Extreme Learning Machine Soft-Sensor Model With Different Activation Functions on Grinding Process Optimized by Improved Black Hole Algorithm

W. XIE[1], **J. S. WANG**[1,2], **(Member, IEEE), C. XING**[1], **S. S. GUO**[1], **M. W. GUO**[1], **AND L. F. ZHU**[1]

[1]School of Electronic and Information Engineering, University of Science and Technology Liaoning, Anshan 114051, China
[2]National Financial Security and System Equipment Engineering Research Center, University of Science and Technology Liaoning, Anshan 114051, China

Corresponding author: J. S. Wang (wang_jiesheng@126.com)

ABSTRACT Aiming at predicting the key economic and technical indicators (Granularity and Ore content)in the grinding production process, the extreme learning machine (ELM) soft-sensor model with different activation functions on grinding process optimized by improved black hole (BH) algorithm was proposed. Based on the selected auxiliary variables for the soft-sensor model of the grinding, the KPCA method is used to reduce the dimensionality of the high-dimensional data. In order to investigate the influence of different activation functions on the prediction accuracy of the ELM model, seven continuous function (Arctan, Hardim, Morlet, ReLu, Sigma, Sin and Tanh) are used as the activation function of the ELM neural network to establish soft-sensor models respectively. For the shortcomings that ELM model weights and offset values are arbitrarily given so as to result in the low prediction accuracy and low reliability, an improved BH algorithm based on the golden sine operator and the Levy flight operator (GSLBH) was used to optimize the parameters of the ELM neural network. Simulation results show that the model has better generalization and prediction accuracy, and can meet the real-time control requirements of the grinding process.

INDEX TERMS Grinding process, soft-sensor, ELM neural network, black hole algorithm, activation function.

I. INTRODUCTION

In the process of processing iron ore, the grinding and classifying process is a very important step in the industrial process of the concentrator. This process is to grind ore with different sizes to a certain particle size, and to separate useful minerals and impurity minerals so that these minerals are dissociated into monomers, which is beneficial to mineral selection. Therefore, some production indicators (granularity, impurity content, iron content) during the grinding process will have a great impact on the subsequent operations (especially for flotation operation) and even the economic and technical indicators of the entire concentrator [1]. The grinding process is complicated and affected by many factors, including changes

The associate editor coordinating the review of this manuscript and approving it for publication was Zhipeng Cai.

in ore characteristics entering the circuit, ore hardness, particle size distribution, mineral composition, the flow rate of water into the circuit and changes in the properties of the vortex feed slurry, etc. The grinding process is characterized by serious non-linearity, strong coupling, and large lag. Due to field conditions and lack of mature detection devices, some key economic and technical indicators (grinding size and iron content) of the grinding process are difficult to obtain in real time, and it is difficult to achieve direct quality closed-loop control [2]–[5]. The soft-sensor method can effectively solve the problem of estimating process quality indicators that are difficult to measure online [6]. So many experts and scholars have applied soft-sensor methods to industrial processes [7]–[10].

In view of the complicated grinding and classifying process, many experts and scholars have proposed various

W. Xie *et al.*: ELM Soft-Sensor Model With Different Activation Functions on Grinding Process Optimized by Improved BH Algorithm

IEEE *Access*

soft-sensing models to predict different production indicators. In the early research of soft-sensing modeling of the grinding process, due to the limitations of theoretical knowledge and computing equipment, some experts tried to analyze the entire grinding process clearly and adopt the algebraic equations or differential equations to establish a mechanism model as a soft-sensor model of the grinding process to predict the ore granularity [11]–[13]. When modeling the mechanism, it is necessary to assume that all working conditions are ideal. But in the actual industrial production process, the industrial production environment is subject to all kinds of unknown disturbances all the time, which will lead to a large error when tracking a certain industrial production index with the method of mechanism modeling. In order to make the soft-sensing model based on mechanism modeling adapt to changes in various environments, Sbarbaro et al. introduced the adaptive weight parameters in the mechanism model for the first time [14], so that the mechanism model has a certain ability to adapt to changes in industrial environments. However, due to the poor robustness of the mechanism model, it still cannot cope with large changes in the industrial environment, and the prediction error of the model still cannot be maintained in a small range. Therefore, in recent years, few people have studied the mechanism based soft-sensor modeling methods. In order to make the soft-sensing model still maintain high prediction accuracy when the industrial environment changes, some experts use case-based reasoning (CBR) models to predict the granularity of the grinding process [15], [16]. This method can cope with the changes of the industrial environment very well, and the calculation efficiency of this method is relatively high. However, this method requires accurate classification of samples during the training process, the characteristics of the data collected in a complex industrial environment are diverse, and it is difficult to effectively classify data samples with a more appropriate clustering algorithm, which will cause a large error when test samples are carried out case reuse operations, and the prediction accuracy of soft-sensing models is greatly reduced.

The use of neural networks to build a soft-sensing model of the grinding process shows great advantages. Neural networks only need to focus on the input and output of the model when building a soft-sensing model, and rarely focus on specific industrial processes. The neural network has strong generalization ability. Once a more suitable structure is determined, it is only necessary to adjust some parameters in the model to make it have better prediction accuracy. So many researchers are very willing to adopt neural network to set up the soft-sensing model of the grinding process to predict the various production indicators. Some researchers have used RBF neural network as a soft-sensing model and obtained some prediction results of ore granularity [17], [18]. Support vector machine (SVM) is also a more commonly used soft-sensing modeling method [19], [20], but the accuracy of prediction is not very good. In order to improve the training accuracy of the soft-sensing model of neural network, Wang et al. used the shuffled frog leap algorithm (SFLA)

to optimize the structural parameters of wavelet neural network [21], which greatly improved the prediction accuracy of wavelet neural network. In order to take advantage of the prediction advantages of various neural network models, Shen et al. proposed a hybrid model based on three different neural networks [22], and proposed a model switching indicator to determine which neural network model should be used to predict the granularity of the ore at a certain moment. However, this method has a higher calculation cost when training the model and reduces the work efficiency. Chai et al. proposed an online robust random vector functional link network (online robust RVFLN)[23], which enhanced the adaptability of neural networks and achieved better prediction results. Some scholars use the extreme learning machine (ELM) to predict the ball mill fill level in the milling process [24], [25], but the error is large. So far, the use of ELM models to predict the ore granularity and pulp iron content in the grinding and classifying process is still relatively rare.

A single hidden layer feed-forward network (SLFN) can be used as a general-purpose approximator, and the extreme learning machine (ELM) is a type of SLFN [26]. ELM algorithm is a very efficient learning algorithm proposed by Huang *et al.* [27]. The neural network structure of the ELM model is similar to the BP neural network, but it has a fixed number of network layers, with only three layers, namely the input layer, the hidden layer, and the output layer. The weight w and bias p between the input layer and the hidden layer of the ELM model are given randomly. The weight between the hidden layer and the output layer is obtained by solving the least squares solution of the ELM neural network model. In recent years, ELM is often widely used to solve classification problems and regression problems due to its short training time, high execution efficiency, and strong generalization ability [28], [29]. Zhang used the improved ELM model to classify images [30], [31], gas types [32], and beverage types [33], respectively, and all achieved better classification results. Li et al. used the improved ELM model to accurately predict the transient stability of power systems [34], [35]. Because the training method of the ELM model is relatively fixed, the traditional ELM model can only enhance the generalization ability by increasing the number of neurons in the hidden layer. But increasing the prediction accuracy of a neural network by increasing the number of neurons is limited.When the number of neurons in the ELM model is large, its generalization ability may even be destroyed, so it is necessary to find other methods to improve the prediction performance of the ELM model. The ELM model is a type of neural network, which means that it can also adjust the weights and bias between the input layer and the hidden layer by finding the minimum value of the error function, so that the prediction accuracy of the ELM model is improved. Then natural heuristic algorithm with strong exploaration can also be used to find the optimal structural parameters corresponding to the minimum value of the ELM error function.
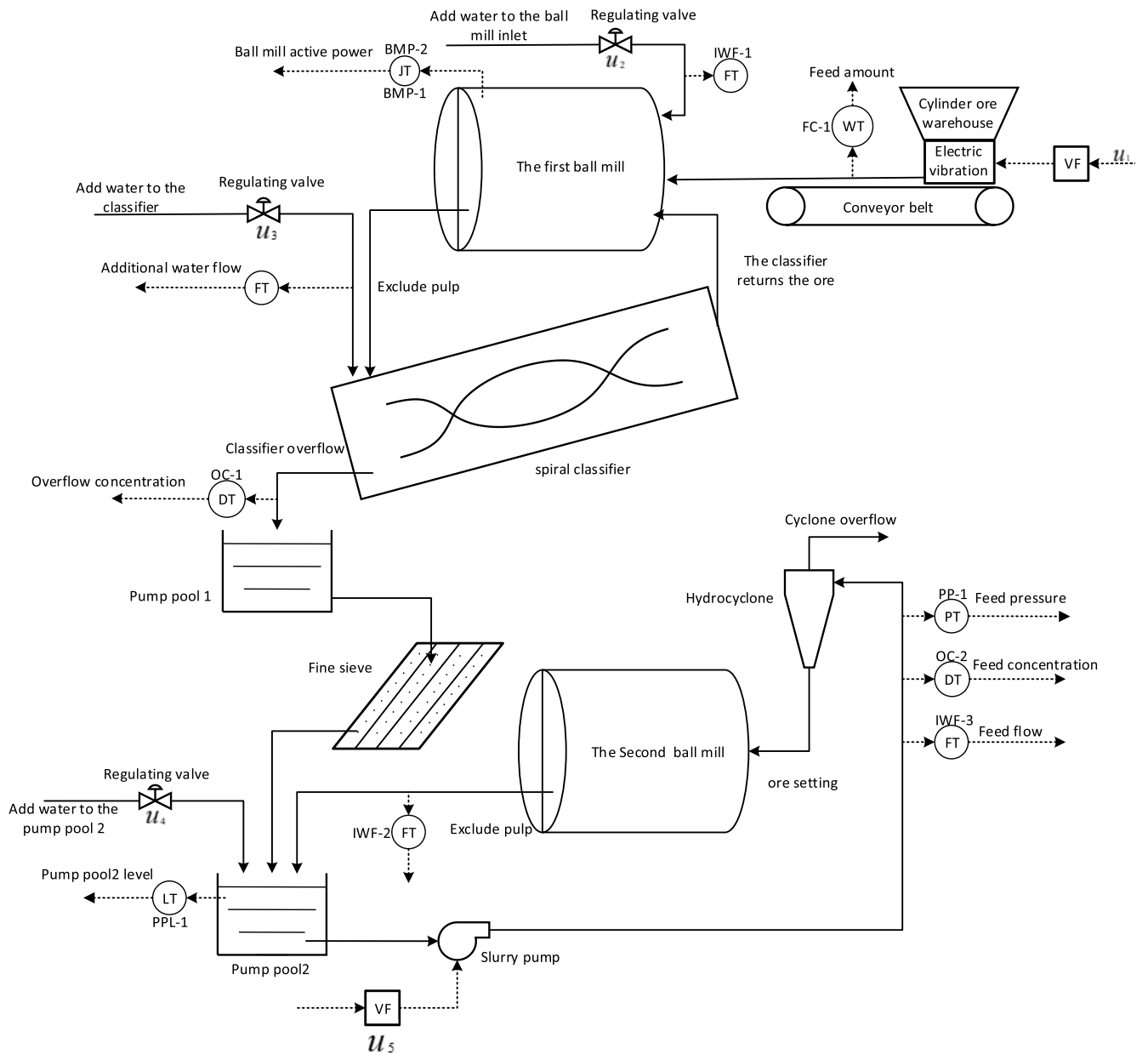
FIGURE 1. Technique flowchart of grinding process.

In this paper, the ELM model is used to predict the ore granularity and pulp iron content during grinding and classifying process. The specific operation process is as follows:

(1) Considering the co-linearity and redundancy between the secondary variable, this paper first adoptes the KPCA method to perform dimensionality reduction on the data to reduce the complexity of the data and improve the data utilization rate.

(2) The improved black hole algorithm (GSLBH) is an optimization algorithm with strong global search ability [37], in the process of training ELM, this paper will use GSLBH to optimize the error function of ELM, and find the weight w

and bias p corresponding to the minimum value of the error function through GSLBH.

(3) Huang believes that any continuous function can be used as the excitation function of the ELM model [27]. In order to explore the influence of different excitation functions on the prediction accuracy of ELM models, seven continuous functions of Arctan, Hardim, Morlet, ReLu, Sigma, Sin, Tanh were used as the excitation functions of the ELM neural network, and these ELM models were optimized by GSLBH.

The simulation experiments are carried out to verify the prediction accuracy of the improved ELM model, and the

W. Xie et al.: ELM Soft-Sensor Model With Different Activation Functions on Grinding Process Optimized by Improved BH Algorithm

IEEE Access

computational cost of the improved ELM model is discussed. The structure of this paper is described as follows. The second section introduces the technique of grinding process and soft-sensor model. The third section introduces the ELM soft-sensor model with different activation functions. The fourth section performs experimental simulation and results analysis. The fifth section draws the conclusions.

II. TECHNIQUE OF GRINDING PROCESS AND SOFT-SENSOR MODEL

A. TECHNIQUE OF GRINDING PROCESS

The grinding process of hematite in this paper is the research object, and the grinding and classifying process was described. The technique of the grinding process is shown in Fig. 1. The grinding process generally consists of two parts. The first part is a closed-circuit grinding process consisted of a ball mill and a spiral classifier and the second part is a two-stage closed-circuit grinding process consisted of a two-stage ball mill and a hydrocyclone. In Fig. 1, u_1 indicates the frequency of the mine-feeding inverter, u_2, u_3 and u_4 indicates the opening degree of the water valves, u_5 represents the frequency of the ore slurry pump inverter.

The rough-processed iron ore is put into a cylindrical ore bin, and then sent through a belt to a ball mill for grinding. At the same time, an appropriate amount of water needs to be added to the entrance of the ball mill to ensure that the ball mill can work normally. After a period of grinding, the ore slurry produced by the ball mill passes through the chute and enters the spiral classifier. Iron ore with a larger particle size settles faster in the classifier and is quickly sent to the chute on the top of the classifier as the it rotates, and the ore returns to one-stage ball mill through the chute to continue grinding. The remaining slurry in the classifier that contains only small-sized iron ore is called classifier overflow product. The overflow product enters the next fine sieving process. Classifier overflow slurry enters the mine pump pool after fine sieving process, the ore slurry in the pump pool is driven into the hydrocyclone by an electric pump. The hydrocyclone classifies the iron ore overflow slurry with particle size that meets the standards as the final product. Larger particles of ore enter the two-stage ball mill through the grit opening to continue grinding. The entire grinding and classifying process will repeat the above process.

B. SOFT-SENSOR MODELING OF GRINDING PROCESS

After careful analysis on the grinding process, ten process variables were selected as secondary variables for set up the soft-sensor model. The position for obtaining auxiliary variables is indicated in the corresponding position in the technique flowchart Fig. 1 and described in Table 1. The circles in Fig. 1 represent the measuring instruments for different variables. The target variables of the soft-sensor model are the grinding granularity and iron content. Information of secondary variables and target variables are listed in Table 1.

TABLE 1. Variables information of grinding process.

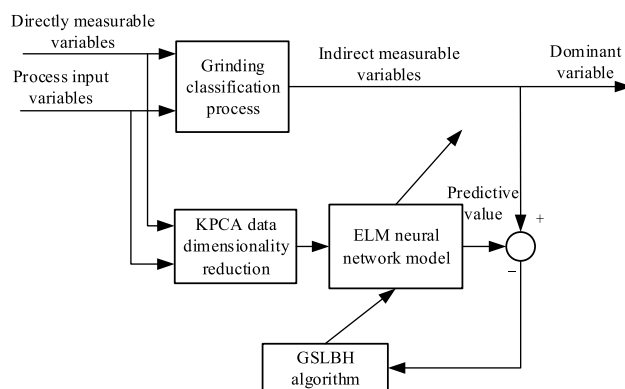| Variable type	Variable	Variable symbol	Unit
Secondary variable	Feeding capacity	FC-1	kg
	Inlet water flow	IWF-1	m³/s
	Outlet water flow	IWF-2	m³/s
	Pump pool level	PPL-1	m
	Pump pressure	PP-1	kpa
	Current	BMP-1	A
	Feed flow of cyclone	IWF-3	m³/s
	Overflow concentration	OC-1	%
	Ball mill power	BMP-2	KW
	Feed concentration	OC-2	%
Objective variable	Ore content		%
	Granularity		%

FIGURE 2. Soft-sensor model structure for grinding process.

The soft-sensor model data of the grinding process are listed in Table 2.

The 10 process variables introduced above are used as the input of the soft-sensor model, the grinding granularity and iron content are used as the output, and the ELM neural network optimized by the improved black hole algorithm was used to fit the nonlinear relationship between the input and output variables, then the soft-sensor model of key economic and technical indicators in the grinding process was established, whose structure is shown in Fig. 2.

C. DATA DIMENSION REDUCTION BASED ON KPCA METHOD

For neural network soft-sensing models, the excessive number of input vector dimensions makes the network topology large and the training process complicated [27]. At the same time, there may be redundancy between various variables, which will interfere with the prediction process of the neural network, so high-dimensional data information needs to be reduced. Thus, the KPCA method was adopted to realize the data reduction. KPCA is a kind of dimensionality reduction method based on kernel function. It uses the kernel function to perform non-linear mapping from low-dimensional space to

TABLE 2. Soft sensor modeling data for the grinding process.

	Secondary variable						Objective variable					
	Feeding capacity	Inlet water flow	Outlet water flow	Pump pool level	Pump pressure	Current	Feed flow of cyclone	Overflow concentration	Ball mill power	Feed concentration	Ore content	Granularity
1	134.796	81.35	62.253	1.584	0.042	63.000	128.156	74.326	1119.311	70.364	64.880	58.621
2	134.578	82.547	63.424	1.609	0.043	58.000	127.717	73.325	1118.638	69.314	64.376	58.992
3	133.633	82.465	63.718	1.595	0.042	66.000	127.911	74.364	1121.663	69.415	63.096	57.694
4	135.702	82.475	64.687	1.592	0.041	63.000	127.561	74.631	1118.561	67.314	67.216	58.323
5	135.412	82.457	62.413	1.617	0.041	60.000	127.454	75.254	1119.878	68.141	63.576	59.922
6	134.794	83.347	63.081	1.606	0.043	59.000	128.932	73.344	1118.749	70.564	63.736	54.773
...	...	...	...	...	...	...	...	...	...	...	...	...
1000	136.364	82.145	84.341	1.534	0.043	60.000	126.347	72.384	1120.379	70.364	63.742	56.314

TABLE 3. Contribution percentage of principal components.

Number	Variance percentage(%)	Cumulative variance percentage(%)
1	0.2448	0.2448
2	0.2035	0.4483
3	0.1738	0.6221
4	0.1087	0.7308
5	0.0947	0.8255
6	0.0625	0.8880
7	0.0501	0.9381
8	0.0442	0.9823
9	0.0160	0.9983
10	0.0017	1

TABLE 4. Expression of activation functions for ELM.

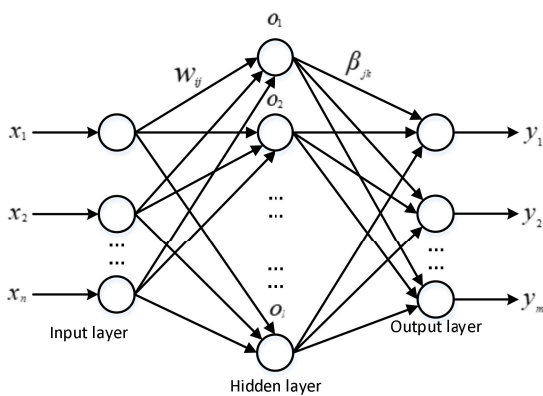| Function	Expression
arctan	$f_1(x) = \arctan(x)$
hardlim	$f_2(x) = \begin{cases} 1 & x > 0 \\ 0 & x \le 0 \end{cases}$
Morlet	$f_3(x) = \cos(1.75x)e^{-x^2/2}$
ReLu	$f_4(x) = \begin{cases} x & x > 0 \\ 0 & x \le 0 \end{cases}$
sigma	$f_5(x) = \dfrac{1-e^x}{1+e^x}$
sin	$f_6(x) = \sin(x)$
tanh	$f_7(x) = \dfrac{e^x - e^{-x}}{e^x + e^{-x}}$

FIGURE 3. Structure of typical feed-forward neural network with single hidden layer.

high-dimensional space. In this high-dimensional space, linear dimensionality reduction is performed by PCA [38].The specific process of KPCA dimension reduction is introduced as follows.

Step 1: Calculate the kernel matrix. The output sample of KPCA is $X = \{x_1, x_2, \cdots, x_n\}$, X is the original data set containing 10 secondary variable, then each element in the kernel matrix can be expressed as:

$$K_{ij}(x_i, x_j) = \exp(-\|x_i - x_j\|^2 / l_1^2) \tag{1}$$

where, $1 < i < n$, $1 < j < n$, and l_1 is the kernel width.

Step 2: Centralize the kernel matrix. The calculation method to centralize the kernel matrix can be expressed as:

$$K^* = K - \frac{1}{n}\theta\theta'K - \frac{1}{n}K\theta\theta' + \frac{1}{n}(\theta'K\theta)\theta\theta' \tag{2}$$

where, θ is an n-dimensional vector whose element values are all 1.

Step 3: Decompose eigenvalue. Let φ be a non-linear mapping. Then φ is used to transform input samples to feature space V, which is recorded as $V = \{\varphi(x_1),$
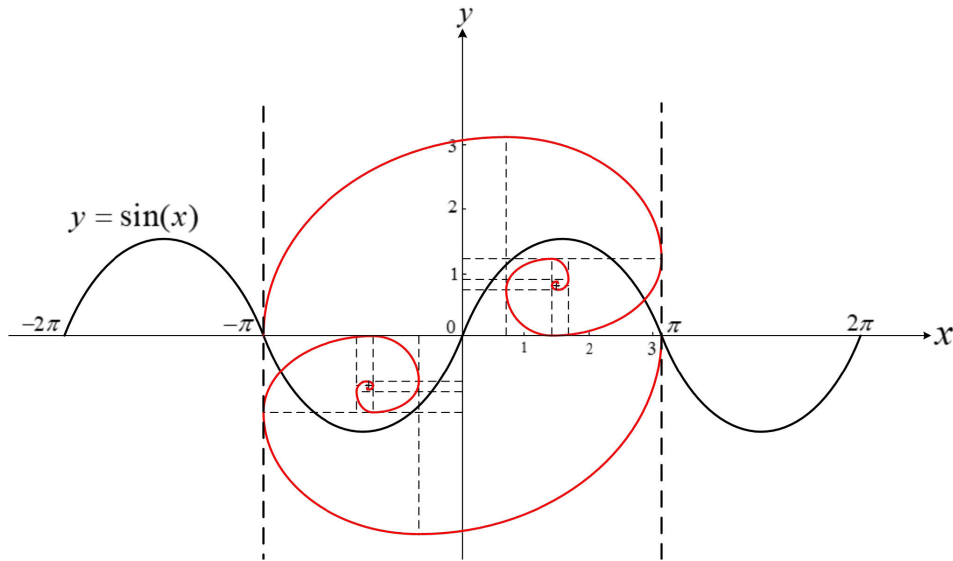
W. Xie *et al.*: ELM Soft-Sensor Model With Different Activation Functions on Grinding Process Optimized by Improved BH Algorithm

IEEE *Access*

FIGURE 4. Searching path of golden sine algorithm.
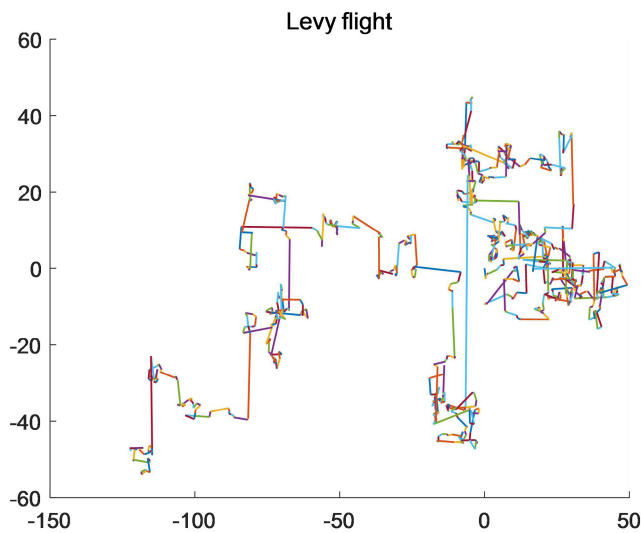
FIGURE 5. Trajectory of Levy flight in two-dimensional space.

TABLE 5. Quantitative indicators of soft-sensor model.

Index	Function
Maximun positive error	$MPE = \max\{(\hat{y} - y), 0\}$
Maximun negative error	$MNE = \min\{(\hat{y} - y), 0\}$
Root mean square error	$RMSE = \left[\dfrac{1}{n}\sum\limits_{i=1}^{n}(\hat{y}-y)^2\right]^{\frac{1}{2}}$
R^2	$R^2 = \dfrac{\sum\limits_{i=1}^{n}(\hat{y}_i - \overline{y})^2}{\sum\limits_{i=1}^{n}(y_i - \overline{y})^2}$

$\varphi(x_2), \ldots, \varphi(x_n)\}$. Suppose the following equation holds:

$$\sum_{i=1}^{n} \varphi(x_i) = 0 \tag{3}$$

Then in the feature space V, the covariance matrix C can be calculated by:

$$C = \frac{1}{n}\sum_{i=1}^{n} \varphi(x_i)\,\varphi(x_i)^T \tag{4}$$

Step 4: Solve Eigenvalues and sort the eigenvalue vectors according to the size of the eigenvalues. The eigenvalues and eigenvectors of Eq. (4) satisfy:

$$\lambda v = Cv \tag{5}$$

The linear expression of the sample vector of v mapped to the feature space C is:

$$v = \sum_{i=1}^{n} \alpha_i \varphi(x_i) \tag{6}$$

Step 5: Unitize eigenvectors. The mapping of sample $\varphi(x_i)$ on v can be expressed as:

$$
\begin{aligned}
j_k(x) &= \langle v, \varphi(x) \rangle \\
&= \sum_{i=1}^{n} \alpha_i^k \langle \varphi(x_i), \varphi(x) \rangle \\
&= \sum_{i=1}^{n} \alpha_i^k K(x, x_i)
\end{aligned}
\tag{7}
$$

where, $j_k(x)$ is kth non-linear principal component of φ.

Step 6: Determine the contribution percentage of the principal components. Assuming that the number of extracted principal elements is m, the calculation of the cumulative

IEEE *Access*

W. Xie *et al.*: ELM Soft-Sensor Model With Different Activation Functions on Grinding Process Optimized by Improved BH Algorithm
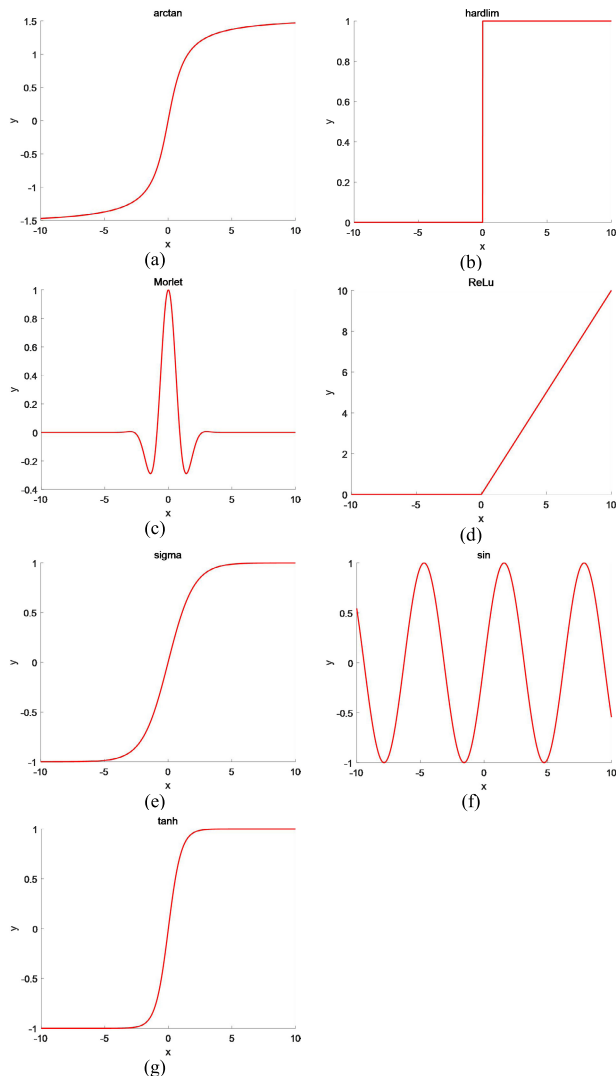

FIGURE 6. Activation functions. (a) arctan function (b) hardlim function (c) Morlet function (d) ReLu function (e) sigma function (f) sin function (g) tanh function.

contribution percentage E can be realized by:

$$E = \frac{\sum\limits_{k=1}^{m} \lambda_k}{\sum\limits_{k=1}^{n} \lambda_k} \quad (E > 85\%) \tag{8}$$

Step 7: Calculate the load vector after the principal components are extracted.

$$h_i^s = \sum_{w=1}^{n} K(x_i, x_w) j_w^s \tag{9}$$

where, $1 \leq i \leq n$, $1 \leq s \leq m$.

The input data of the soft-sensor model of grinding process were processed by above KPCA method and the results are shown in Table 3. It can be seen from Table 3 that the cumulative contribution percentage of the first 6 principal elements is already greater than 85%, so the first 6 principal components are used as the input of the ELM soft-sensor
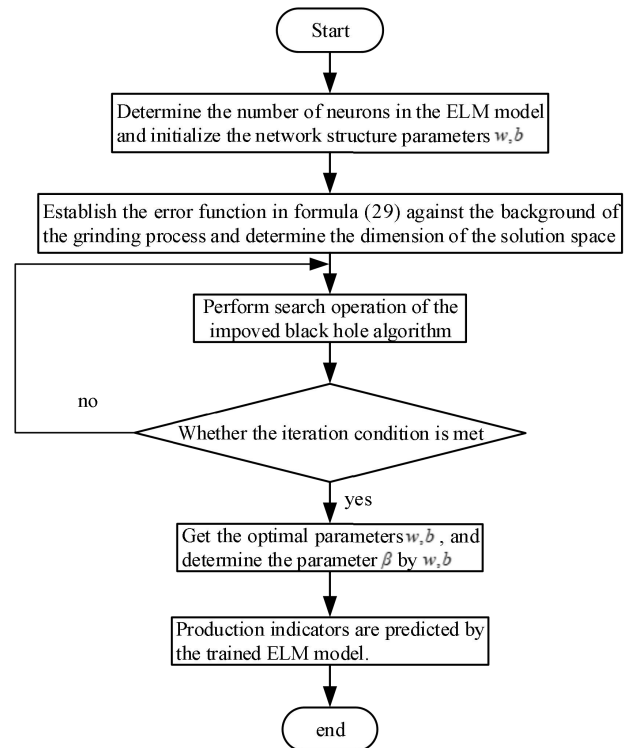

FIGURE 7. ELM optimization procedure based on improved black hole algorithm.

model. The complexity of the ELM neural network is greatly simplified, and the execution efficiency is also improved.

III. TECHNIQUE OF GRINDING PROCESS AND SOFT-SENSOR MODEL
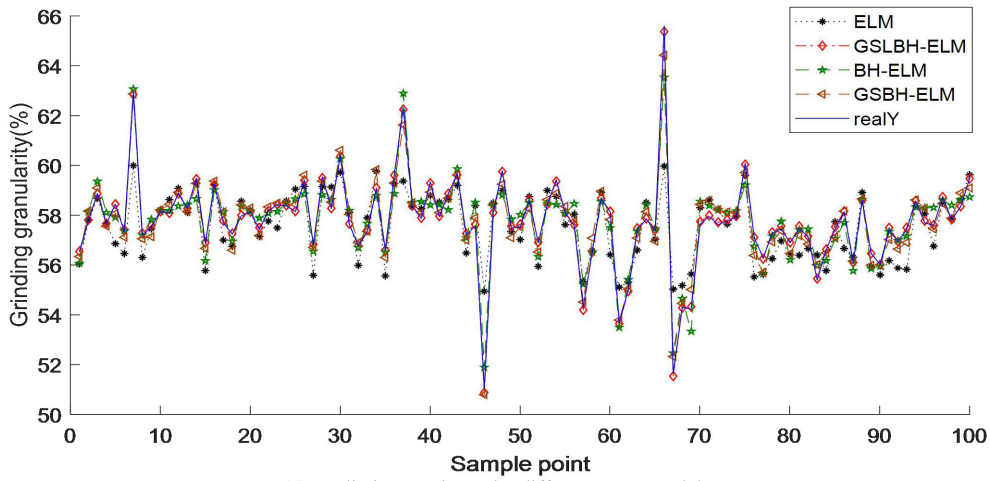### A. ELM NEURAL NETWORK

ELM is a neural network model with high learning efficiency proposed by Huang *et al.* [27], which is a typical SLFN. The structure schematic diagram of the ELM model is shown in Fig. 3.

Its structure is a fully connected structure, but it has only three layers, namely the input layer, the hidden layer, and the output layer. Let the ELM model have n neurons in the input layer, corresponding to n input variables, l neurons in the hidden layer, and m neurons in the output layer. The connection weight w between the input layer and the hidden layer can be described as:

$$w = \begin{bmatrix} w_{11} & w_{12} & \cdots & w_{1n} \\ w_{21} & w_{22} & \cdots & w_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ w_{l1} & w_{l2} & \cdots & w_{ln} \end{bmatrix}_{l \times n} \tag{10}$$

The connection weight β between the hidden layer and the output layer can be described as:

$$\beta = \begin{bmatrix} \beta_{11} & \beta_{12} & \cdots & \beta_{1m} \\ \beta_{21} & \beta_{22} & \cdots & \beta_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ \beta_{l1} & \beta_{l2} & \cdots & \beta_{lm} \end{bmatrix}_{l \times m} \tag{11}$$

W. Xie *et al.*: ELM Soft-Sensor Model With Different Activation Functions on Grinding Process Optimized by Improved BH Algorithm

IEEE *Access*

(a) Prediction results under different ELM models

(b)Prediction errors under different ELM models

FIGURE 8. Simulation results under activation function arctan.

where, β_{jk} represents the connection weight between the jth neuron in the hidden layer and the kth neuron in the output layer.

Let the threshold of hidden layer neurons b as:

$$b = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_l \end{bmatrix}_{l \times 1} \quad (12)$$

Let there be a training set of Q samples, the input matrix X and the output matrix Y are defined as:
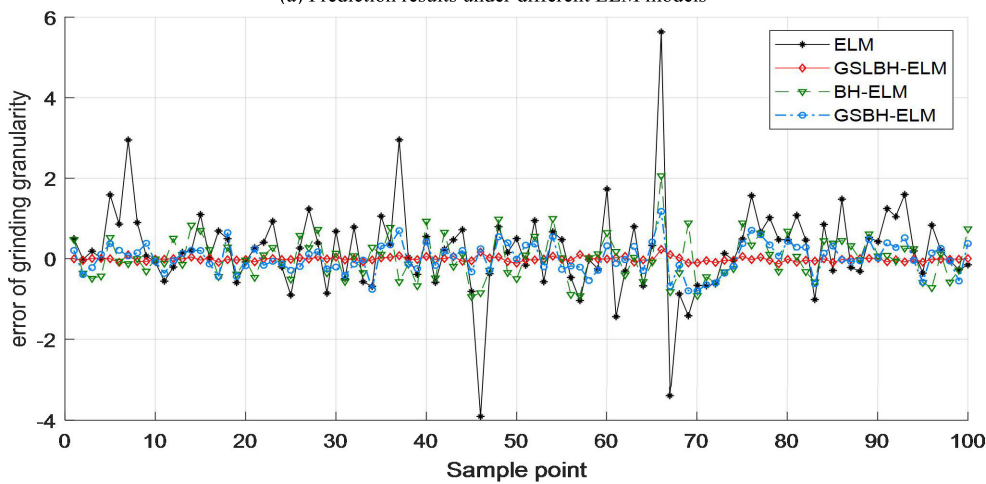
$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1Q} \\ x_{21} & x_{22} & \cdots & x_{2Q} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nQ} \end{bmatrix}_{n \times Q}$$

$$Y = \begin{bmatrix} y_{11} & y_{12} & \cdots & y_{1Q} \\ y_{21} & y_{22} & \cdots & y_{2Q} \\ \vdots & \vdots & \ddots & \vdots \\ y_{m1} & y_{m2} & \cdots & y_{mQ} \end{bmatrix}_{m \times Q} \quad (13)$$

Let the activation function of the hidden layer neuron be $g(x)$. It can be seen from the structure information of the ELM model in Fig. 3, the network output T can be calculated by:

$$T = \begin{bmatrix} t_1 & t_2 & \cdots & t_Q \end{bmatrix}_{m \times Q}, \quad t_j = \begin{bmatrix} t_{1j} \\ t_{2j} \\ \vdots \\ t_{mj} \end{bmatrix}_{m \times 1}$$

$$= \begin{bmatrix} \sum_{i=1}^{l} \beta_{i1} g(w_i x_j + b_i) \\ \sum_{i=1}^{l} \beta_{i2} g(w_i x_j + b_i) \\ \vdots \\ \sum_{i=1}^{l} \beta_{im} g(w_i x_j + b_i) \end{bmatrix}_{m \times 1} \quad j = 1, 2, \ldots, Q \quad (14)$$

where, $w_i = [w_{i1} \ w_{i2} \ \cdots \ w_{in}]$, $x_i = [x_{i1} \ x_{i2} \ \cdots \ x_{in}]^T$.

At this time, Eq. (14) can be expressed as:

$$H\beta = T' \quad (15)$$
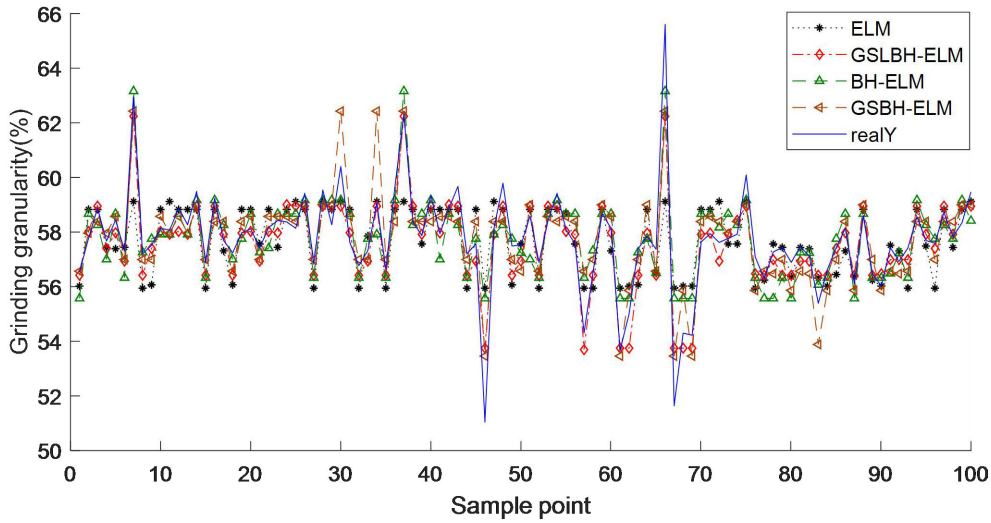
(a) Prediction results under different ELM models

(b)Prediction errors under different ELM models

FIGURE 9. Simulation results under activation function hardlim.

where, T' is the transpose of T. H is the hidden layer output matrix of the neural network, and its expression can be expressed as:

$$
\begin{aligned}
&H(w_1\, w_2 \cdots w_l\, b_1\, b_2 \cdots b_l\, x_1\, x_2 \cdots x_Q) \\
&= \begin{bmatrix}
g(w_1\, x_1 + b_1)(w_2\, x_1 + b_2)(w_l\, x_1 + b_l) \\
g(w_1\, x_2 + b_1)(w_2\, x_2 + b_2)(w_l\, x_2 + b_l) \\
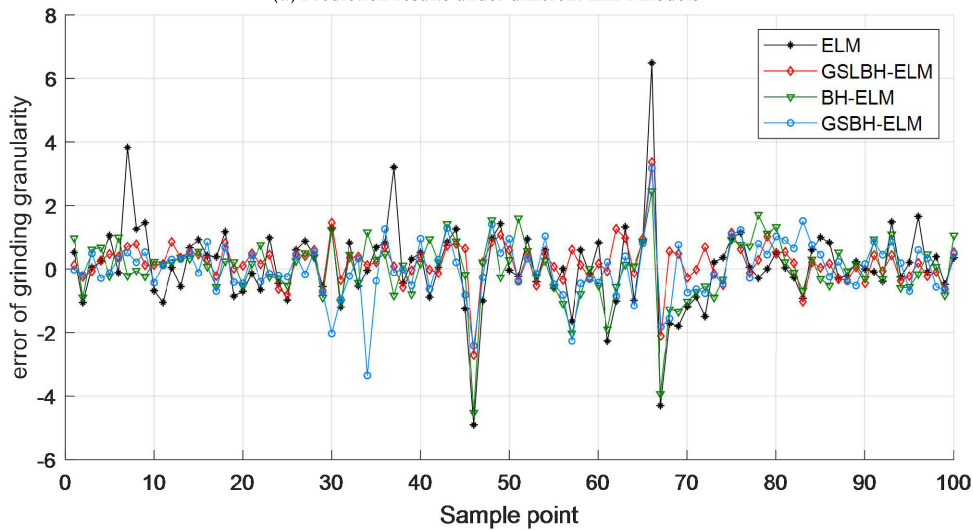\vdots \\
g(w_1\, x_Q + b_1)(w_2\, x_Q + b_2)(w_l\, x_Q + b_l)
\end{bmatrix}_{Q \times l}
\end{aligned}
\tag{16}
$$

According to the theory described by Huang *et al.* [27], when the number of hidden layer neurons is equal to the number of training set samples, for any w and b, the ELM model can approximate the value of the training sample with zero error, that is to say:

$$
\sum_{j=1}^{Q} \left\| t_j - y_j \right\| = 0
\tag{17}
$$

where, $y_j = [y_{1j}\ y_{2j}\ \cdots\ y_{mj}]^T$ ($j = 1, 2, \ldots, Q$). Because the training set sample Q is often large, in order to reduce the amount of calculations, the number of hidden neurons K is usually a number smaller than Q. the ELM model training error can approximate an arbitrary value $\varepsilon > 0$, that is to say:

$$
\sum_{j=1}^{Q} \left\| t_j - y_j \right\| < \varepsilon
\tag{18}
$$

When the activation function $g(x)$ is infinitely differentiable, the parameters of the ELM model do not need to be fully adjusted, w and b can be randomly selected before training, and remain unchanged during the training process, and the connection weight β between the hidden layer and the output layer can be obtained by solving the least squares solution of the following equations, that is to say:

$$
\min_{j=1} \left\| H\beta - T' \right\|
\tag{19}
$$

W. Xie *et al.*: ELM Soft-Sensor Model With Different Activation Functions on Grinding Process Optimized by Improved BH Algorithm

IEEE *Access*

(a) Prediction results under different ELM models
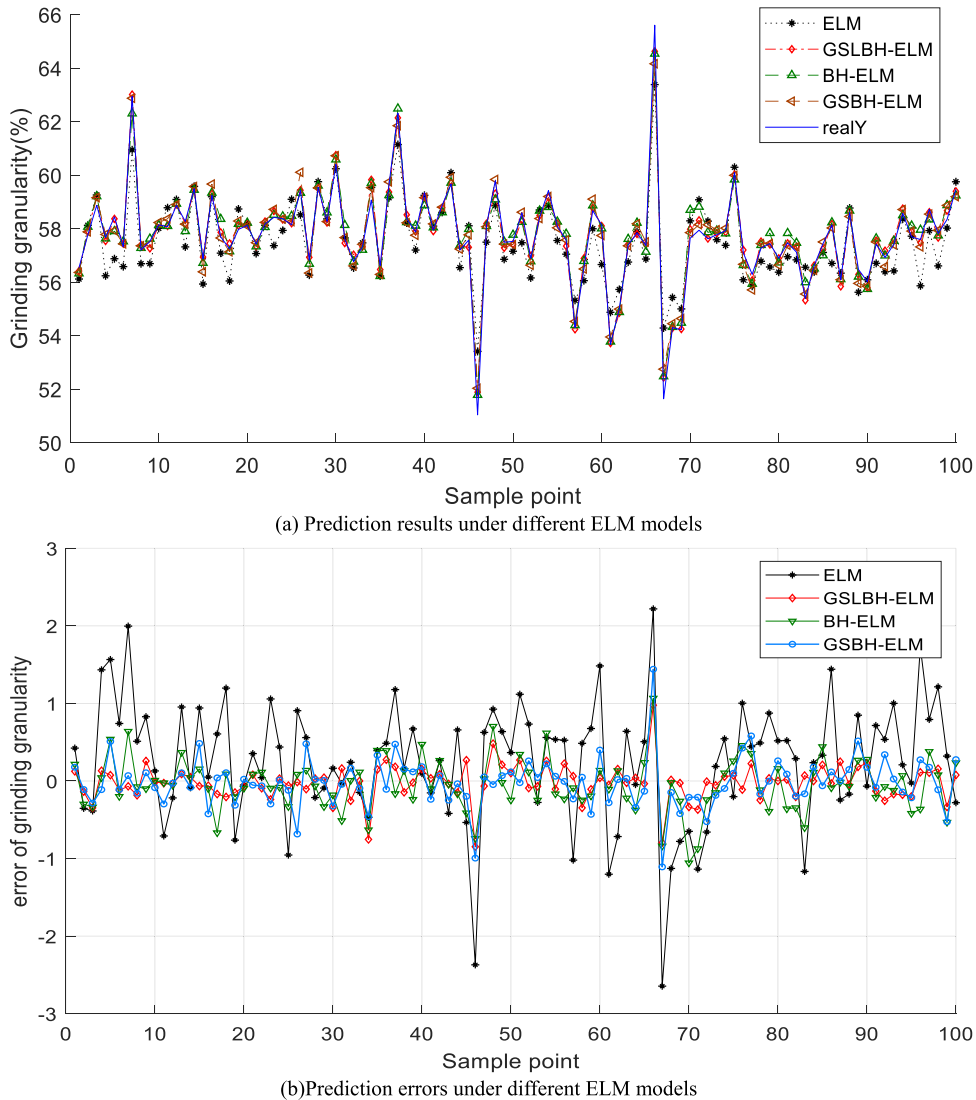

(b)Prediction errors under different ELM models

FIGURE 10. Simulation results under activation function ReLu.

The solution of the Eq (19) can be calculated by:

$$\hat{\beta} = H^+ T' \tag{20}$$

where, H^+ is the Moore-Penrose generalized inverse of the hidden layer output matrix H.

From the above analysis, the weight w and bias b of the ELM model are randomly generated. Only the number of hidden layer neurons, the activation function and parameter β of the hidden layer neurons can be determined. The ELM model training process can be divided into the following steps:

1) Determine the number of neurons, randomly set the connection weight w between the input layer and the hidden layer and the offset value b of the hidden layer neurons.

2) Select an infinitely differentiable function as the activation function of the hidden layer neurons, and then calculate the hidden layer output matrix H.

3) Calculate the output layer weights $\hat{\beta}$: $\hat{\beta} = H^+ T'$.

A large number of experiments show that the ELM model has fast learning speed and strong generalization ability. It can use not only many non-linear functions, but also non-differentiable functions, and even some discontinuous functions can be used as activation functions.

B. IMPROVED BLACK HOLE ALGORITHM

The black hole (BH) algorithm is a new type of natural heuristic algorithm that is inspired by the laws of motion of celestial bodies like "black holes" in the universe. Xie and Wang introduced the golden sine operator and Levy flight operator (GSLBH) in the black hole algorithm, which greatly enhanced the exploaration of the black hole algorithm [36]. The Gold-SA algorithm (Gold-SA) is a mathematical heuristic algorithm with strong global search ability proposed by Tanyildizi *et al.* [39]. The continuous movement of points on a sine function is equivalent to the continuous scanning of points on a unit circle, which is similar to the process of agents
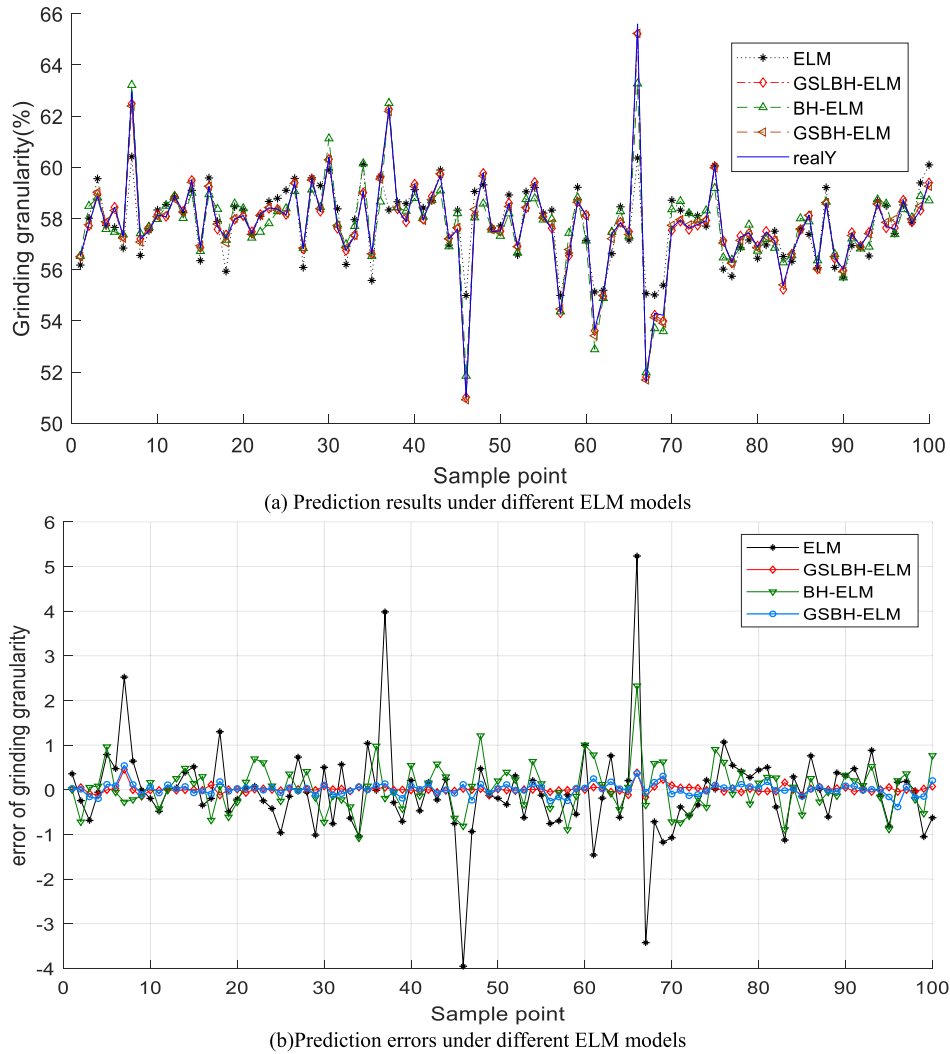
FIGURE 11. Simulation results under activation function sigma.

in the natural heuristic algorithm searching for an optimal solution in a certain range of solution space. However, these agents are not searched according to the standard sine function trajectory, but a golden sine line proposed by Tanyildizi. The shape of this searching path in a two-dimensional space is shown in Fig. 4 [36], and the red curve is the searching path.

In Gold-SA, the agents search the optimal solution according to the golden sine line described as follows.

$$x_i(t+1) = x_i(t) * |\sin(r_1)|$$
$$- r_2 * \sin(r_1) * |m_1 * D - m_2 * x_i(t)| \quad (21)$$

where, $x_i(t+1)$ is the position of agent x_i at the next iteration, $x_i(t)$ is the current position of the agent, r_1 is a random number between $[0, 2\pi]$, r_2 is a random number between $[0, \pi]$, D is the global optimal position in Gold-SA, m_1 and m_2 are coefficients obtained through the golden section method. These coefficients can reduce the search space and improve the search efficiency of particles, and make particles

move from the current position to the target position. This searching mode shown in Eq. (21) may be named as a golden sine operator described as:

$$x_i(t+1) = x_i(t) * |\sin(r_1)|$$
$$- r_2 * \sin(r_1) * |m_1 * D - m_2 * x_i(t)| \quad (22)$$

where, $m_1 = -\pi + (1 - \tau) * \pi$, $m_2 = -\pi + \tau * \pi$. Here m_1 and m_2 are specified as constants. The remaining parameters in Eq. (22) have the same meanings as the parameters in Eq. (21).

Levy flight was proposed by French mathematician Paul Levy. Many scholars have since found that the predation and hunting methods of most creatures are in line with the Levy flight process. It is generally believed that Levy flight can improve the efficiency and accuracy of predation [40]–[42]. The trajectory of Levy flight in a two-dimensional space is shown in Fig. 5. The agents moved 1000 times in this two-dimensional space. The mechanism of Levy flight is
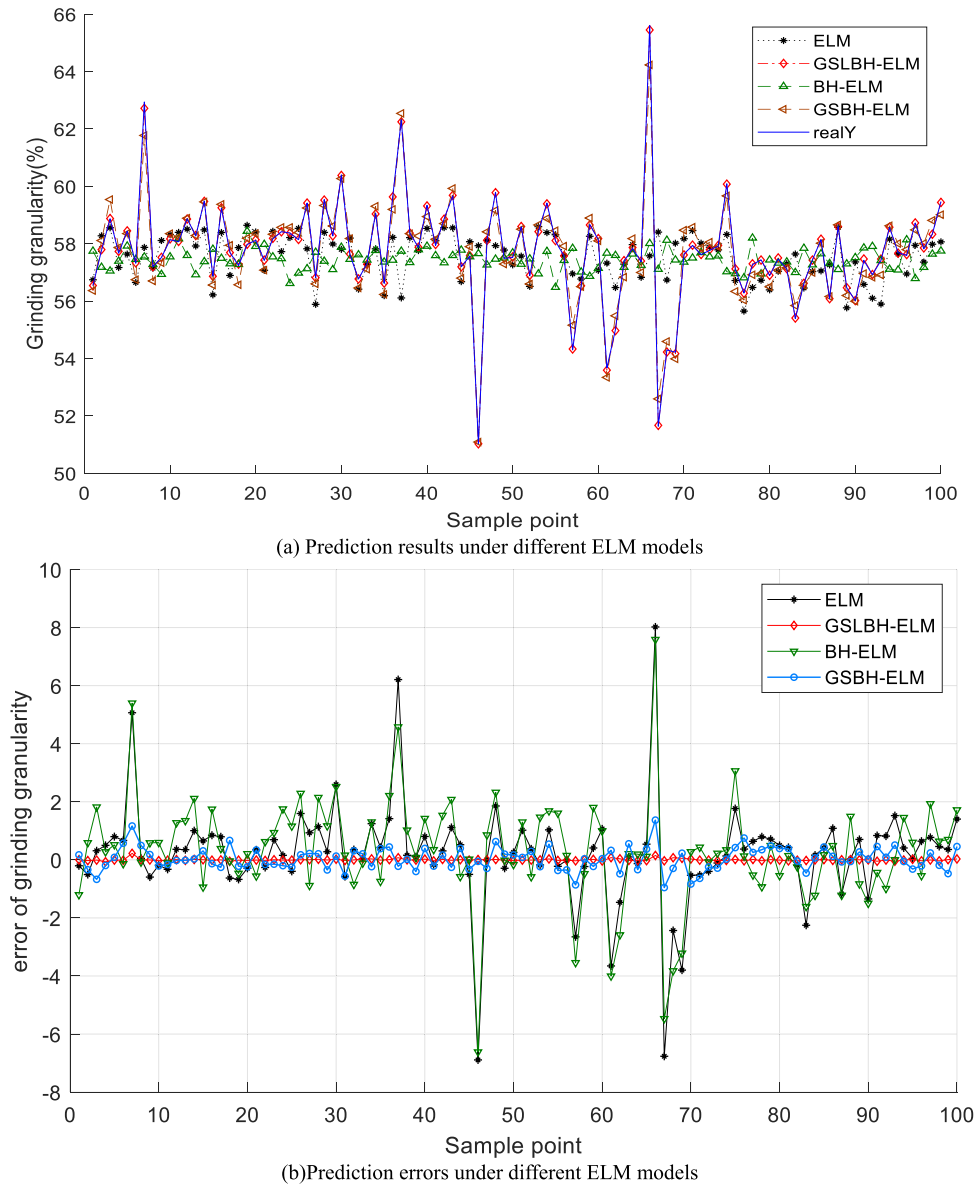
W. Xie *et al.*: ELM Soft-Sensor Model With Different Activation Functions on Grinding Process Optimized by Improved BH Algorithm

IEEE *Access*

(a) Prediction results under different ELM models

(b)Prediction errors under different ELM models

FIGURE 12. Simulation results under activation function sin.

long-term and short-distance migration, which increases the diversity of the population. The organic combination of short-distance and long-distance jumps greatly enhances the exploaration of the population [36]. By reducing the motion step of the Levy flight, the Levy flight becomes a local search operator with strong exploitation, and the improved Levy flight operator can be expressed as:

$$x_i(t+1) = x_i(t) + 0.015 * \text{sign}(rand - 1/2) \otimes s \quad (23)$$

where, 0.015 is the step size control factor, which can ensure that agents search in a small area, $\text{sign}(rand - 1/2)$ is three random numbers, whose values are $-1, 0, 1$, s is the step size vector for Levy flight, and $\otimes$ represents the multiplication among elements. The procedure of the improved black hole algorithm can be described as follows.

Step 1: Initialize the population. Let the initial population number be N. The agents position in the population is initialized by Eq. (24).

$$X_i(t+1) = X_{min} + rand \times (X_{max} - X_{min}) \quad (24)$$

where, $X_i(t+1)$ represents the position of the next iteration of the i-th agent, X_{max} and X_{min} represent the upper and lower boundaries of the particle, $rand$ represents a random number between [0, 1] so that agents can be uniformly distributed in the solution space.

Step 2: Find out the current optimal position. Calculate the fitness function value of the agent based on its current position. Let the function optimized by the improved algorithm be $f(x)$, then the fitness of the agent can be calculated by:

$$fit_i = f(x_i) \quad (25)$$
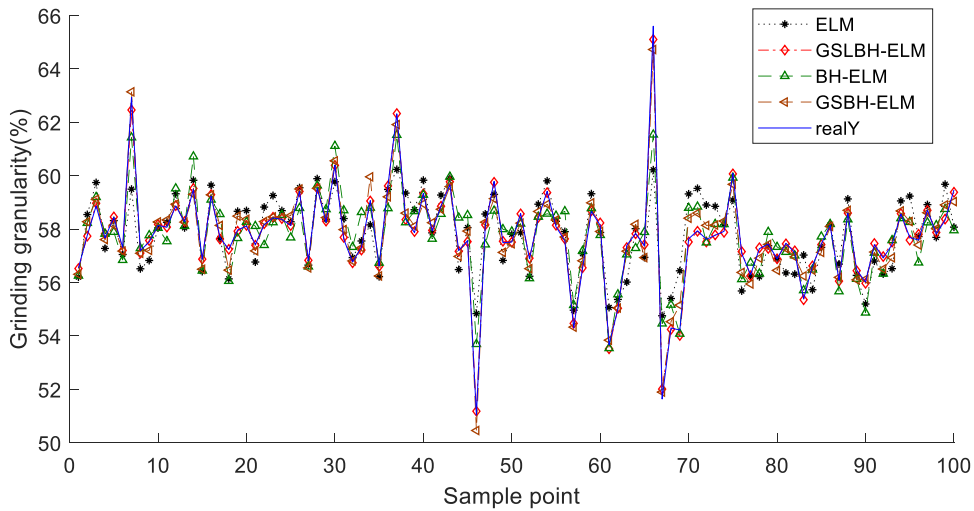
IEEE *Access*

W. Xie *et al.*: ELM Soft-Sensor Model With Different Activation Functions on Grinding Process Optimized by Improved BH Algorithm

(a) Prediction results under different ELM models
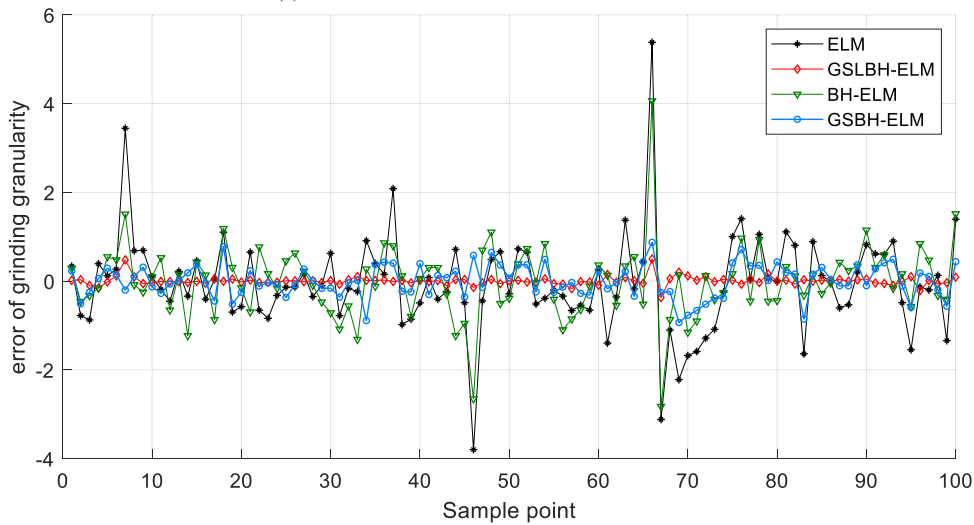
(b)Prediction errors under different ELM models

FIGURE 13. Simulation results under activation function thah.

Select the minimum value of the fitness value from these agents as the current optimal value G_{gest}. So the position corresponding to G_{gest} is the current optimal position X_{gest} for the agent search.

Step 3: Search the optimal solution. Let the agents move to the current optimal solution according to the black hole algorithm search principle shown in Eq. (26).

$$x_i(t+1) = x_i(t) + rand \times (X_{gest} - x_i(t)) \quad (26)$$

Step 4: Perform the ''swallow'' operation of BH algorithm. In the black hole algorithm, an ''event horizon'' R_s is built around the agent corresponding to the optimal position X_{gest}, which is defined as:

$$R_s = \frac{|G_{gest}|}{\sum_{i=1}^{N} |fit_i|} \quad (27)$$

The distance D_i from other agents in solution space to X_{gest} is expressed by Euclidean distance:

$$D_i = \sqrt{(X_{gest} - X_i)^2} \quad (28)$$

When the distance from agent i to X_{gest} is greater than D_i, the position of agent i does not change. When the distance from particle i to X_{gest} is less than D_i, the agent ''disappears'' from the solution space. In order to keep the population of agents as N, a agent is randomly generated at any position in the solution space.

Step 5: Perform Levy flight operation. Let the agents in the solution space perform the Levy flight search operation according to Eq. (23).

Step 6: Perform golden sine operation. Let the agents in the solution space perform the golden sine search operation according to Eq. (22).
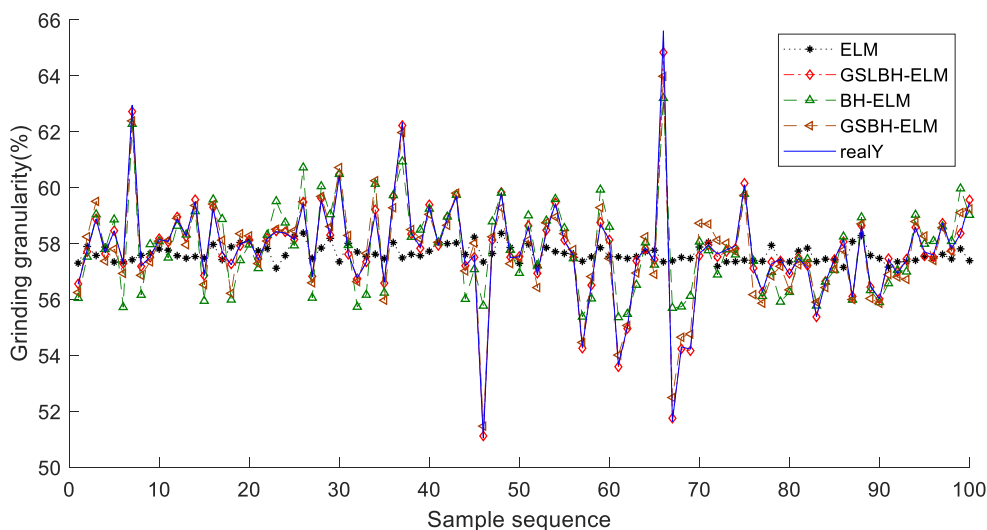
W. Xie *et al.*: ELM Soft-Sensor Model With Different Activation Functions on Grinding Process Optimized by Improved BH Algorithm

IEEE *Access*

(a) Prediction results under different ELM models
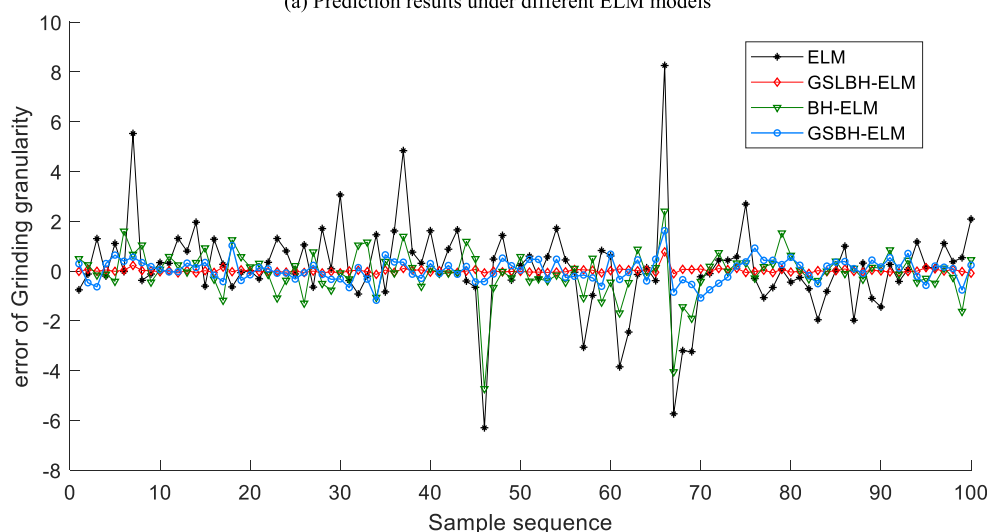
(b)Prediction errors under different ELM models

FIGURE 14. Simulation results under activation function Morlet.

Step 7: Perform from Step 2 to Step 7 in a loop until the iteration conditions are met.

Step 8: Output the optimal position X_{gest} and solution G_{gest} found by the improved BH algorithm.

C. IMPROVED BLACK HOLE ALGORITHM TO OPTIMIZE ELM NEURAL NETWORK

1) IMPROVED BLACK HOLE ALGORITHM TO ADJUST STRUCTURAL PARAMETERS OF ELM NEURAL NETWORK

Since the weight w and bias b of the ELM model are randomly generated and the training method of the ELM model is relatively simple, this model can only improve prediction accuracy by increasing the number of hidden layer neurons and the prediction accuracy is limited. This paper attempts to use the improved black hole algorithm to find the minimum value of the ELM model training error function so as to obtain

the weight w and bias b corresponding to the minimum value. This paper also try to find some activation functions that can improve the prediction accuracy of ELM model.

The training method of the ELM model is relatively simple. The weight w and bias b between the input layer and the hidden layer generally take random numbers between [0,1]. The weight β between the output layers of the hidden layer is obtained by obtaining the least square solution of $\min_{j=1} \|H\beta - T'\|$. Let the ELM model be a non-linear function $Y_p = F(x)$, and at this time the number of ELM neurons is the same as the number of ELM model neurons in Section A of Part III. Let the number of training samples be M, a_i be one of the M samples and the real output corresponding to this sample be y_i. Let the row vector consisting of the structural parameters w and b between the input layer and the hidden layer in the ELM model be $C = [w_1 \cdots b_l]_{1 \times (l \times n + l)}$, then

IEEE Access

W. Xie *et al.*: ELM Soft-Sensor Model With Different Activation Functions on Grinding Process Optimized by Improved BH Algorithm

(a) Fitness function curve of ELM model with arctan function optimized by black hole algorithm.

(b) Fitness function curve of ELM model with hardlim function optimized by black hole algorithm.

(c) Fitness function curve of ELM model with ReLu function optimized by black hole algorithm.

(d) Fitness function curve of ELM model with sigma function optimized by black hole algorithm.

(e) Fitness function curve of ELM model with sin function optimized by black hole algorithm.

(f) Fitness function curve of ELM model with thah function optimized by black hole algorithm.

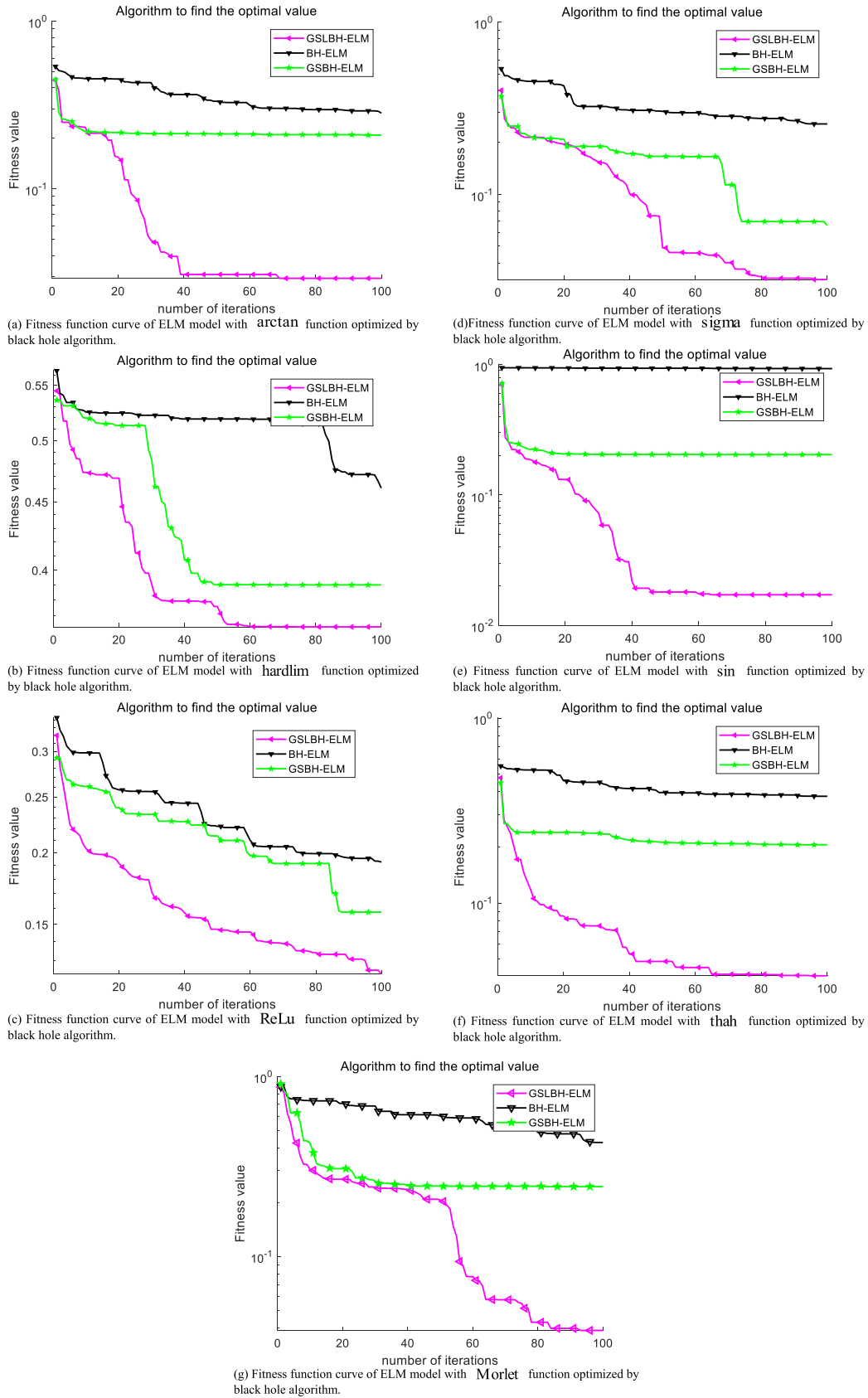(g) Fitness function curve of ELM model with Morlet function optimized by black hole algorithm.

FIGURE 15. Fitness function curve of ELM model with different activation function optimized by black hole algorithm.

W. Xie *et al.*: ELM Soft-Sensor Model With Different Activation Functions on Grinding Process Optimized by Improved BH Algorithm

IEEE *Access*

TABLE 6. Quantitative evaluation index of grinding granularity prediction.

| Function	Training method	MPE	MNE	R^2	RMSE
sigma	ELM	3.9538	-5.2345	0.68132	1.0436
	BH-ELM	1.0818	-2.3276	0.91939	0.52487
	GSBH-ELM	0.38807	-0.54337	0.99494	0.13154
	GSLBH-ELM	0.16684	-0.45189	0.99803	0.082005
sin	ELM	5.5441	-6.1817	0.32139	1.5229
	BH-ELM	6.0721	-7.4879	-0.023459	1.8702
	GSBH-ELM	0.84289	-1.2929	0.95542	0.39035
	GSLBH-ELM	0.13007	-0.4389	0.99857	0.069983
arctan	ELM	3.9055	-5.6343	0.63543	1.1162
	BH-ELM	0.94278	-2.0599	0.91819	0.52877
	GSBH-ELM	0.79683	-1.1755	0.95938	0.37261
	GSLBH-ELM	**0.12647**	**-0.23178**	**0.99903**	**0.057468**
ReLu	ELM	2.645	-2.2189	0.79607	0.83484
	BH-ELM	1.0561	-1.0672	0.96466	0.34755
	GSBH-ELM	1.1076	-1.4405	0.97011	0.3196
	GSLBH-ELM	0.84765	-0.97681	0.98386	0.23484
hardlim	ELM	4.9055	-6.48	0.47964	1.3336
	BH-ELM	4.5306	-2.4433	0.71865	0.98058
	GSBH-ELM	3.358	-3.172	0.76998	0.88663
	GSLBH-ELM	2.7085	-3.3611	0.85519	0.70351
Morlet	ELM	6.2967	-8.2575	0.04062	1.8107
	BH-ELM	4.7295	-2.4039	0.73749	0.94718
	GSBH-ELM	1.1736	-1.6224	0.94022	0.45202
	GSLBH-ELM	0.13384	-0.76905	0.99726	0.096744
thah	ELM	3.7979	-5.379	0.64094	1.1078
	BH-ELM	2.8242	-4.0612	0.79757	0.83176
	GSBH-ELM	0.93325	-0.87376	0.96044	0.36768
	GSLBH-ELM	0.37048	-0.49727	0.99698	0.10154

an error function about C can be constructed as:

$$error = \frac{1}{M} \sum_{i=1}^{M} (F(a_i, C) - y_i)^2 \qquad (29)$$

The improved black hole algorithm can be used to optimize the defined *error* function. The algorithm execution steps have been described in Section B of Part III, and the fitness function at this time is the *error* function. The argument of the *error* function is C. According to the structure of the ELM model, we can know that the number of parameters in C is $(l \times n + l)$, so the dimension of the search agent is $(l \times n + l)$. The optimal value finally found by the algorithm is the minimum value of the *error* function, and the parameters w and b corresponding to the minimum value are reasonable model

parameters. The prediction error at this time has reached the minimum.

2) IMPROVED BLACK HOLE ALGORITHM TO ADJUST STRUCTURAL PARAMETERS OF ELM NEURAL NETWORK

The activation function $g(x)$ of the ELM model is generally *sigma* function. According to the ELM model theory, the activation function $g(x)$ in the neural network model can be any continuous function [27]. This paper selects 7 different continuous functions as the activation function of the ELM model for carry out simulation experiments. The expressions and function images of these seven functions are shown in Table 4 and Fig. 6, respectively.
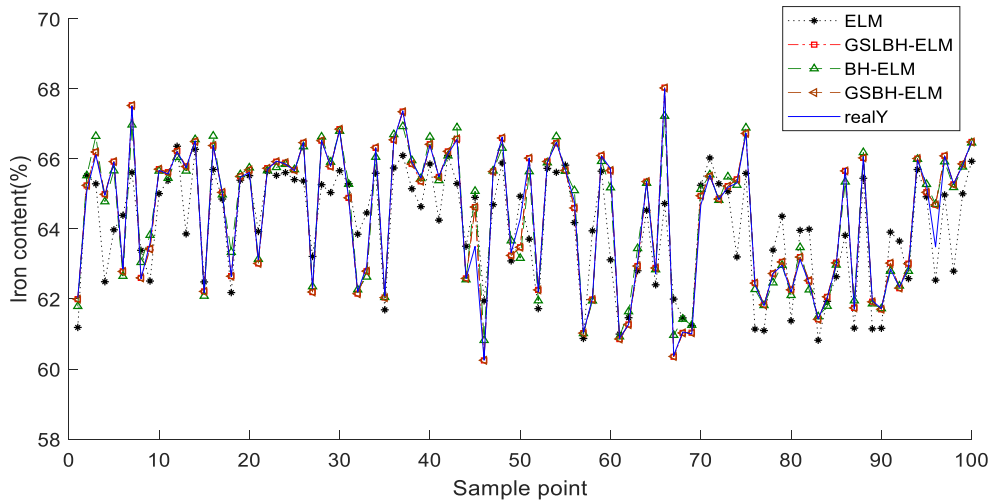
(a) Prediction results under different ELM models
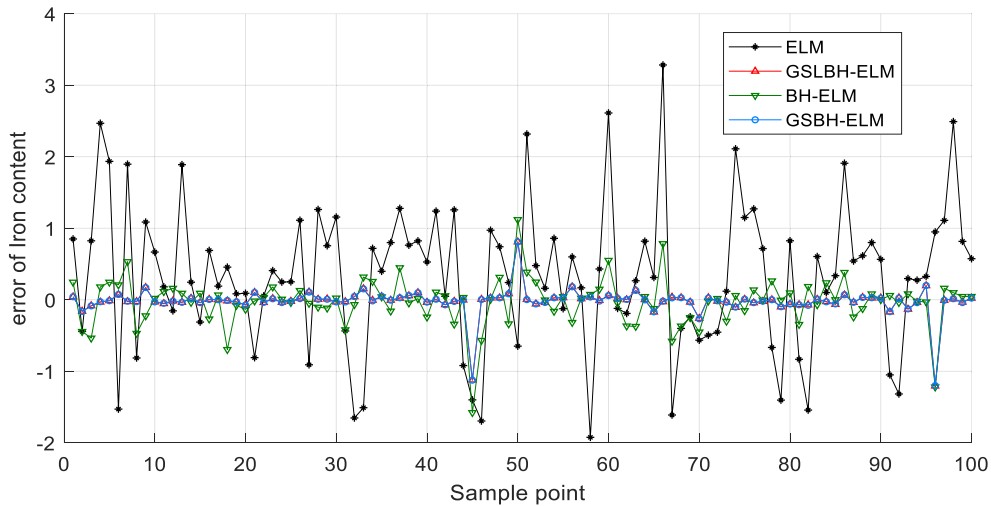
(b)Prediction errors under different ELM models

FIGURE 16. Simulation results under activation function arctan.

One thing needs special explanation here. When the activation function $g(x)$ is a wavelet basis function *Morlet*, the function *Morlet* can be expressed as:

$$g(j) = g_j \left(\frac{\sum\limits_{i=1}^{k} w_{ij} x_i - b_j}{a_j} \right) \qquad j = 1, 2, ..., l \qquad (30)$$

where, a_j is the scaling factor of the *Morlet* function, which is also the parameter that needs to be adjusted. When the algorithm optimizes the ELM model, it is necessary to adjust not only w and b, but also the scaling factor a. The dimension of the searching agents becomes $(l \times n + 2l)$. For other activation functions, the dimensions of the search particle should remain the same as those specified in the previous section.

3) ALGORITHM FLOWCHART
The pseudo code of the ELM model optimized by the improved black hole algorithm is shown as follows.

Determine the number of neurons in the ELM model.

Establish the error function in Eq. (29) based on the grinding process and the determine dimensions of the space.

For $i = 1 : iter$

Initialize the position of the population according to Eq. (24)($C_i = [w_1, ..., b_l]_{1 \times (l \times n + l)}$).

Input the population C of the coordinate and the training sample set M into the *error* function of Eq. (29) and calculate the its fitness value.

Perform the black hole search operation according to the Eq. (26).

Perform the "swallow" operation of the black hole algorithm.

Perform the Levy flight search operation according to Eq. (23).
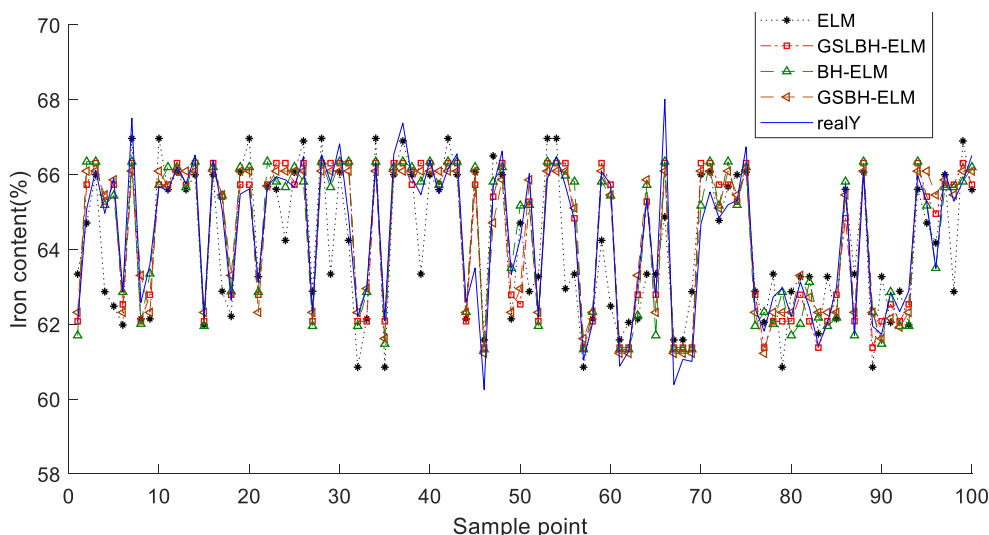
W. Xie *et al.*: ELM Soft-Sensor Model With Different Activation Functions on Grinding Process Optimized by Improved BH Algorithm

IEEE *Access*

(a) Prediction results under different ELM models

(b)Prediction errors under different ELM models

FIGURE 17. Simulation results under activation function hardlim.

Perform golden sine search operation according to Eq. (22).

End For

Get the optimal position $C_{best} = [w_{1best}, \ldots, b_{lbest}]_{1 \times (l \times n + l)}$ corresponding to the optimal fitness value of *error* function.

Determine the parameter β between the hidden layer and the output layer by w_{best}, b_{best} and Eq. (20).
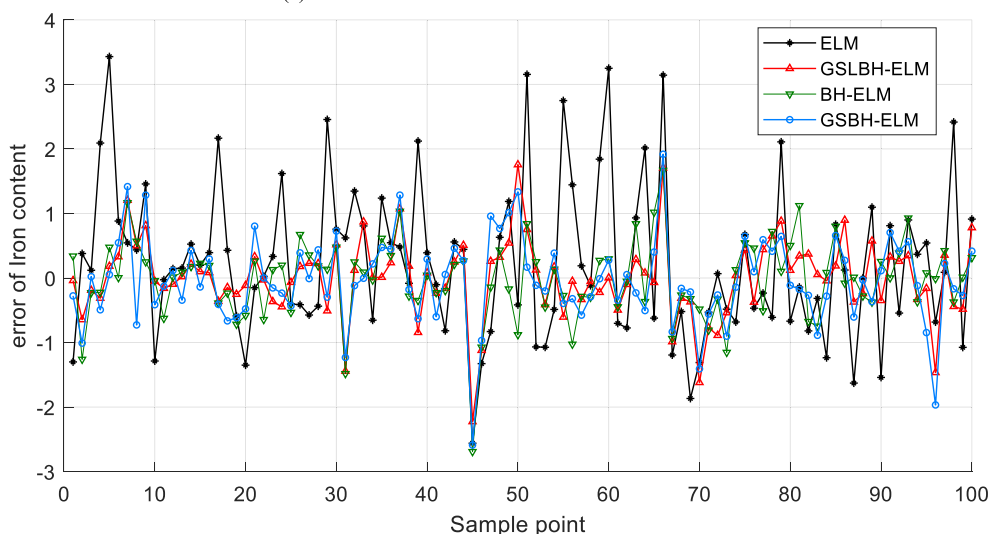
Input the test sample into the trained ELM model to obtain the test result.

Evaluate prediction results.

The algorithm flowchart of the ELM model optimized by the improved black hole algorithm is shown in Fig. 7.

IV. SIMULATION EXPERIMENT AND RESULT ANALYSIS

In this paper, simulation experiments are carried out with the background of grinding process of a beneficiation plant. Hematite is the main processing material in this beneficiation plant. The technique principle of the hematite grinding process has been introduced in details in the second section. After grinding and classifying process, the size of the ore will meet the specified production requirements. Seven kinds of ELM neural network soft-sensor models (GSLBH-ELM) based on improved black hole algorithm with different activation functions will be established to predict the iron content and grinding granularity that reach 200 mesh (0.0074mm) content during the grinding process. In order to clearly indicate the prediction accuracy of each soft-sensor model, four quantitative indicators are selected to evaluate the prediction accuracy. The calculation of these four indicators is shown in Table 5, where $\hat{y}$ represents the predicted value of the soft-sensor model and y represents the actual output of the test samples. Among these indicators, the first three indicators *MPE*, *MNE*, *RMSE* are used to reflect the deviation between the predicted value and the true value. R^2 is adopted

IEEE Access

W. Xie *et al.*: ELM Soft-Sensor Model With Different Activation Functions on Grinding Process Optimized by Improved BH Algorithm

(a) Prediction results under different ELM models

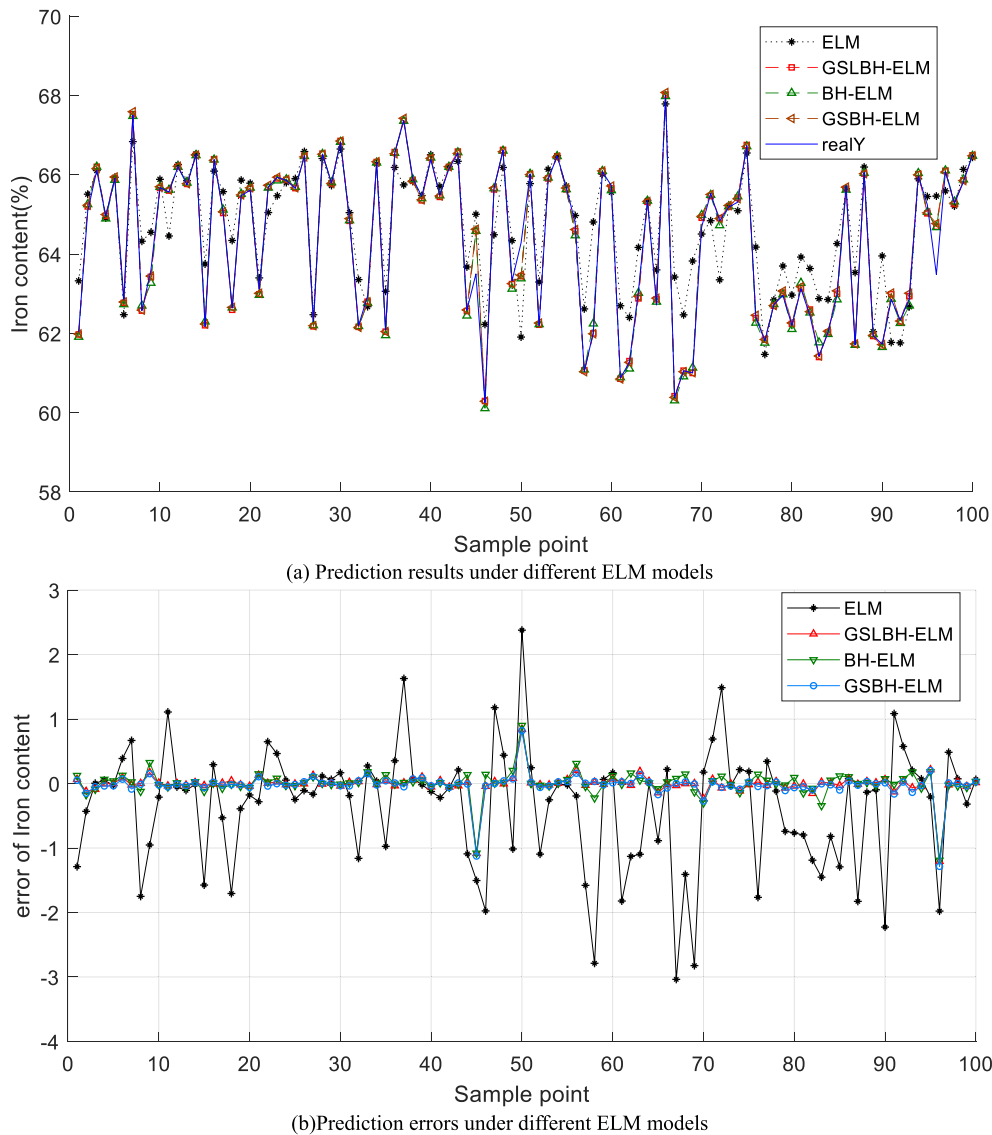(b)Prediction errors under different ELM models

FIGURE 18. Simulation results under activation function ReLu.

to evaluate the degree of fit of the soft-sensor model. This indicator is a number between [0, 1], and the larger its value, the better the prediction curve fits.

The production data of grinding process are collected during the one-month, and 1000 sets of experimental samples are selected. These sample sets are divided into 900 training samples and 100 test samples. The training samples are used to train the ELM model, and the test samples are used to test whether the prediction accuracy of the soft-sensor model is accurate. The input variables of the ELM model are reduced to 6 by KPCA dimensionality reduction method. Therefore, the number of input neurons in ELM is 6, the number of hidden layer neurons is 8, and the number of output layer neurons is 1, which is the iron content or grinding granularity.

In order to explore which of the two search operators has a significant effect on the black hole algorithm, this paper

compares the prediction results of the original ELM model, ELM optimized by black hole algorithm (BH-ELM), ELM model optimized by black hole algorithm with golden sine operator (GSBH-ELM), and ELM model optimized by black hole algorithm with golden sine operator and Levy flight operator (GSLBH-ELM) to illustrate the optimization ability of GSLBH on the ELM model. The population of these three types of black hole algorithms is set to 30 and the number of iterations is set to 100.

A. PREDICTION RESULTS AND ANALYSIS ON GRINDING GRANULARITY

Fig. 8 to Fig. 15 reflect the prediction results of the ore granularity by adopting four soft-sensor models (ELM, BH-ELM, GSBH-ELM, and GSLBH-ELM) under 7 different activation functions. All figures of Fig. 8 to Fig. 15 are divided into
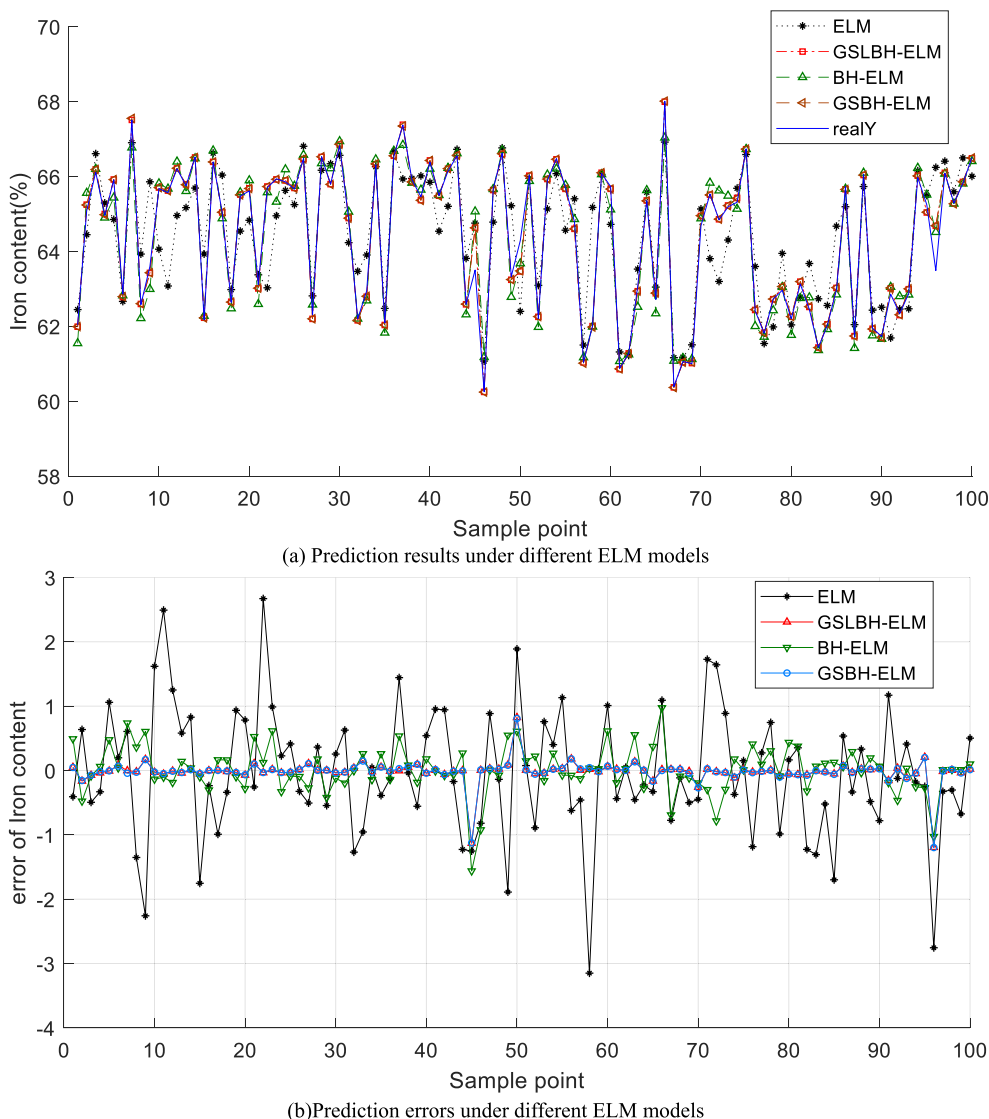
W. Xie *et al.*: ELM Soft-Sensor Model With Different Activation Functions on Grinding Process Optimized by Improved BH Algorithm

IEEE *Access*

(a) Prediction results under different ELM models

(b)Prediction errors under different ELM models

FIGURE 19. Simulation results under activation function sigma.

two sub-figures (a), (b) to show the predictive results and predictive errors,Fig. 15 shows the fitness convergence curves under different algorithms to optimize ELM soft-sensor for predicting grinding granularity. The quantitative indicators of seven soft-sensor models for predicting ore granularity are listed in Table 6, which are the average of 10 simulation experiments. Seen from the simulation results in Fig. 8 to Fig. 14 and the numerical results in Table 5, when predicting the grinding granularity, after the optimization algorithm adjusts the parameters of the ELM model, the prediction accuracy of all ELM models have been improved. But for different activation functions, the optimization effect of the algorithm is different, and exploration abilities of BH, GSBH, and GSLBH are also different. It can be seen form Fig. 15 that the optimization ability of GSLBH is significantly better than GSBH, and the optimization ability of GSBH is significantly better than BH, which leads to the results of the 7 groups of

experiments that the prediction accuracy of GSLBH-ELM is significantly better than GSBH-ELM and BH-ELM. At the same time, Some phenomena can also be observed. The global search ability of the pure BH algorithm is poor, and it is easy to fall into the local optimal solution, which leads to the poor fitting effect of some BH-ELM, and some results have worse prediction accuracy than the original ELM model, such as the BH-ELM model where the activation function is sin. When the activation function is different, the optimization results of the algorithm for the ELM model are also different. From the quantitative indicators listed in Table 5, it can be seen that GSLBH-ELM has better optimization results for ELM models whose activation functions are sigma, sin, arctan, and Morlet. When the activation function is arctan, the prediction accuracy of GSLBH-ELM is the highest. From all the above results, the golden sine operator improves exploration ability of the BH algorithm, and the Levy flight operator
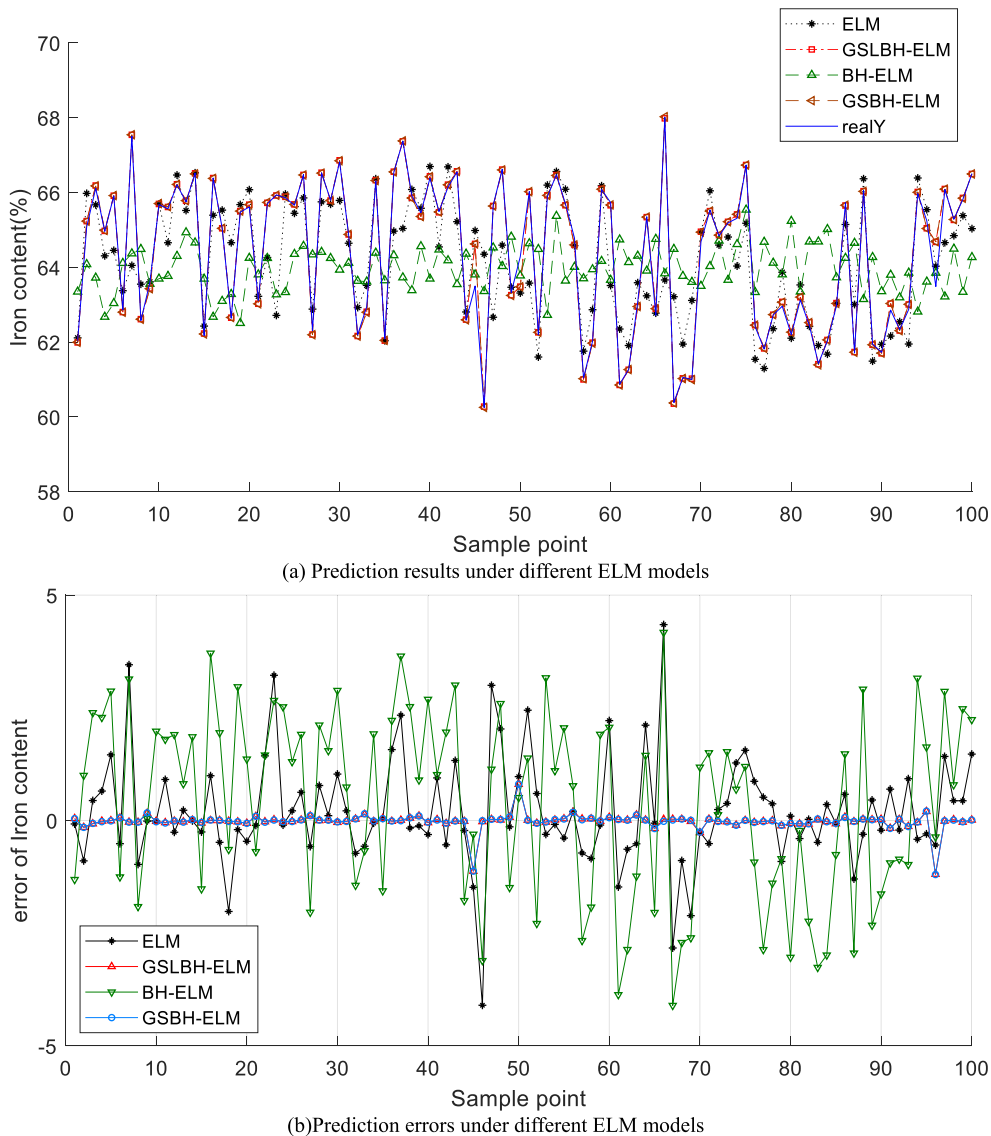
IEEE *Access*

W. Xie *et al.*: ELM Soft-Sensor Model With Different Activation Functions on Grinding Process Optimized by Improved BH Algorithm

(a) Prediction results under different ELM models

(b)Prediction errors under different ELM models

FIGURE 20. Simulation results under activation function sin.

further enhances exploitation ability of the BH algorithm in a certain range. Two operators make the prediction ability of GSLBH-ELM for grinding granularity greatly enhanced.

B. PREDICTION RESULTS AND ANALYSIS ON ORE IRON CONTENT

Fig. 16 to Fig. 23 reflect the prediction results of the ore iron content by adopting four soft-sensor models (ELM, BH-ELM, GSBH-ELM, and GSLBH-ELM) under 7 different activation functions. All figures are divided into two sub-figures (a), (b) to show the predictive results and predictive errors, Fig. 23 shows the fitness convergence curves under different algorithms to optimize ELM soft-sensor for predicting the ore iron content. The quantitative indicators of seven soft-sensor models for predicting ore granularity are

listed in Table 7, which are the average of 10 simulation experiments.

From the simulation results in Fig. 16 to Fig. 22 and the numerical results in Table 6, when predicting the ore iron content, after the optimization algorithm adjusts the parameters of the ELM model, the prediction accuracy of all ELM models have been improved. But for different activation functions, the optimization effect of the algorithm is different, and exploaration ability of BH, GSBH, and GSLBH are also different. It can be seen form Fig. 23 that the optimization ability of GSLBH is similar to that of GSBH, and the optimization ability of GSLBH and GSBH is significantly better than BH, which led to the results of the 7 groups of experiments showing that the prediction accuracy of GSLBH-ELM and GSBH-ELM was significantly better than that of BH-ELM.
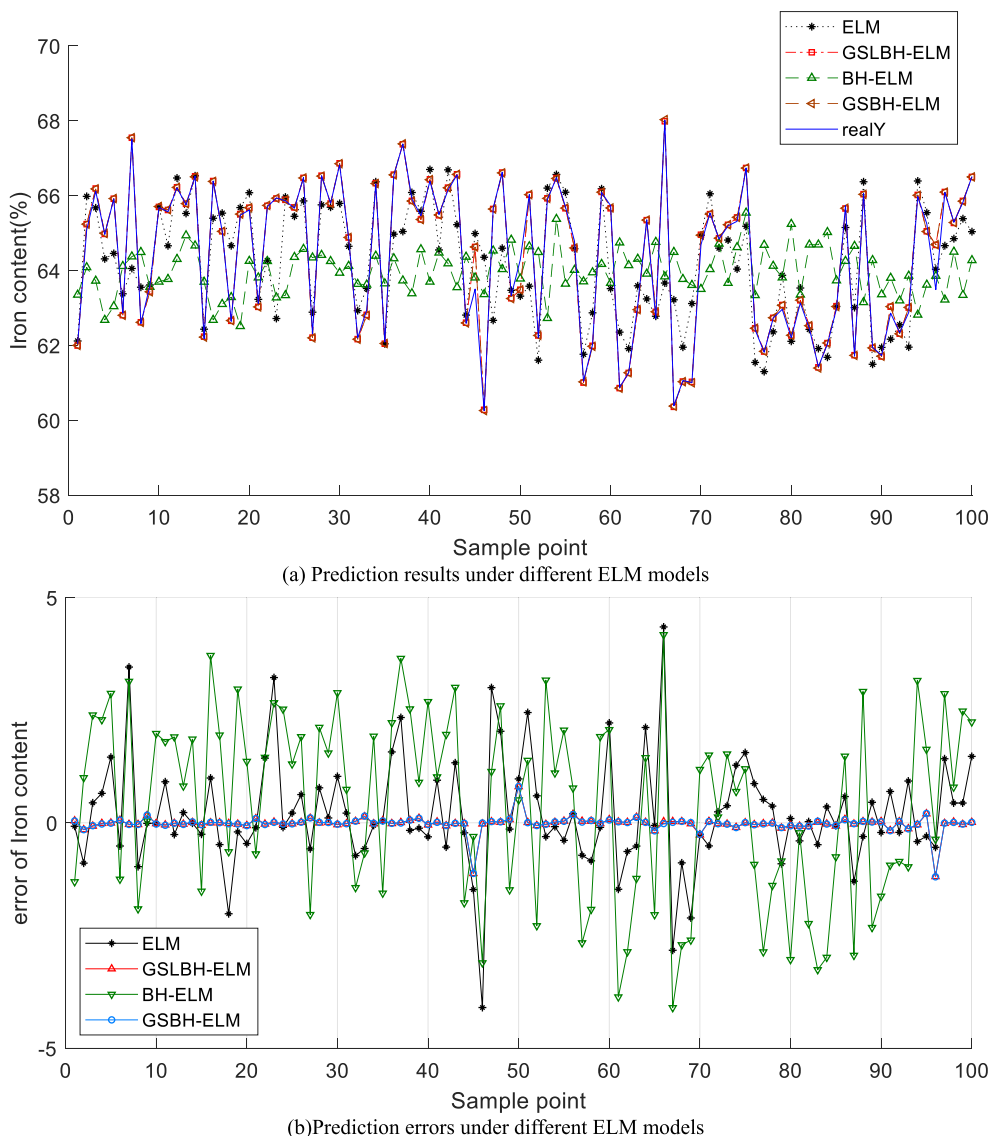
W. Xie *et al.*: ELM Soft-Sensor Model With Different Activation Functions on Grinding Process Optimized by Improved BH Algorithm

IEEE *Access*

(a) Prediction results under different ELM models

(b)Prediction errors under different ELM models

FIGURE 21. **Simulation results under activation function thah.**

At the same time, some phenomena can also be observed, the global search ability of the pure BH algorithm is poor, and it is easy to fall into the local optimal solution, which results in many BH-ELM fitting effects are poor, and the prediction accuracy is even worse than the original ELM model, such as the BH-ELM model where the activation function are sin and thah. From the quantitative indicators in Table 6, it can be seen that when the activation function is different, the optimization results of the algorithm for the ELM model are also different. In addition to the hardlim function, the GSLBH-ELM is more ideal for ELM models with other activation functions, and the prediction accuracy is higher. When the activation function is sin, GSLBH-ELM has the highest prediction accuracy. Based on the above experimental results, when predicting the iron ore content, the golden sine operator plays a leading role in the optimization performance of the BH algorithm.

Levy flight has a small improvement in the optimization ability of the BH algorithm. This improvement has greatly enhanced the ability of GSLBH-ELM to predict the iron ore content.

Combining all the above experimental results can lead to such a conclusion, although the optimization effect of the improved black hole algorithm on ELM models with different activation functions will be different, after the parameters of the ELM model are adjusted using the algorithm, the prediction ability of the ELM model for the production index of the grinding process will be greatly improved.

C. COMPUTATIONAL COST ANALYSIS
In order to investigate whether the implementation efficiency of the proposed ELM meets the actual industrial environment, this paper tests the training time of various ELM models.
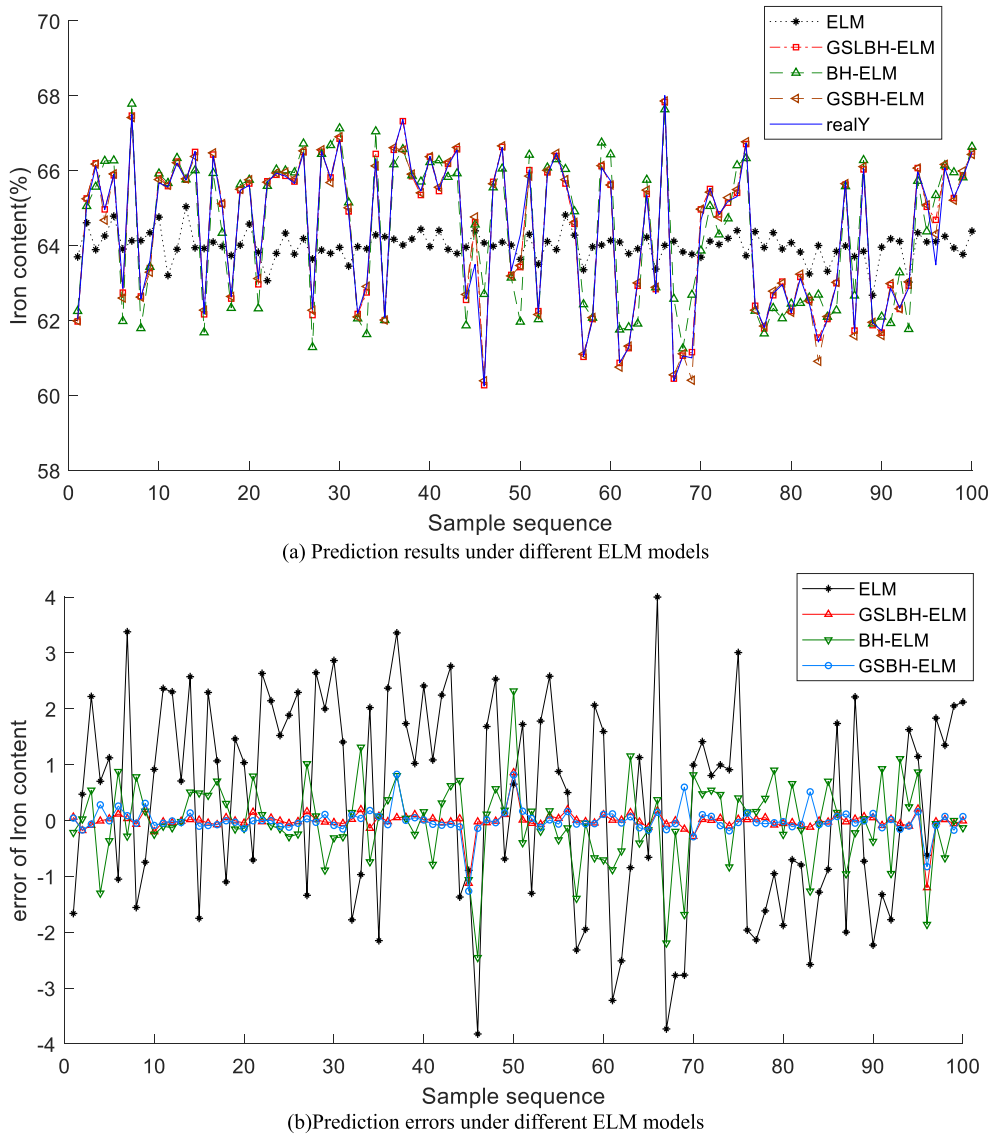
IEEE *Access*

W. Xie *et al.*: ELM Soft-Sensor Model With Different Activation Functions on Grinding Process Optimized by Improved BH Algorithm

(a) Prediction results under different ELM models

(b)Prediction errors under different ELM models

FIGURE 22. Simulation results under activation function Morlet.

The test results are shown in Table 8 and Table 9, and the time unit is seconds (s).

Seen from the experimental results, the length of the training time of the ELM model predicting the two production indicators is very similar. In the case of different excitation functions, the training time of the simple ELM model is basically the same, which are very short. When an optimization algorithm is used to adjust the structural parameters of the ELM model, the training time of the ELM model whose excitation function is Morlet function is obviously longer than the training time of other ELM models, because the ELM model not only needs to optimize its own structural parameters, it also needs to optimize the scaling factors a_j mentioned in Section C of Part III, which will increase the solution space dimension of the optimization algorithm and

make the search time of the optimization algorithm longer. Because the execution process of the two operators takes time, the training time of GSBLB-ELM is longer than that of GSBH-ELM, and the training time of GSBH-ELM is longer than that of BH-ELM. According to the experimental results in Table 8 and Table 9, the training time of GSLBH-ELM with Morlet function is the longest, and the training time is 3.053s and 3.1265s respectively. In the process of industrial soft-sensing modeling, the training time of the model is generally not allowed to exceed 60s, and the calculation cost of GSLBH-ELM is still within the specified range. Therefore, the improved ELM soft-sensing model proposed in this paper is suitable for monitoring the production indicators of the grinding process in actual industrial production.

W. Xie *et al.*: ELM Soft-Sensor Model With Different Activation Functions on Grinding Process Optimized by Improved BH Algorithm

IEEE*Access*

(a) Fitness function curve of ELM model with arctan function optimized by black hole algorithm.

(b) Fitness function curve of ELM model with hardlim function optimized by black hole algorithm.

(c) Fitness function curve of ELM model with ReLu function optimized by black hole algorithm.

(d) Fitness function curve of ELM model with sigma function optimized by black hole algorithm.

(e) Fitness function curve of ELM model with sin function optimized by black hole algorithm.

(f) Fitness function curve of ELM model with thah function optimized by black hole algorithm.

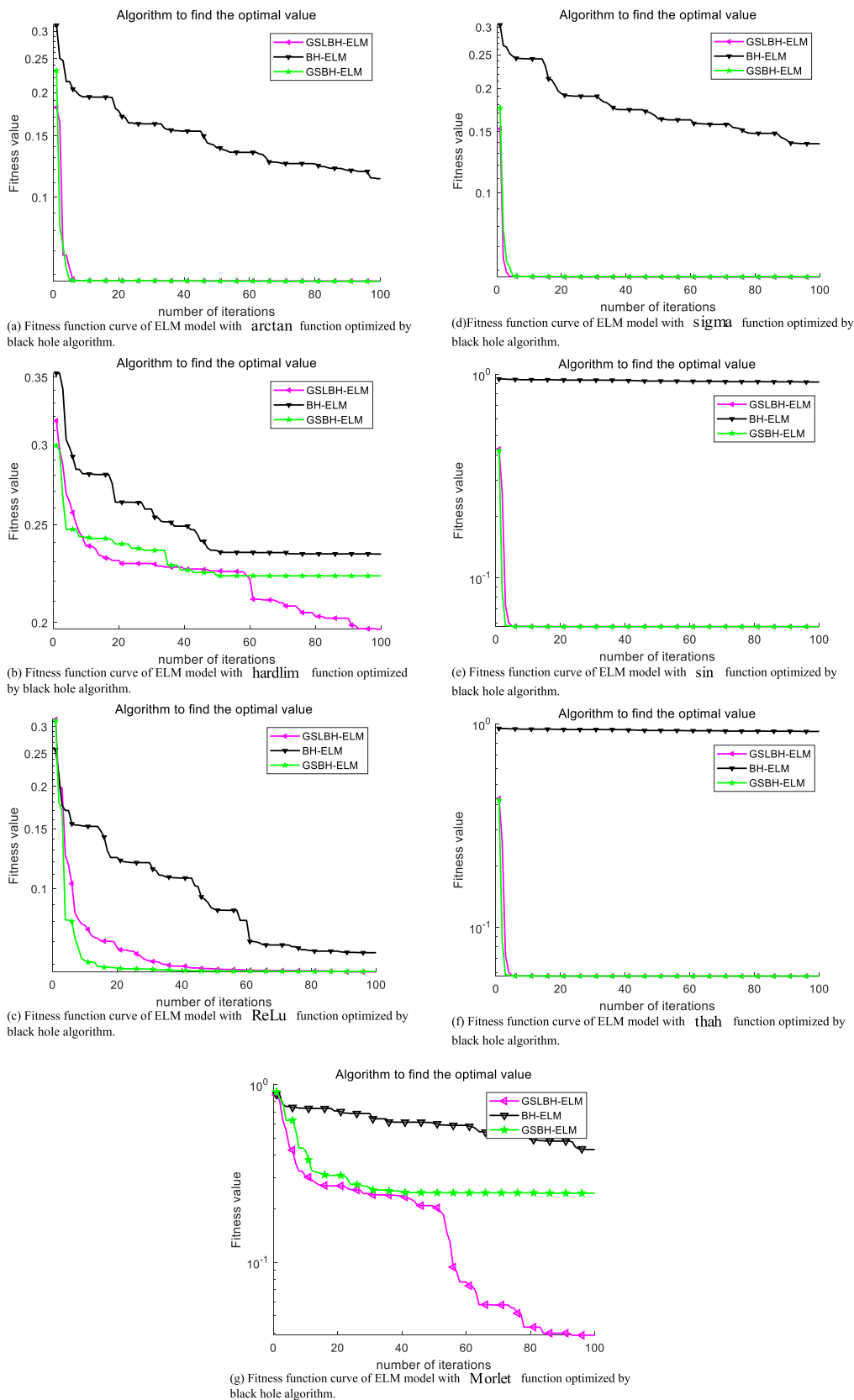(g) Fitness function curve of ELM model with Morlet function optimized by black hole algorithm.

FIGURE 23. Fitness function curve of ELM model with different activation function optimized by black hole algorithm.

IEEE Access

W. Xie *et al.*: ELM Soft-Sensor Model With Different Activation Functions on Grinding Process Optimized by Improved BH Algorithm

TABLE 7. Quantitative evaluation index of grinding granularity prediction.

Function	Training method	MPE	MNE	R^2	RMSE
sigma	ELM	3.1497	-2.6709	0.73695	0.99935
	BH-ELM	1.5617	-0.97375	0.96423	0.36852
	GSBH-ELM	1.1977	-0.80819	0.98994	0.19543
	GSLBH-ELM	1.1953	-0.82313	0.98987	0.19613
sin	ELM	4.0993	-4.346	0.60094	1.2309
	BH-ELM	4.1035	-4.1729	-0.16059	2.0991
	GSBH-ELM	1.1991	-0.80636	0.98995	0.1953
	GSLBH-ELM	**1.1929**	**-0.80278**	**0.99857**	**0.19483**
arctan	ELM	1.9238	-3.283	0.69404	1.0778
	BH-ELM	1.5774	-1.1204	0.96966	0.33938
	GSBH-ELM	1.2001	-0.80859	0.9899	0.19579
	GSLBH-ELM	1.2061	-0.80866	0.98988	0.19601
ReLu	ELM	3.0371	-2.378	0.73454	1.0039
	BH-ELM	1.1901	-0.90067	0.98819	0.21175
	GSBH-ELM	1.2849	-0.81874	0.98929	0.20162
	GSLBH-ELM	1.2062	-0.84039	0.98723	0.19512
hardlim	ELM	2.5602	-3.4302	0.60757	1.2206
	BH-ELM	2.6908	-1.6668	0.90066	0.61413
	GSBH-ELM	2.5888	-1.9176	0.88205	0.66918
	GSLBH-ELM	2.2253	-1.7552	0.90317	0.60631
Morlet	ELM	3.8217	-3.9996	0.048318	1.9008
	BH-ELM	4.7295	-2.4039	0.85768	0.73506
	GSBH-ELM	1.2647	-0.82713	0.98582	0.23205
	GSLBH-ELM	1.2027	-0.85689	0.9891	0.20347
thah	ELM	2.5084	-1.8113	0.82955	0.80443
	BH-ELM	2.0116	-1.6645	0.91962	0.55242
	GSBH-ELM	1.202	-0.81683	0.98984	0.19641
	GSLBH-ELM	1.2105	-0.81396	0.98986	0.19618

TABLE 8. Training time of ELM model for predicting ore granularity.

	ELM	BH-ELM	GSBH-ELM	GSLBH-ELM
sigma	0.0006768	0.5612	1.0813	1.777
sin	0.0025442	0.49084	0.92034	1.6518
arctan	0.0025459	0.45242	0.96224	1.6582
ReLu	0.0024785	0.40959	0.89416	1.6035
hardlim	0.0023822	0.43708	0.92314	1.6231
Morlet	0.0012938	0.97251	1.7516	3.053
thah	0.0022054	0.56301	1.0856	1.7977

TABLE 9. Training time of ELM model for predicting pulp iron content.

	ELM	BH-ELM	GSBH-ELM	GSLBH-ELM
sigma	0.0007338	0.49646	1.0146	1.7175
sin	0.0030778	0.50329	0.97306	1.7943
arctan	0.10146	0.46165	0.99828	1.7308
ReLu	0.0025851	0.40985	0.87537	1.5444
hardlim	0.10537	0.43893	0.90977	1.5965
Morlet	0.0009433	0.98719	1.8102	3.1265
thah	0.0005915	0.56761	1.1359	1.8728

W. Xie *et al.*: ELM Soft-Sensor Model With Different Activation Functions on Grinding Process Optimized by Improved BH Algorithm

IEEE *Access*

V. CONCLUSION

This paper first introduces the grinding and classifying process in details. After carefully analyzing the ore processing process in combination with a certain background of grinding process and certain theoretical knowledge, ten process variables were selected as secondary variables to predict the granularity and iron content of the grinding process. Then the KPCA method was adopted to carry out the data dimension reduction. Then the ELM neural network was used to set up the soft-sensor model and 7 different continuous functions are selected as the ELM activation functions. An improved black hole algorithm was used to optimize these 7 different soft sensor model, and then these models were used to predict the granularity and iron content of the grinding process. Through simulation experiments, this paper can draw the followed conclusions.

(1) After using BH and improved BH to optimize the weight w and bias b of the ELM model, the ELM model will show different prediction accuracy, where the prediction accuracy of GSLBH-ELM is better than the prediction accuracy of GSBH-ELM, BH-ELM and ELM.

(2) When the ELM model uses different activation functions, its prediction results for the two production indicators will also be different. When the ore granularity is predicted, the ELM model with the arctan function as the activation function has the highest prediction accuracy, and the RMSE of GSLBH-ELM reaches 0.057468. When pulp iron content is predicted, the ELM model with the sin function as the activation functions has the highest prediction accuracy, and the RMSE of GSLBH-ELM reaches 0.19483.

(3) Although using optimization algorithms to optimize the parameters of the ELM model increases the computational cost, the added cost is still within an acceptable range.

The industrial environment of the grinding and classifying process is very complicated, and the industrial status may change at any time, the future work is mainly divided into two parts. (1) Based on the work that has been done, we try to improve the structure of the ELM model and introduce some adaptive technologies (such as moving window, just-in-time) so that it can maintain a good generalization ability when the industrial state changes. (2) In the database of the grinding grading process, a large number of samples have not been labeled. In the future work, the semi-supervised model of ELM will be studied, so that these unlabeled samples can also be used to predict the production indicators of the grinding process.

REFERENCES

[1] H. G. Wen and M. Y. Ni, "Multiple neural networks-based soft sensing model for particle size of grinding product," *Metal Mine*, vol. 2, no. 344, pp. 47–49, 2005.

[2] Y. Sun, Y. Wang, X. Liu, C. Yang, Z. Zhang, W. Gui, X. Chen, and B. Zhu, "A novel Bayesian inference soft sensor for real-time statistic learning modeling for industrial polypropylene melt index prediction," *J. Appl. Polym. Sci.*, vol. 134, no. 40, p. 45384, Oct. 2017.

[3] S. Zheng, J. Guo, N.-Z. Shi, and G.-L. Tian, "Likelihood-based approaches for multivariate linear models under inequality constraints for incomplete data," *J. Stat. Planning Inference*, vol. 142, no. 11, pp. 2926–2942, Nov. 2012.

[4] Z.-G. Su and P.-H. Wang, "Regression analysis of belief functions on interval-valued variables: Comparative studies," *Soft Comput.*, vol. 18, no. 1, pp. 51–70, Jan. 2014.

[5] M. Kano, S. Lee, and S. Hasebe, "Two-stage subspace identification for softsensor design and disturbance estimation," *J. Process Control*, vol. 19, no. 2, pp. 179–186, Feb. 2009.

[6] J. L. Godoy, J. L. Marchetti, and J. R. Vega, "An integral approach to inferential quality control with self-validating soft-sensors," *J. Process Control*, vol. 50, pp. 56–65, Feb. 2017.

[7] H. A. L. Kiers and A. K. Smilde, "A comparison of various methods for multivariate regression with highly collinear variables," *Stat. Meth. Appl.*, vol. 16, no. 2, pp. 193–228, Jul. 2007.

[8] W. Shao, Z. Ge, and Z. Song, "Soft-sensor development for processes with multiple operating modes based on semisupervised Gaussian mixture regression," *IEEE Trans. Control Syst. Technol.*, vol. 27, no. 5, pp. 2169–2181, Sep. 2019.

[9] C. Shang, X. Gao, F. Yang, and D. Huang, "Novel Bayesian framework for dynamic soft sensor based on support vector machine with finite impulse response," *IEEE Trans. Control Syst. Technol.*, vol. 22, no. 4, pp. 1550–1557, Jul. 2014.

[10] Z. Ge, "Supervised latent factor analysis for process data regression modeling and soft sensor application," *IEEE Trans. Control Syst. Technol.*, vol. 24, no. 3, pp. 1004–1011, May 2016.

[11] A. Casali, G. Vallebuona, M. Bustos, G. Gonzalez, and P. Gimenez, "A soft-sensor for solid concentration in hydrocyclone overflow," *Minerals Eng.*, vol. 11, no. 4, pp. 375–383, Apr. 1998.

[12] A. Casali, G. Gonzalez, F. Torres, G. Vallebuona, L. Castelli, and P. Gimenez, "Particle size distribution soft-sensor for a grinding circuit," *Powder Technol.*, vol. 99, no. 1, pp. 15–21, Sep. 1998.

[13] A. Casali, G. Gonzalez, G. Vallebuona, C. Perez, and R. Vargas, "Grindability soft-sensors based on lithological composition and on-line measurements," *Minerals Eng.*, vol. 14, no. 7, pp. 689–700, Jul. 2001.

[14] D. Sbarbaro, P. Ascencio, P. Espinoza, F. Mujica, and G. Cortes, "Adaptive soft-sensors for on-line particle size estimation in wet grinding circuits," *Control Eng. Pract.*, vol. 16, no. 2, pp. 171–178, Feb. 2008.

[15] P. Zhou, "Soft-sensor approach with case-based reasoning and its application in grinding process," *Control Decis.*, vol. 21, no. 6, p. 646, 2006.

[16] P. Zhou, S.-W. Lu, and T. Chai, "Data-driven soft-sensor modeling for product quality estimation using case-based reasoning and fuzzy-similarity rough sets," *IEEE Trans. Autom. Sci. Eng.*, vol. 11, no. 4, pp. 992–1003, Oct. 2014.

[17] Y. Zhang and Z. Wang, "Soft sensor model based on RBF neural network for particle size of grinding circuit and implementation," *Comput. Meas. Control*, vol. 25, no. 1, pp. 38–39, 2017.

[18] X.-L. Huang, "Application of improved RBF neural network to the prediction of grinding production index," *Control Eng. China*, vol. 15, no. 5, pp. 560–563, 2008.

[19] J. Tang, L.-J. Zhao, H. Yue, and T.-Y. Chai, "Soft sensor for ball mill load based on multi-source data feature fusion," *J. Zhejiang Univ.*, vol. 44, no. 7, pp. 1406–1413, 2010.

[20] J. Tang, "Soft sensor modeling of ball mill load via principal component analysis and support vector machines," in *Advances in Neural Network Research and Applications*. Berlin, Germany: Springer, 2010, pp. 803–810.

[21] J.-S. Wang, Y. Yang, and S. Shi-Feng, "SFLA-WNN soft-sensor modeling and reconfiguration of grinding process based on model migration," *J. Shanghai Jiaotong Univ.*, vol. 46, no. 12, pp. 1951–1955, 2012.

[22] J.-S. Wang and S. Na-Na, "Hybrid multiple soft-sensor models of grinding granularity based on cuckoo searching algorithm and hysteresis switching strategy," *Sci. Program.*, vol. 2015, Dec. 2015, Art. no. 146410.

[23] W. Dai, Q. Liu, and T. Chai, "Particle size estimate of grinding processes using random vector functional link networks with improved robustness," *Neurocomputing*, vol. 169, pp. 361–372, Dec. 2015.

[24] Y. Kang, M.-C. Lu, and G.-W. Yan, "Soft sensor for ball mill fill level based on DBN-ELM model," *Instrum. Technique Sensor*, vol. 4, pp. 73–75, 2015.

[25] C. Ruihui and Y. Gaowei, "Soft sensor for ball mill fill level based on OBE-ELM model," *J. North Univ. China, Natural Sci.*, vol. 38, no. 5, pp. 574–579, 2017.

[26] S. Zhang, Z. Liu, X. Huang, and W. Xiao, "A modified residual extreme learning machine algorithm and its application," *IEEE Access*, vol. 6, pp. 62215–62223, 2018.

[27] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, "Extreme learning machine: Theory and applications," *Neurocomputing*, vol. 70, nos. 1–3, pp. 489–501, Dec. 2006.

[28] Y. Liu, L. Zhang, P. Deng, and Z. He, "Common subspace learning via cross-domain extreme learning machine," *Cogn. Comput.*, vol. 9, no. 4, pp. 555–563, Aug. 2017.

[29] T. Guo, L. Zhang, and X. Tan, "Neuron pruning-based discriminative extreme learning machine for pattern classification," *Cogn. Comput.*, vol. 9, no. 4, pp. 581–595, Aug. 2017.

[30] L. Zhang and D. Zhang, "Robust visual knowledge transfer via extreme learning machine-based domain adaptation," *IEEE Trans. Image Process.*, vol. 25, no. 10, pp. 4959–4973, Oct. 2016.

[31] L. Zhang and D. Zhang, "Evolutionary cost-sensitive extreme learning machine," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 12, pp. 3045–3060, Dec. 2017.

[32] L. Zhang and D. Zhang, "Domain adaptation extreme learning machines for drift compensation in E-nose systems," *IEEE Trans. Instrum. Meas.*, vol. 64, no. 7, pp. 1790–1801, Jul. 2015.

[33] L. Zhang, X. Wang, G.-B. Huang, T. Liu, and X. Tan, "Taste recognition in e-tongue using local discriminant preservation projection," *IEEE Trans. Cybern.*, vol. 49, no. 3, pp. 947–960, Mar. 2019.

[34] Y. Li and Z. Yang, "Application of EOS-ELM with binary Jaya-based feature selection to real-time transient stability assessment using PMU data," *IEEE Access*, vol. 5, pp. 23092–23101, 2017.

[35] Y. Zhang, T. Li, G. Na, G. Li, and Y. Li, "Optimized extreme learning machine for power system transient stability prediction using synchrophasors," *Math. Problems Eng.*, vol. 2015, Nov. 2015. Art. no. 529724.

[36] W. Xie, J. S. Wang, and Y. Tao, "Improved black hole algorithm based on golden sine operator and levy flight operator," *IEEE Access*, vol. 7, pp. 161459–161486, 2019.

[37] W. Z. Sun and J. S. Wang, "Elman neural network soft-sensor model of conversion velocity in polymerization process optimized by chaos whale optimization algorithm," *IEEE Access*, vol. 5, pp. 13062–13076, 2017.

[38] Z. Chen, S. Han, and K. Wang, "Genetic association test based on principal component analysis," *Stat. Appl. Genet. Mol. Biol.*, vol. 16, no. 3, pp. 189–198, 2017.

[39] E. Tanyildizi and G. Demir, "Golden sine algorithm: A novel math-inspired algorithm," *Adv. Elect. Comput. Eng.*, vol. 17, no. 2, pp. 71–78, 2017.

[40] C. T. Brown and S. Larry Liebovitch Rachel Glendon, "Lévy flights in dobe Ju/'hoansi foraging patterns," *Human Ecol.*, vol. 35, no. 1, pp. 129–138, Feb. 2007.

[41] A. M. Reynolds and M. A. Frye, "Free-flight odor tracking in Drosophila is consistent with an optimal intermittent scale-free search," *PLoS ONE*, vol. 2, no. 4, p. e354, Apr. 2007.

[42] I. Pavlyukevich, "Lévy flights, non-local search and simulated annealing," *J. Comput. Phys.*, vol. 226, no. 2, pp. 1830–1844, Oct. 2007.

W. XIE is currently pursuing the master's degree with the School of Electronic and Information Engineering, University of Science and Technology Liaoning, Anshan, China.

J. S. WANG (Member, IEEE) received the B.Sc. and M.Sc. degrees in control science from the University of Science and Technology Liaoning, China, in 1999 and 2002, respectively, and the Ph.D. degree in control science from the Dalian University of Technology, China, in 2006. He is currently a Professor and a Master's Supervisor with the School of Electronic and Information Engineering, University of Science and Technology Liaoning. His main research interests are the modeling of complex industry process, intelligent control, and computer integrated manufacturing.

C. XING received the B.Sc. degree in automation from the University of Science and Technology Liaoning, China, in 2011, and the M.Sc. degree in computer science and engineering from Hanyang University, South Korea, in 2013. She is currently pursuing the Ph.D. degree. Her main research interest is modeling of complex industry process.

S. S. GUO is currently pursuing the master's degree with the School of Electronic and Information Engineering, University of Science and Technology Liaoning, Anshan, China.

M. W. GUO is currently pursuing the master's degree with the School of Electronic and Information Engineering, University of Science and Technology Liaoning, Anshan, China.

L. F. ZHU is currently pursuing the master's degree with the School of Electronic and Information Engineering, University of Science and Technology Liaoning, Anshan, China.

● ● ●