 10.1109/ACCESS.2021.3052960.
22. Bengio, Y. Learning deep architectures for AI. Found Trends Mach Learn 2019, 2, 1–127 https://doi.org/10.1561/2200000006.
23. Tanyildizi, E.; Demir, G. Golden Sine Algorithm: A Novel Math-Inspired Algorithm. Adv. Electr. Comput. Eng. 2017, 17, 71-78. https://doi.org/10.4316/AECE.2017.02010.
24. Liu, J.S.; Ma, Y.X.; Li, Y. Improved Butterfly Algorithm for Multi-dimensional Complex Function Optimization Problem. Acta Electronic Sinca 2021, 49, 1068-1076. https://doi.org/10.12263 /DZXB.20200148.
25. Mirjalili, S. SCA: A Sine Cosine Algorithm for solving optimization problems. Knowl Based Syst, 2016, 96, 120-133. https://doi.org/10.1016/j.knosys.2015.12.022.
26. Arora, S.; Singh, S. Butterfly optimization algorithm: a novel approach for global optimization. Soft Comput. 2019, 23, 715 − 734. https://doi.org/10.1007/s00500-018-3102-4.
27. Mirjalili, S.; Lewis, A. The Whale Optimization Algorithm. Adv. Eng. Softw. 2016, 95, 51-67. https://doi.org/10.1016/j.advengsoft.2016.01.008.
28. Nadimi-Shahraki, M.H.; Taghian, S.; Mirjalili, S. An improved grey wolf optimizer for solving engineering problems. Expert Syst. Appl. 2021, 166, 113917.1-113917.25. https://doi.org/10.1016/j.eswa.2020.113917.
29. Naik, M.K.; Panda, R.; Abraham, A. Normalized square difference based multilevel thresholding technique for multispectral images using leader slime mould algorithm. J King Saud Univ-Com, 2020. https://doi.org/10.1016/j.jksuci.2020.10.030.
30. Naik, M.K.; Panda, R.; Wunnava, A. A leader Harris hawks optimization for 2-D Masi entropy-based multilevel image thresholding. Multimed. Tools. Appl. 2021, 1-41. https://doi.org/ 10.1007/s11042-020-10467-7.
31. Naik, M.K.; Panda, R.; Abraham, A. Adaptive opposition slime mould algorithm. Soft Comput. 2021, 1-17. https://doi.org/ 10.1007/s00500-021-06140-2.
32. Zhang, M.J.; Long, D.Y.; Qin, T.; Yang, J. A Chaotic Hybrid Butterfly Optimization Algorithm with Particle Swarm Optimization for High-Dimensional Optimization Problems. Symmetry 2020, 12, 1800. https://doi.org/10.3390/sym12111800.
33. He, X.Z.; Huang, J.D.; Rao, Y.Q.; Gao, L. Chaotic teaching-learning-based optimization with Lévy flight for global numerical optimization. Comput. Intell. Neurosci. 2016, 2016, 43. https://doi.org/

10.1155/2016/8341275.

34. Liu, X.; Yang, S.; Liu, S.B.; Yang, Y. Performance and mechanism of phosphorus removal by slag ceramsite filler. Process Saf Environ Prot. 2021, 148, 858-866. https://doi.org/ 10.1016/j.psep.2021.02.016.
35. Dayev, Z.; Kairakbaev, A.; Yetilmezsoy, K.; Bahramian, M.; Sihag, P.; Kıyan, E. Approximation of the discharge coefficient of differential pressure flowmeters using different soft computing strategies. Flow Meas Instrum. 2021, 79, 101913. https://doi.org/ 10.1016/j.flowmeasinst.2021.101913 .