Prediction of Endpoint Sulfur Content in KR Desulfurization Based on the Hybrid Algorithm Combining Artificial Neural Network With SAPSO

SIWEI WU1, JIAN YANG1, RUNHAO ZHANG1, AND HIDEKI ONO2
1State Key Laboratory of Advanced Special Steel, School of Materials Science and Engineering, Shanghai University, Shanghai 200444, China
2Faculty of Sustainable Design, University of Toyama, Toyama 930-8555, Japan
Corresponding author: Jian Yang (yang_jian@t.shu.edu.cn)

This work was supported in part by the China Postdoctoral Science Foundation funded Projects under Grant 2019M651467 and in part by the Natural Science Foundation Joint Fund Project of Liaoning Province under Grant 2019-KF-25-06.

ABSTRACT In the present work, the endpoint sulfur content prediction model of Kambara Reactor (KR) desulfurization in the steelmaking process is investigated. For Artificial Neural Network (ANN), the effects of different structure parameters, including the number of hidden layer neurons, activation functions and training functions, on the performance of desulfurization model are studied. The initial weights and biases of the neural network is optimized to further elevated the prediction accuracy of the model. Three models established by using Multiple Linear Regression (MLR), ANN and a hybrid algorithm (artificial neural network optimized by SAPSO, named SAPSO-ANN) are compared by the Correlation Coefficient (R), Root Mean Square Error (RMSE) and Mean Absolute Relative Error (MARE). The results show that in the process of KR desulfurization, the nonlinear model of ANN and SAPSO-ANN has a higher accuracy than the linear model of MLR. Among the three models, the SAPSO-ANN model achieves the highest accuracy with R value of 0.54, RMSE of $2.61 \times 10^{-4}$% and MAER of 0.47, which is selected to analyze the effect of process parameters on the desulfurization rate and design the amount of desulfurization flux in the KR desulfurization process. Experimental results show good agreements with the calculation results, indicating the practicability of the model.

INDEX TERMS Artificial neural network, particle swarm optimization, multiple linear regression, prediction of endpoint sulfur content, KR desulfurization process.

I. INTRODUCTION

In the steelmaking process, since the complex physical changes and chemical reactions take place, it has the characteristics of high temperature, multi-variables and nonlinearity for relationship between the input process parameters and the output steelmaking results [1], [2]. Traditionally, the mechanism model is established by solving partial differential equation based on the theories of metallurgical transportation, chemical reaction and mass conservation. However, it has shortcomings of complexity in modeling, poor universality, low accuracy and difficulty in maintenance.

Benefited from the large amount of industrial data available to be collected, lots of process control models have been developed by using intelligent algorithms. In the process of hot metal pretreatment, Zhang et al. [3] applied an optimization neural network model based on grey theory to predict endpoint temperature of hot metal in the process of dephosphorization. In the process of converter blowing, researchers have established prediction models to forecast temperature [2], [4]–[8], carbon content [2], [4], [5], [8], phosphorus content [1], [9], [10], manganese content [11] of endpoint and the amount of desulfurization flux [12], achieving reasonable results. In the second blowing period of basic oxygen furnace steelmaking, Han and Zhao [13] has combined the adaptive-network-based fuzzy inference system and robust relevance vector machine to calculate the
additional amounts of oxygen and coolant. In the process of Ladle Furnace (LF) refining, endpoint temperature of molten steel have been predicted by using the ensemble extreme learning machine based on modified AdaBoost. RT algorithm [14], [15], the back propagation neural network combined with expert system [16], the combined method of case-based reasoning and Bayesian belief network [17] and the feature-weighted model optimized by mutual learning cuckoo search [18]. As most of the existing temperature prediction models are built on small-scale data, Wang et al. [19], [20] proposed a bootstrap feature subsets ensemble regression trees method to establish temperature prediction model of molten steel in LF refining on the large-scale data, improving the accuracy and generality of the model. In the process of Ruhstahl Hausen (RH) refining, Feng et al. [21] proposed an improved case-based reasoning method to predict the endpoint temperature of molten steel, acquiring a higher hit rate than that of ordinary case-based reasoning method. Although the data-driven models have made some progress in endpoint results of hot metal pretreatment dephosphorization, converter blowing, LF refining and RH refining, the research on prediction of hot metal pretreatment of Kambara Reactor (KR) desulfurization is still insufficient. Since there are so many model parameters to be determined during establishing the endpoint sulfur content prediction model of hot metal, it is difficult to establish a data-driven model suitable for accurately predicting desulfurization endpoint based on hot metal pretreatment data. Moreover, for neural network of KR desulfurization, parameters such as number of hidden layer neurons, activation function, training function or weights and biases of network, have significant influences on network performance, and the influences of these parameters on the performance of the network model are not clear yet.

This paper aims at establishing a data-driven model to predict the endpoint sulfur ([S]_{end}) content of hot metal during KR desulfurization process. Based on the industrial data collected from hot metal pretreatment production process in a steelmaking plant, the data cleaning strategy for reducing redundant and removing abnormal data was developed. The influences of network parameters, such as number of hidden layer neurons, activation function (‘tansig’, ‘logsig’ and ‘purelin’) and training function (‘traingd’, ‘trainda’, ‘traindx’, ‘trainbr’, ‘trainlm’, ‘trainrp’, ‘traingdm’ and ‘trainscg’), on network performance were studied. In order to avoid the influences of random weights on network performance, a hybrid algorithm, which is a Particle Swarm Optimization revised by Simulation Anneal algorithm (SAPSO), was applied to optimize the initial weights and biases of neural networks. In order to verify the performance of the proposed model, the present model was compared with the Artificial Neural Network (ANN) model and the traditional statistical model of Multiple Linear Regression (MLR) model. The prediction errors of the three data-driven models were quantitatively evaluated by using Correlation Coefficient (R), Root Mean Square Error (RMSE) and Mean Absolute Relative Error (MARE). Finally, the SAPSO-ANN model (combination of SAPSO and ANN model) was implemented to calculate the amount of desulfurization flux to optimize the KR desulfurization process and the experiments were carried out to verify the feasibility and practicability of the model.

The rest of the paper is organized as follows. In Section 2, a brief description of KR desulfurization process during hot metal pretreatment was introduced. In Section 3, the theories of multiple linear regression, artificial neural network and particle swarm optimization algorithm were presented. These algorithms were applied or combined to establish prediction models for [S]_{end} content during KR desulfurization process in Section 4. The performances of the established models were compared and the best performance model was used to optimize the KR desulfurization process in Section 5. In the last section, we make some conclusions about this study.

II. A DESCRIPTION OF KR DESULFURIZATION PROCESS
KR stirring method is one of the most commonly used economical means for hot metal desulfurization pretreatment [22], [23]. Schematic diagram of KR desulfurization is shown in Figure 1. In process of KR desulfurization, an impeller with a protective covering of refractory materials is inserted into a certain depth below the liquid level of hot metal ladle [24], [25]. After adding desulfurization flux into the hot metal, the impeller is rotated to stir the hot metal and make its level form a whirlpool. The desulfurization flux particles are dispersed in the end region of the slurry blade due to the turbulence of hot metal and “spit out” along the radius direction. Then the desulfurization flux particles are suspended, rotated around the axis or floated in hot metal. By stirring, the desulfurization flux fully contacts and reacts with hot metal, reducing the sulfur content of hot metal and achieving the purpose of desulfurization. The lime is applied as desulfurization flux in this study. CaO reacts with S in hot metal to form high melting point product of CaS transferring into slag with the aid of C in the hot metal. And the chemical reaction in hot metal is as following [26].

\[
\text{CaO(s)} + \text{[S]} + \text{[C]} = \text{(CaS)} + \text{CO(g)} \quad \Delta G^0 = -86670 - 68.96T
\]
Solid CaO absorbs $[S]$ from hot metal very quickly. Carbon in hot metal is reductants, which can combined with oxygen generated by reaction to form CO gas. Therefore, the products of desulfurization are CaS and CO. Lime desulfurization is an endothermic reaction. Under the conditions of high temperature, high alkalinity, reducing atmosphere (low oxygen potential) and high $[C]$, $[Si]$, $[P]$ contents that increase the activity coefficient of sulfur, the desulfurization process is promoted. In terms of desulfurization mechanism, CaO particles contact with $[S]$ in hot metal to form a slag shell of CaS, which hinders the diffusion of $[S]$ and $[O]$ in the hot metal and results in the desulfurization slowing down. After stirring stops, the slag that floats up to the surface of hot metal is removed to separate the sulfur from the hot metal.

There are several process parameters that affect the performance of KR desulfurization process. Hot metal temperature has a great effect on desulfurization, as desulfurization process is an endothermic reaction, which is easy to occur at high temperature. In addition, high temperature is beneficial to reducing the viscosity of slag so that the fluidity of slag is increased, which improves the kinetic conditions of desulfurization [27]. The amount of desulfurization flux with a suitable range is also an important factor affecting desulfurization. If the amount of desulfurization flux is too small, the reaction will be insufficient. If the amount of desulfurization flux is too large, utilization efficiency of desulfurization flux will be reduced, resulting in an extra consumption of desulfurization flux to increase the desulfurization cost. Furthermore, it will also generate more slag during the desulfurization process, increasing temperature drop and the difficulty of slag treatment. With the increase of the initial sulfur ($[S]_{ini}$) content in hot metal, the activity of $[S]$ increases, resulting that the reaction of desulfurization occurs easily. The stirring time is also an important factor affecting desulfurization performance. Too short stirring time leads to insufficient desulfurization reaction, while too long stirring time results in resulphurization occurrence and long desulfurization treatment period. Increasing the rotation speed of the impeller can accelerate the mass transfer to promote the desulfurization reaction. However, it will also increase the power consumption, the processing cost and temperature drop of hot metal. During the desulfurization process, the impeller needs to be inserted into an appropriate depth. If the insertion depth is too shallow, the level fluctuation of hot metal is serious, and the desulfurization flux can not reach the bottom of the hot metal ladle, resulting in the insufficient of desulfurization. If the insertion depth is too deep, the hot metal at the bottom of the ladle fluctuates greatly during stirring. The shear stress causes great friction between the hot metal and the bottom of the ladle. The refractory on the inner wall of the ladle falls off and enters the hot metal, contaminating the hot metal. Moreover, the stirring of the hot metal is insufficient and the desulfurization is not complete.

**III. BRIEF DESCRIPTIONS OF MODELING TECHNIQUES**

**A. ARTIFICIAL NEURAL NETWORK**

ANN is a widely used machine learning approach to establish the relationship between input information and output information, which simulates the flow of information inside the brain [28]. It defines a set of neurons as functions to process
the information input the network. Figure 2 shows a typical structure of a three layer back-propagation neural network. The neural network contains one input layer, one output layer and one hidden layer. Each layer contains several neurons. The neurons in each layer connect with a certain weight, $w_{ij}$, to all the neurons in the next layer and each neuron has a bias, $b_i$. Neurons in the same layer are disconnected from each other. For a neuron, there is an activation function to process the input data. During the training, these weights and biases will be adjusted to minimize the mean squared error of the network’s output. By adjusting the weights and biases, the network prediction capability is improved [29], [30].

In the network, each neuron implements a basic computation and produces the output [31]. The input data will be processed by the neurons of input layer, then the output can be obtained, which will be used as the input for the next layer. The whole network output value can be obtained by using the following equations [32].

$$o^h = \phi^h \left( \sum_{j=1}^{m} w_{ij}^h x_j + b^h \right)$$ \hspace{1cm} (1)

$$y = o^o = \phi^o \left( \sum_{j=1}^{k} w_{jh}^o o^h + b^o \right)$$ \hspace{1cm} (2)

where $x$ is the input and $y$ is the output; $o^h$ and $o^o$ are the output of hidden layer neuron and output layer neuron, respectively; $w_{ij}^h$ is weight between input layer neuron and hidden layer neuron; $w_{ij}^h$ is weight between hidden layer neuron and output layer neuron; $b^h$ and $b^o$ are biases of hidden layer neuron and output layer neuron, respectively; $m$ is the number of input variables; $k$ is the number of hidden layer neurons; $\phi^h$ and $\phi^o$ are the activation function of hidden layer neuron and output layer neuron, respectively. The output of current layer neuron is used as an input to the neurons of next layer. The activation function for each neuron can be selected from three different expressions, which can be seen in Figure 3. The function, called ‘tansig’, in Figure 3(a) is mathematically equivalent to hyperbolic tangent function and has stronger gradient, or higher derivation, than the standard logistic function (Called ‘logsig’ in Figure 3(b)), which allows faster optimization of the network. The ‘purelin’ function is a pure linear function, which is usually used in the output layer.

In this work, a Mean Squared Error (MSE) function is used to calculate the network output error, which can be expressed as Eq. (3).

$$\text{MSE} = \frac{1}{N} \sum_{i=1}^{N} (y_i - t_i)^2 \hspace{1cm} (3)$$

where $y_i$ and $t_i$ represent the output of the network and the target value, respectively. $N$ denotes the number of experimental data employed in the investigation. The MSE is used to help to determine the changes of weights and biases of network. It will be recalculated until the MSE decrease below the predefined value or the other stop condition reaches [33]. The weights updating of network can be realized according to Eq. (4) and Eq. (5) [34]:

$$w_{jk}^{ho}(i+1) = w_{jk}^{ho}(i) + \eta \left( t_k - o^o_k \right) \phi'^o \left( net^o_k \right) o^h_j$$ \hspace{1cm} (4)

$$w_{ij}^{ho}(i+1) = w_{ij}^{ho}(i) - \eta \phi'^h \left( net^h_j \right) \times \sum_{i=1}^{M} \left( t_k - o^o_k \right) \phi'^o \left( net^o_k \right) w_{jk}^{ho}(i+1)$$ \hspace{1cm} (5)

where $w_{ij}^{ho}$ respects the weight between the $i$-th neuron of the input layer and the $j$-th neuron of the hidden layer; $w_{jk}^{ho}$ respects the weight between the $j$-th neuron of the hidden layer and the $k$-th neuron of the output layer; $t_k$ is the target value of the $k$-th output neuron; $o^o_k$ is the $k$-th output neuron of the output layer; $o^h_j$ is the $j$-th neuron of the hidden layer; $net^h_j$ is the $j$-th input of the hidden layer; $net^o_k$ is the $k$-th input of the output layer; $\eta$ is the learning rate; $M$ is the number of output layer neurons; $ii$ is the current iteration number of the given sample.

B. PARTICLE SWARM OPTIMIZATION

Particle Swarm Optimization (PSO), proposed by Eberhart and Kennedy in 1995 [35], is a sociologically inspired global optimization approach, which is based on the motion of a flock of birds in search of food. The basic idea of the PSO algorithm is to firstly generate a group of random particles or solutions, and then the particles or proposed solutions evolve.
by iteration, finally moving towards the optimal solution of the problem [29].

In PSO, each individual of the swarm is considered as a particle with a position, \( x \), and a velocity, \( v \), in a multi-dimensional space. The particle dimension is equal to the number of variables. The initial particle position, \( x_0 \), and velocity, \( v_0 \), are chosen randomly. The value of the fitness function is calculated for each particle to evaluate the performance of the solution. The velocity and position of particle are updated by taking into account the fitness value to achieve the global best position [29]. By competing for the best position among the particle swarm, PSO algorithm has a very fast convergence speed. However, the competition among particles affects the solution diversity of the particle swarm, which may lead to the algorithm falling into local optimum. To avoid this problem, a parallel algorithm, simulated annealing algorithm, is introduced into PSO to improve the algorithm diversity [36]. Assuming that \( t \) is the anneal temperature. The initial temperature \( t_0 \) can be expressed as Eq. (6).

\[
t_0 = \frac{f(p_g)}{\ln 5}
\]  

where \( p_g \) is the global best position in the entire swarm, \( f \) is the fitness function. In the process of iteration, the Eq. (7) can be used to calculate the temperature drop.

\[
t_{j+1} = \lambda t_j
\]  

where \( \lambda \) is temperature drop coefficient. According to Eq. (8), the fitness of each particle at the current temperature, TF, can be determined.

\[
TF(p_i) = \frac{e^{-\frac{f(p_i)-f(p_g)}{t_j}}}{\sum_{j=1}^{NP} e^{-\frac{f(p_i)-f(p_j)}{t_j}}}
\]  

where \( NP \) is the number of the particle swarm.

In each iteration process, each particle updates the velocity and location by tracking the current optimal position of the particle and that of the particle in the entire swarm. Updating the position and velocity are done by using Eq. (9) to Eq. (11).

\[
x_i(i+1) = x_i(ii) + v_i(ii + 1)
\]

\[
v_i(ii + 1) = \xi \left[ v_i(ii) + c_1r_1 [p_i - x_i(ii)] + c_2r_2 [p_g - x_i(ii)] \right]
\]

\[
\xi = \frac{2 - (c_1 + c_2) - \sqrt{(c_1 + c_2)^2 - 4(c_1 + c_2) \sum_{i=1}^{m} (x_i - y_i)^2}}{2}
\]

where \( x_i \) is the position of \( i \)-th particle in the swarm; \( ii \) is the number of current iterations; \( v_i \) is the velocity of \( i \)-th particle in the swarm; \( c_1 \), called the ‘cognitive’ factor, is the particle’s cognition of its own knowledge; \( c_2 \), called the ‘social’ factor, is a learning factor, which means the particle knowledge of the entire swarm; \( r_1 \) and \( r_2 \) are the random number between 0 and 1; \( p_i \) is the local best position of the particle; \( p_g \) is the global best position of the particle [37].

IV. ESTABLISHMENT OF A PREDICTION MODEL FOR ENDPOINT SULFUR CONTENT

A. DATA CLEANING

Experimental data are collected for establishing the prediction model for endpoint sulfur content during hot metal pretreatment of KR desulfurization process. For the purpose of obtaining a final model with strong robustness and high prediction accuracy, 15 key variables are synthetically picked out. However, the original data inevitably contain abnormal data like outliers, redundant data or extra-large dispersion distribution, which will cause a misleading prediction. In order to obtain a robust model and reliable analysis results, the data must be cleaned before modeling.

In our work, Pauto criterion is implemented for data cleaning [38]. Assuming \( X \) is data of a certain variable, \( X = [X_1, X_2, \ldots, X_i, \ldots, X_N] \) (\( N \) is the number of the data). The average and standard deviation of \( X \) can be calculated by Eq. (12) to Eq. (14).

\[
\mu = \frac{1}{N} \sum_{i=1}^{N} X_i
\]

\[
V_i = X_i - \mu
\]

\[
\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (X_i - \mu)^2}
\]

where \( \mu \) is the average of \( X \); \( V_i \) is the residual error for \( i \)-th data; \( \sigma \) is the standard deviation of \( X \). Eq. (13) is used to calculate the residual error of data. Data satisfying the condition of \( |V_i| > 3\sigma \) will be removed as abnormal data. That is to say, the abnormal data are not contained in the range of \( [\mu - 3\sigma, \mu + 3\sigma] \). After the current abnormal data are removed, all the remaining data will be recalculated until all the data satisfy the condition of \( X_i \in[\mu - 3\sigma, \mu + 3\sigma] \). Data of each variable will be treated in this way.

In addition to the abnormal data in the original dataset, redundant data also need to be eliminated. Therefore, hierarchical clustering is applied to divide the data into several groups. Hierarchical clustering starts with all of the samples in one cluster and forms the sequence by successively splitting clusters. Euclidean distance metric is selected as distance measurement. The Euclidean distance between point \( x \) and point \( y \) is expressed as Eq. (15).

\[
d(x, y) = \sqrt{\sum_{i=1}^{m} (x_i - y_i)^2}
\]

where \( m \) is the number of variables. By calculating the Euclidean distance between data, the data are divided into several groups. The number of groups is controlled by the thresholds of main variables. The threshold represents the critical range for each variable in a group [39]. If the ranges of all the variables in a group are smaller than the threshold vector, then the clustering stops. Otherwise, the clustering continues to divide the data into more groups. During the process of data pretreatment, the thresholds are set as 0.001 mass% for [S]end content, 0.005 mass% for [S]ini content, 500 kg for lime weight, 50 kg for fluorite weight, 20 t for hot metal weight,
100°C for hot metal temperature and 100°C for hot metal temperature after treatment. Then, for data in one group, data of each dimension containing the minimum value or the maximum value are selected instead of all the data in the group used for modeling [40]. In this way, the redundant data are eliminated. Figure 4 shows abnormal data removed for hot metal temperature and lime weight. Data beyond the dash lines represent abnormal data judged by Pauta criterion. Therefore, the abnormal data are removed to improve the data quality. Figure 5 shows relationship among the data of $[S]_{\text{end}}$ content, lime weight and $[S]_{\text{ini}}$ content. It can be seen that the data distribute in a reasonable range, which indicates that the data can be used to establish a model with strong robustness.

After data cleaning, there are 1521 data remained. Based on the metallurgical principle, 15 variables are selected to establish the model of the $[S]_{\text{end}}$ content, which are $[S]_{\text{ini}}$ content, lime weight, fluorite weight, hot metal weight, hot metal temperature, hot metal temperature after treatment, deslaging weight, stirring time, hot metal height, medium stirring speed, medium stirring time, medium stirring depth, high stirring speed, high stirring time and high stirring depth. The symbols of these variables and the data used in establishing $[S]_{\text{end}}$ content prediction model are list in Table 1.

The process parameter variables have different order of magnitude, which will result in failure in modeling. In order to eliminate the different magnitude of each variable, the data are normalized to $[0, 1]$ by using the following equation.

$$ X_{\text{norm}} = \frac{X_i - X_{\text{min}}}{X_{\text{max}} - X_{\text{min}}} $$  \hspace{1cm} (16) 

where $X_{\text{norm}}$ is the normalized data, $X_i$ is the original data; $X_{\text{min}}$ and $X_{\text{max}}$ are minimum and maximum of original data, respectively. The data are randomly divided into two parts approximately 3:1. The data of 1141 cases are applied for model establishment, while the remaining data of 380 cases are applied for testing the model performance. The quantitative relationship between the $[S]_{\text{end}}$ content and the process parameters can be described as a function of $y = \psi(x_1, x_2, \ldots, x_{15})$. In this paper, the relationships are carried out by using MLR, ANN and SAPSO-ANN, respectively.

### B. MLR MODEL

Multiple linear regression is a common statistical modeling method, which is widely used in predictions. It can be used to investigate the relationship between dependent variables and multiple independent variables, and establish quantitative expressions between indicators and multiple influence factors. In this paper, based on the cleaned data above, MLR model is established for predicting $[S]_{\text{end}}$ content during KR desulphurization. The mathematical expression of the $[S]_{\text{end}}$ content of hot metal and process parameters can be written as Eq. (17).

$$ y = a_0 + a_1 x_1 + a_2 x_2 + \cdots + a_{15} x_{15} $$  \hspace{1cm} (17)
### TABLE 1. Data used in establishing $[S]_{\text{end}}$ content prediction model.

| Variable                                      | Symbol | Minimum | Maximum | Mean   | Standard deviation |
|-----------------------------------------------|--------|---------|---------|--------|--------------------|
| $[S]_{\text{end}}$ content/mass%              | $y$    | 0.0005  | 0.0033  | 0.00079| 0.00042            |
| $[S]_{\text{in}}$ content/mass%              | $x_1$  | 0.0102  | 0.0589  | 0.02945| 0.00921            |
| Lime weight/kg                                | $x_2$  | 1464    | 4030    | 2761.0 | 445.0              |
| Fluorite weight/kg                            | $x_3$  | 96      | 334     | 213.8  | 41.3               |
| Hot metal weight/t                            | $x_4$  | 282     | 322     | 302.0  | 6.8                |
| Hot metal temperature/°C                      | $x_5$  | 1250    | 1469    | 1360.4 | 37.6               |
| Hot metal temperature after treatment/°C      | $x_6$  | 1234    | 1431    | 1332.6 | 33.5               |
| Deslaging weight/t                            | $x_7$  | 0.2     | 14.9    | 6.74   | 2.91               |
| Stirring time/min                             | $x_8$  | 6       | 18      | 11.4   | 1.2                |
| Hot metal height/mm                           | $x_9$  | 6561    | 7408    | 6979.0 | 164.1              |
| Medium stirring speed/(r/min)                 | $x_{10}$| 54      | 120     | 99.1   | 14.8               |
| Medium stirring time/min                      | $x_{11}$| 1      | 7       | 2.1    | 0.5                |
| Medium stirring depth/mm                      | $x_{12}$| 1353   | 1493    | 1466.4 | 24.7               |
| High stirring speed/(r/min)                   | $x_{13}$| 66      | 120     | 113.2  | 9.3                |
| High stirring time/min                        | $x_{14}$| 5      | 15      | 9.3    | 1.1                |
| High stirring depth/mm                        | $x_{15}$| 1721   | 1822    | 1775.7 | 25.0               |

### TABLE 2. Values of the coefficients in MLR.

| Coefficient | Value | Coefficient | Value | Coefficient | Value | Coefficient | Value |
|-------------|-------|-------------|-------|-------------|-------|-------------|-------|
| $a_0$       | 0.4804| $a_1$       | -0.1390| $a_2$       | -0.0869| $a_3$       | 0.0892|
| $a_4$       | -0.0252| $a_5$       | -0.0799| $a_6$       | -0.0161| $a_7$       | -0.0203|
| $a_8$       | 0   | $a_9$       | 0.0626| $a_{10}$    | -0.0351| $a_{11}$    | -0.0840|
| $a_{12}$    | -0.2267| $a_{13}$    | -0.0572| $a_{14}$    | -0.0191| $a_{15}$    | 0.0461|

where $a_i$ is the coefficient to be determined corresponding to $x_i$. The values of the coefficients are solved by MATLAB software and the results are list in Table 2.

### C. ANN MODEL

The design of artificial neural network consists the number of neurons in hidden layer, the activation function in each layer and the network training method and so on. Choosing a suitable number of hidden layer neurons is important for a successful neural network. Too many neurons make the network complex for computation, and too few neurons cause inaccuracy of the model. Activation function can process the data input into the network, which has a significant influence on the convergence of the neural network. Different training functions affect the performance of the network by updating network weights in different ways. Researches have proved that a three layer neural network can achieve infinite approximation of arbitrary functions [41]. Hence, a three layer neural network is applied in this paper and a number of experiments are performed to select out the optimal network structure. The input variables are 15 process parameters and the output variable is $[S]_{\text{end}}$ content of hot metal. During the experiments, the networks with different number of hidden layer neurons (from 2 to 20), different activation functions of hidden layer and output layer neurons (‘tansig’, ‘logsig’ and ‘purelin’) and different network training functions (‘traingd’, ‘traingda’, ‘traingdx’, ‘trainbr’, ‘trainlm’, ’trainrp’, ‘traingdm’, ‘traingcg’) are established. The maximum training epoch is 2000 and the error goal for normalized training data is $10^{-7}$. Other parameters are set as defaults values of MATALB2018b. An early stop technique is implemented to avoid over-fitting problem. The network training continues until the error goal or the maximum epoch is reached. In order to avoid the influence of random initial weights and biases on network performance, models with each structure parameters run three times with random seeds of 0, 5 and 10, respectively. During the process, R and MSE in Eq. (18) and Eq. (19) are applied to evaluate the performance of the models.

$$R = \frac{\sum_{i=1}^{N} (M_i - \bar{M}) (P_i - \bar{P})}{\sqrt{\sum_{i=1}^{N} (M_i - \bar{M})^2 \sum_{i=1}^{N} (P_i - \bar{P})^2}} \quad (18)$$

$$\text{MSE} = \frac{1}{N} \sum_{i=1}^{N} (P_i - M_i)^2 \quad (19)$$

where $\bar{M}$ and $\bar{P}$ represent the mean values of measured and predicted values, respectively. $N$ denotes the number of data employed in the investigation.

All the neural network with different structure parameters combination are carried out for selecting best performance model on testing data, while the performance of models on
FIGURE 6. (a) R and (b) MSE of ANN with different number of hidden layer neurons both on training data and testing data.

FIGURE 7. (a) R and (b) MSE of ANN with different activation functions for hidden layer neurons and output layer neurons both on training data and testing data.

FIGURE 8. (a) R and (b) MSE of ANN with different training functions both on training data and testing data.
training data is used as reference for selecting the optimal network. Partial performance of the calculation results is shown in Figure 6 to Figure 8, in which the structure parameters of neural networks are set as follows. The number of hidden layer neurons is 4. The activation function is ‘tansig’ for hidden layer neuron and ‘pureline’ for output layer neuron. The training function is ‘trainbr’. When investigating a certain parameter, the other parameters are remained as above. Figure 6 shows R and MSE of ANN with different number of hidden layer neurons both on training data and testing data. As the number of hidden layer neurons increases, the performance of neural network on training data improves, while the performance of neural network on testing data improves slightly or even worsens because of the increase of network complexity. It can be seen that the ANN with hidden layer neuron number of 4 can achieve a highest R and a lowest MSE. Figure 7 shows R and MSE of ANN with different activation functions for hidden layer neuron and output layer neuron both on training data and testing data. In the Figure 7, ‘t-t’ represents ‘tansig’ function for both hidden layer neurons and output layer neurons; ‘t-l’ represents ‘tansig’ function for hidden layer neurons and ‘logsig’ function for output layer neurons; ‘l-t’ represents ‘logsig’ function for hidden layer neurons and ‘tansig’ function for output layer neurons; ‘l-l’ represents ‘logsig’ function for both hidden layer neurons and output layer neurons; ‘t-p’ represents ‘tansig’ function for hidden layer neurons and ‘purelin’ function for output layer neurons; ‘p-t’ represents ‘purelin’ function for hidden layer neurons and ‘tansig’ function for output layer neurons; ‘p-l’ represents ‘purelin’ function for hidden layer neurons and ‘logsig’ function for output layer neurons; ‘p-p’ represents ‘purelin’ function for both hidden layer neurons and output layer neurons. One can see that the ANN with activation function of ‘tansig’ for hidden layer neuron and ‘purelin’ for output layer neuron can achieve a best performance. In the Figure 8, ‘traingd’ represents a gradient descent algorithm; ‘traingda’ represents an adaptive learning algorithm; ‘traingdx’ represents an adaptive learning algorithm with momentum terms; ‘trainbr’ represents a Bayesian regularization algorithm; ‘trainlm’ represents a Levenberg-Marquardt
algorithm; ‘trainrp’ represents an elastic back propagation algorithm; ‘traindm’ represents a gradient descent algorithm with momentum terms; ‘trainscg’ represents a normalized conjugate gradient algorithm. The ANN with training function of ‘trainbr’ achieves a highest $R$ and a lowest MSE. In summary, ANN with the hidden layer neuron number of 4, activation function of ‘tansig’ for hidden layer neuron and ‘purelin’ for output layer neuron, training function of ‘trainbr’ achieve the best performance.

D. SAPSO-ANN MODEL

The initial weights and biases selection of ANN have great influence on the prediction accuracy. From Figure 6 to Figure 8, it should be noticed that ANN models generating weights and biases using different random seeds perform differently. In order to further improve the prediction accuracy of the network, a SAPSO algorithm is applied to optimize initial weights and biases of ANN model. And then the SAPSO-ANN model is used for predicting $[S]_{\text{end}}$ content of hot metal. Figure 9 shows the flow chart of SAPSO-ANN model.

The ANN model with best performance in Section 4.3 is further optimized. In the process of initialization of SAPSO algorithm, the particle is encoded according to the number of the weights and biases of the ANN. The total number of weights and biases can be obtained by Eq. (20).

$$n_{\text{total}} = n_{\text{input}} \times n_{\text{hidden}} + n_{\text{hidden}} \times n_{\text{output}} + n_{\text{output}}$$ (20)

where $n_{\text{input}}$, $n_{\text{hidden}}$ and $n_{\text{output}}$ are the number of neurons in input layer, hidden layer and output layer, respectively. Thus, each particle in the swarm is a combination of weights and biases of network [42]. The fitness function can be defined as the network prediction errors.

During the calculation of SAPSO, the position of particle in initial particle swarm is randomly generated within the range of $[-1, 1]$. The parameters used in the SAPSO are set as follows. The size of particle swarm is set as 20. The coefficients $c_1$ and $c_2$ are both set as 2. The temperature drop coefficient is set as 0.6. The SAPSO is run for 1000 epochs to provide enough time for the exchange of information between the particles. Figure 10 shows fitness value change with the iterative epoch of the SAPSO algorithm. It can be seen that the best fitness value decreases with the iterative epoch increases, finally achieving the smallest value of 0.018196.

V. RESULTS AND ANALYSIS

In order to make a comparative analysis on the three models, another 190 data are collected to test the model’s performance.

Figure 11 shows prediction absolute error of the three models. It can be seen that all the models can achieve the prediction of $[S]_{\text{end}}$ content of hot metal. The MLR model predict the $[S]_{\text{end}}$ content of hot metal with a small absolute error. Due to the extremely nonlinear fitting capability of ANN, the prediction absolute error of $[S]_{\text{end}}$ content of hot metal decreases. After the initial weights and biases of ANN are optimized by SAPSO algorithm, the accuracy of the model is improved, acquiring a high prediction accuracy.
Figure 12 shows box plot of prediction error of the three models. The box plot shows the mean value and fluctuation range of the prediction error of the three models. It can be seen the SAPSO-ANN model obtains the smallest prediction mean error of 5.32×10⁻⁵% comparing with the ANN model with prediction mean error of 6.34×10⁻⁵% and the MLR model with prediction mean error of 6.60×10⁻⁵%. Besides, the error range of ANN and SAPSO-ANN models are smaller than that of MLR model. The standard deviations of prediction error of ANN model and SAPSO-ANN model are also smaller than MLR model’s.

Figure 13 shows distribution of prediction error of the three models. Most of the prediction results are concentrated in areas with small errors. The prediction hit rate is over 98% with the error of ±0.0005% for all the three models. When considering the prediction hit rate with the error of ±0.0003%, ±0.0005%, ±0.0007% and ±0.0009%, the non-linear models, ANN and SAPSO-ANN, perform better than linear model, MLR, does. Among the three models, the SAPSO-ANN obtains the best performance with the hit rate of 98.95% with the error of ±0.0005%.

To further quantify the model prediction error, R is used to measure the model’s fit performance. RMSE and MARE are applied to evaluate the model prediction accuracy, which can be calculated by Eq. (21) and Eq. (22).

\[
RMSE = \sqrt{\frac{\sum_{i=1}^{N} (P_i - M_i)^2}{N}} \tag{21}
\]

\[
MARE = \frac{1}{N} \sum_{i=1}^{N} \left| \frac{P_i - M_i}{M_i} \right| \tag{22}
\]
where $P_i$ is the predicted value derived from the obtained model and $M_i$ is the measured value. $N$ is the number of data. Figure 14 shows R, RMSE and MARE of the three models, which further quantify and compare the prediction errors of the three models. The R of SAPSO-ANN model is larger than that of ANN model and MLR model, and the MARE and RMSE of SAPSO-ANN model are smaller than those of ANN model and MLR model. The comparison results suggest that the SAPSO-ANN model can obtain the best performance with R value of 0.54, RMSE of $2.61 \times 10^{-4}$% and MAER of 0.47 on prediction of $[S]_{\text{ini}}$ content during hot metal pretreatment of KR desulfurization process.

In order to further verify the rationality of the proposed model, the change in hot metal desulfurization rate with process parameters during KR desulfurization is studied by the SAPSO-ANN model. Figure 15 shows the effect of $[S]_{\text{ini}}$ content, lime weight, hot metal temperature and hot metal weight on desulfurization rate of hot metal. The Figure is obtained by varying two selected inputs, while other variables remain as mean values. The desulfurization rate, $\zeta$, can be obtained by Eq. (23).

$$\zeta = \frac{[S]_{\text{ini}} - [S]_{\text{end}}}{[S]_{\text{ini}}}$$

where $[S]_{\text{ini}}$ and $[S]_{\text{end}}$ are the [S] content of initial and endpoint of hot metal, respectively. It can be clearly seen from the Figure that the relationship of desulfurization rate of hot metal with process parameters is non-linearity. As the $[S]_{\text{ini}}$ content increases from 0.0102 mass% to 0.0589 mass%, the activity of [S] increases, which results in the increase in the desulfurization rate from 85.2% to 92.0%. As the weight of lime of the desulfurization flux increases, the desulfurization rate increases. When the hot metal weight is about 280 t, hot metal temperature increases from 1250 °C to 1469 °C, the desulfurization rate increases from 97% to 97.8%. The model is used to calculate the amount of desulfurization flux to ensure a high desulfurization rate at the most economical cost. Data of five ladles are randomly selected, the amount of lime weight is designed based on the SAPSO-ANN model by drawing the relationship between desulfurization rate, $[S]_{\text{ini}}$ content and lime weight while other variables remain unchanged. Table 3 shows a comparison of experimental results and calculated results based on the SAPSO-ANN model. It can be seen that the calculated results reasonably agree with the experimental results, verifying the applicability of the model.

### Table 3. Comparison of experimental results and calculated results based on the SAPSO-ANN model.

| No. | Data category | $[S]_{\text{ini}}$ content/\% | High stirring depth/mm | Lime weight/kg | $[S]_{\text{end}}$ content of hot metal/\% | Desulfurization rate/\% |
|-----|---------------|-------------------------------|----------------------|---------------|----------------------------------------|------------------------|
| 1   | Calculated    | 0.0255                        | ...                  | 1745          | 1933                                   | 0.00177                | 93.05%                 |
| 2   | Measured      | 0.0224                        | ...                  | 1795          | 2545                                   | 0.00107                | 95.21%                 |
| 3   | Calculated    | 0.0329                        | ...                  | 1798          | 3065                                   | 0.00059                | 98.20%                 |
| 4   | Measured      | 0.0276                        | ...                  | 1796          | 3083                                   | 0.00050                | 98.48%                 |
| 5   | Calculated    | 0.0213                        | ...                  | 1745          | 2021                                   | 0.00101                | 95.28%                 |
| 6   | Measured      |                               |                      | 2010          | 0.00100                                | 95.31%                 |

**FIGURE 15.** Effect of (a) $[S]_{\text{ini}}$ content and lime weight and (b) hot metal temperature and hot metal weight on desulfurization rate of hot metal.
VI. CONCLUSION

In the process of hot metal pretreatment, the establishment of a high accuracy endpoint sulfur content prediction model of hot metal can help to improve the efficiency of desulfurization and the automation control level of desulfurization production. However, there is a complex relationship between process variables and endpoint sulfur content of hot metal in the process of KR desulfurization. Establishing a high accuracy endpoint sulfur content prediction model of hot metal based on traditional mechanism model is difficult. In this paper, the effects of different structure parameters, including the number of hidden layer neurons, activation functions and training functions, on the performance of ANN were investigated. Combined neural network with optimization algorithm, a hybrid algorithm was proposed to establish the endpoint sulfur content prediction model of hot metal in KR desulfurization process. Comparing with traditional MLR model and ANN model, the proposed model achieved a better accuracy. Further, the amount of desulfurization flux was calculated based on the proposed model and the experiments were carried out to verify the practicality and effectiveness of the model. The conclusions can be summarized as follows:

1. To establish end point sulfur content prediction model with low signal-to-noise ratio, the hierarchical clustering and Pauta criterion can be effectively applied to improve the data quality by removing the abnormal data and redundant data in the KR desulfurization process.

2. As a nonlinear fitting model, ANN performs better than linear fitting model such as MLR. The ANN model with hidden layer neuron number of 4, ‘tansig’ function for hidden layer neuron and ‘purelin’ function for output layer neuron, ‘trainbr’ function for training function, can obtain a good performance on endpoint sulfur content prediction with the hit rate of 98.95% for prediction error within ±0.0007%.

3. The hybrid model of SAPSO-ANN is developed based on optimization of the initial weights and biases of the ANN model to further elevate the model prediction accuracy, which can obtain the smallest prediction mean error by comparing with the ANN model and the MLR model.

4. Implementation of the SAPSO-ANN model to analyze the relationship between desulfurization rate and process parameters show that the desulfurization rate is increased with the increase of lime weight, $S_{\text{in}}$, content, hot metal temperature and hot metal weight. The amount of desulfurization flux can be calculated to obtain the desired desulfurization rate and the calculated results can be verified through experiments.

REFERENCES

[1] F. He and L. Zhang, “Prediction model of end-point phosphorus content in BOF steelmaking process based on PCA and BP neural network,” J. Process Control, vol. 66, pp. 51–58, Jun. 2018.

[2] C. Gao, M. Shen, X. Liu, L. Wang, and M. Chen, “End-point prediction of BOF steelmaking based on KNNWTSVR and LWOA,” Trans. Indian Inst. Met., vol. 72, no. 1, pp. 257–270, Jan. 2019.

[3] H.-N. Zhang, A.-J. Xu, J. Cui, D.-F. He, and N.-Y. Tian, “Establishment of neural network prediction model for terminative temperature based on grey theory in hot metal pretreatment,” J. Iron Steel Res. Int., vol. 19, no. 6, pp. 25–29, Jun. 2012.

[4] X. Huang and M. Han, “Greedy kernel components acting on ANFIS to predict BOF steelmaking endpoint,” IFAC Proc. Volumes, vol. 41, no. 2, pp. 11007–11012, 2008.

[5] X. Wang, M. Han, and J. Wang, “Applying input variables selection technique on input weighted support vector machine modeling for BOF endpoint prediction,” Eng. Appl. Artif. Intell., vol. 23, no. 6, pp. 1012–1018, Sep. 2010.

[6] W. Yang, H. J. Meng, Y. J. Huang, and Z. Xie, “Prediction on molten steel end temperature during tapping in BOF based on LS-SVM and PSO,” Adv. Mater. Res., vol. 508, pp. 233–236, Jun. 2012.

[7] H. Liu, B. Wang, and X. Xiong, “Basic oxygen furnace steelmaking endpoint prediction based on computer vision and general regression neural network,” Optik, vol. 125, no. 18, pp. 5241–5248, Sep. 2014.

[8] M. Han and C. Liu, “Endpoint prediction model for basic oxygen furnace steel making based on membrane algorithm evolving extreme learning machine,” Appl. Soft Comput., vol. 19, pp. 430–437, Jun. 2014.

[9] H.-B. Wang, A.-J. Xu, L.-X. Ai, and N.-Y. Tian, “Prediction of endpoint phosphorus content of molten steel in BOF using weighted K-means and GM/D neural network,” J. Iron Steel Res. Int., vol. 19, no. 1, pp. 11–16, Jan. 2012.

[10] Z. Wang, F. Xie, B. Wang, Q. Liu, X. Lu, L. Hu, and F. Cai, “The control and prediction of end-point phosphorus content during BOF steelmaking process,” steel res. int., vol. 85, no. 4, pp. 599–606, Apr. 2014.

[11] Z. Wang, J. Chang, Q.-P. Ju, F.-M. Xie, B. Wang, H.-W. Li, B. Wang, X.-C. Lu, G.-Q. Fu, and Q. Liu, “Prediction model of end-point manganese content for BOF steelmaking process,” ISIJ Int., vol. 52, no. 9, pp. 1585–1590, 2012.

[12] B. Rezaee, “Desulfurization process using Takagi–Sugeno–Kang fuzzy modeling,” Int. J. Adv. Manuf. Technol., vol. 46, nos. 1–4, pp. 191–197, Jan. 2010.

[13] M. Han and Y. Zhao, “Dynamic control model of BOF steelmaking process based on ANFIS and robust relevance vector machine,” Expert Syst. Appl., vol. 38, no. 12, pp. 14786–14798, Nov. 2011.

[14] H. Tian, Z. Mao, and Y. Wang, “Hybrid modeling of molten steel temperature prediction in LF,” ISIJ Int., vol. 48, no. 1, pp. 58–62, 2008.

[15] H.-X. Tian and Z.-Z. Mao, “An ensemble ELM based on modified adaboost.r_t algorithm for predicting the temperature of molten steel in Ladle Furnace,” IEEE Trans. Autom. Sci. Eng., vol. 7, no. 1, pp. 73–80, Jan. 2010.

[16] Q. Li, G. Cao, J. Li, and N. Wang, “Process estimated temperature model of molten steel in LF based on neural network combined with expert system,” Appl. Mech. Mater., vol. 48–49, pp. 853–857, Feb. 2011.

[17] K. Feng, D. He, A. Xu, and H. Wang, “End temperature prediction of molten steel in LF based on CBR-BBN,” Steel Res. Int., vol. 87, no. 1, pp. 79–86, Jan. 2016.

[18] Q. Yang, J. Zhang, and Z. Yi, “Predicting molten steel endpoint temperature using a feature-weighted model optimized by mutual learning cuckoo search,” Appl. Soft Comput., vol. 83, Oct. 2019, Art. no. 105675.

[19] X. Wang, M. You, Z. Mao, and P. Yuan, “Tree-structure ensemble general regression neural networks applied to predict the molten steel temperature in Ladle Furnace,” Adv. Eng. Informat., vol. 30, no. 3, pp. 368–375, Aug. 2016.

[20] X. Wang, P. Yuan, Z. Mao, and M. You, “Molten steel temperature prediction model based on bootstrap feature subsets ensemble regression trees,” Knowl.-Based Syst., vol. 101, pp. 48–59, Jun. 2016.

[21] K. Feng, H. Wang, A. Xu, and D. He, “Endpoint temperature prediction of molten steel in RH using improved case-based reasoning,” Int. J. Minerals, Metallurgy, Mater., vol. 20, no. 12, pp. 1148–1154, 2013.

[22] I.-H. Ji, R.-Q. Liang, and J.-C. He, “Simulation on mixing behavior of desulfurizer and high-sulfur hot metal based on variable-velocity stirring,” ISIJ Int. ISIJ Int., vol. 56, no. 5, pp. 794–802, 2016.

[23] J. He, “Experimental research of external desulfurization in situ mechanical stirring,” J. Iron Steel Res. Int., vol. 18, no. S2, pp. 119–124, 2011.

[24] D.-S. Huang and F.-C. Huang, “Coupled thermo-fluid stress analysis of Kambara reactor with various anchors in the stirring of molten iron at extremely high temperatures,” Appl. Thermal Eng., vol. 73, no. 1, pp. 222–228, Dec. 2014.
Z. Tong, G. Ma, X. Cai, Z. Xue, W. Wang, and X. Zhang, “Characterization and valorization of Kanbara reactor desulfurization waste slag of hot metal pretreatment,” Waste Biomass Valor., vol. 7, no. 1, pp. 1–8, Feb. 2016.

X. Zhu, “Research and application of prediction model of hot metal desulfurization based on RBF neural network,” M.S. thesis, Dept. Mater. Metall., Northeastern Univ., Shenyang, China, 2010.

F. Zhonghua, Y. Ningchuan, W. Yanping, H. Yan, and W. Ling, “Analysis of factors influencing the effect of KR desulfurization,” Ind. Heating, vol. 42, no. 1, pp. 57–59, 2013.

U. Ali, W. Muhammad, A. Brahme, O. Skiba, and K. Inal, “Application of artificial neural networks in micromechanics for polycrystalline metals,” Int. J. Plasticity, vol. 120, pp. 205–219, Sep. 2019.

P. J. G. Nieto, E. García-Gonzalo, J. C. A. Antón, V. M. G. Suárez, R. M. Bayón, and F. M. Martín, “A comparison of several machine learning techniques for the centerline segregation prediction in continuous cast steel slabs and evaluation of its performance,” J. Comput. Appl. Math., vol. 330, pp. 877–895, Mar. 2018.

Z. Tong, G. Ma, X. Cai, Z. Xue, W. Wang, and X. Zhang, “Characterization and valorization of Kanbara reactor desulfurization waste slag of hot metal pretreatment,” Waste Biomass Valor., vol. 7, no. 1, pp. 1–8, Feb. 2016.

X. Zhu, “Research and application of prediction model of hot metal desulfurization based on RBF neural network,” M.S. thesis, Dept. Mater. Metall., Northeastern Univ., Shenyang, China, 2010.

F. Zhonghua, Y. Ningchuan, W. Yanping, H. Yan, and W. Ling, “Analysis of factors influencing the effect of KR desulfurization,” Ind. Heating, vol. 42, no. 1, pp. 57–59, 2013.

U. Ali, W. Muhammad, A. Brahme, O. Skiba, and K. Inal, “Application of artificial neural networks in micromechanics for polycrystalline metals,” Int. J. Plasticity, vol. 120, pp. 205–219, Sep. 2019.

P. J. G. Nieto, E. García-Gonzalo, J. C. A. Antón, V. M. G. Suárez, R. M. Bayón, and F. M. Martín, “A comparison of several machine learning techniques for the centerline segregation prediction in continuous cast steel slabs and evaluation of its performance,” J. Comput. Appl. Math., vol. 330, pp. 877–895, Mar. 2018.