Prediction of lime utilization ratio of dephosphorization in BOF steelmaking based on online sequential extreme learning machine with forgetting mechanism

Runhao Zhang, Jian Yang✉*, Han Sun, and Wenkui Yang*

State Key Laboratory of Advanced Special Steel, School of Materials Science and Engineering, Shanghai University, Shanghai 200444, China

Abstract: The machine learning models of multiple linear regression (MLR), support vector regression (SVR), and extreme learning machine (ELM) and the proposed ELM models of online sequential ELM (OS-ELM) and OS-ELM with forgetting mechanism (FOS-ELM) are applied in the prediction of the lime utilization ratio of dephosphorization in the basic oxygen furnace steelmaking process. The ELM model exhibites the best performance compared with the models of MLR and SVR. OS-ELM and FOS-ELM are applied for sequential learning and model updating. The optimal number of samples in validity term of the FOS-ELM model is determined to be 1500, with the smallest population mean absolute relative error (MARE) value of 0.058226 for the population. The variable importance analysis reveals lime weight, initial P content, and hot metal weight as the most important variables for the lime utilization ratio. The lime utilization ratio increases with the decrease in lime weight and the increases in the initial P content and hot metal weight. A prediction system based on FOS-ELM is applied in actual industrial production for one month. The hit ratios of the predicted lime utilization ratio in the error ranges of ±1%, ±3%, and ±5% are 61.16%, 90.63%, and 94.11%, respectively. The coefficient of determination, MARE, and root mean square error are 0.8670, 0.06823, and 1.4265, respectively. The system exhibits desirable performance for applications in actual industrial production.

Keywords: basic oxygen furnace steelmaking; machine learning; lime utilization ratio; dephosphorization; online sequential extreme learning machine; forgetting mechanism

1. Introduction

Basic oxygen furnace (BOF) steelmaking is an important process in integrated steel companies. BOF steelmaking mainly functions to control carbon content and temperature and eliminate impurities [1]. As a harmful element for most steel grades, phosphorus results in the cold brittleness and decreased strength of steel products. To reduce the negative influence of phosphorus, the phosphorus contents must to be reduced to low values as much as possible.

Lime with calcium oxide (CaO) as the main component is usually used as the dephosphorization flux. The dicalcium silicate ($2CaO \cdot SiO_2$) in the slag phase can easily combine with the oxidation product of phosphorus (P_2O_5) to form the stable solid solution of $2CaO \cdot SiO_2 - 3CaO \cdot P_2O_5$ [2]. From a thermodynamic perspective, the dephosphorization reaction benefits from the low temperature in the early stage of the BOF steelmaking process [3–4]. However, a low temperature retards the melting rate of lime particles. Moreover, a layer of $2CaO \cdot SiO_2$ with a high melting point is generated on the surface of lime particles, which inhibits the inner part of lime particles from melting further [5]. As a consequence, a part of

the lime is excluded in the dephosphorization reaction during the BOF process, which results in a low lime utilization ratio.

The precise prediction of the lime utilization ratio is helpful to deepen the understanding of the melting behavior of lime particles in BOFs, which benefits process optimization. However, due to the high temperature and complex physicochemical behavior, the models based on metallurgical mechanisms generally give low prediction accuracies [6–7]. In recent years, the technology of machine learning shows its outstanding capability in establishing relationships between inputs and outputs. The popularization of computers and informatization leads to the generation of large amounts of data daily in steelmaking plants. Therefore, data-driven models have been established based on machine learning algorithms for endpoint predictions in hot metal pretreatment [8], BOF steelmaking [9–11], electric-arc furnace steelmaking [12–14], secondary refining [15–17], continuous casting [18–19], and so on. As a part of machine learning, artificial neural network (ANN) is the core of current hotspots in the field of artificial intelligence. Among various types of ANNs, the backpropagation neural network (BPNN), extreme learning machine (ELM), convolutional neural network, and deep

---

🍀 Springer

neural network (also called deep learning) are commonly used in handling prediction tasks in steelmaking processes. The advantage of ELM lies in its extremely fast training speed. Chen *et al.* [20] applied ELM to predict the endpoint temperature of vacuum degassing refining and showed a lower prediction error than the BPNN model. The ELM model has an extremely short training time of 0.029 s, and the model BPNN requires 4.018 s to train.

Most investigations on dephosphorization behavior prediction in the BOF steelmaking process focus on the prediction of the endpoint phosphorus content [6,21–24]. Liu *et al.* [25] proposed a prediction model for BOF endpoint phosphorus and oxygen contents that combines principal component analysis, genetic algorithm, and BPNN. Zhou *et al.* [26] established a monotone-constrained BPNN model for endpoint phosphorus content prediction. Acosta *et al.* [27] applied relevance vector regression and support vector regression (SVR) optimized by a self-adaptive differential evolution algorithm for regression to model the phosphorus endpoint content in the BOF steelmaking process based on industrial data. All the above models present high prediction accuracies and low errors, which indicates that the machine learning models can predict well the endpoint phosphorus content in BOFs. However, the lime utilization ratio of dephosphorization in BOFs lacks research attention. In addition, most investigations that use machine learning methods in steelmaking processes focus on the prediction accuracy but barely give attention on the variable importance [28].

The present article aims to establish a prediction model of the lime utilization ratio of dephosphorization in BOFs. Industrial production data from a steelmaking plant are collected. Data preprocessing methods, such as variable selection, abnormal data elimination, and data normalization, are employed. Three basic machine learning models, including those of multiple linear regression (MLR), SVR, and the original ELM, are compared. Based on the original ELM model, the modified ELM models, that is, online sequential ELM (OS-ELM) and the OS-ELM model with forgetting mechanism (FOS-ELM), are proposed to learn from the samples successively and incrementally. The optimal number of samples in the validity term of the FOS-ELM model is determined, and the variable importance of the proposed FOS-ELM model is quantized based on the mean impact value (MIV). Finally, the system for the prediction of lime utilization ratio is developed based on FOS-ELM, and a notably good performance is achieved.

This paper proposes three innovation points: (1) Innovation in the research subject. The prediction of lime utilization ratio during dephosphorization in the BOF process using machine learning models has not been conducted. (2) Innovation in research methodology. Most previous investigations train and test the machine learning models using randomly divided data. In the present paper, the proposed FOS-ELM can deal with time sequential data, which show a pattern close to data generation in steelmaking plants. Furthermore, outdated samples can be removed in the FOS-ELM model, which cannot be achieved by other machine learning models.

(3) Innovation in research results. The variable importance is quantified using the MIV to evaluate the importance of process parameters in the lime utilization ratio.

2. Data preprocessing

Experimental data are collected from Hebei Iron and Steel Group Co., Ltd. in China for 10 months. The original database contains the BOF steelmaking production data with more than 13000 heats. In the database, the steel grades include the Q345 structural steel, SWRCH35K cold heading steel, HRB400EC rebar, and so on. As the original data cannot be directly trained using the machine learning models, data preprocessing is conducted.

The lime utilization ratio of dephosphorization (η, %) indicates the proportion of CaO amount used for dephosphorization to the total CaO amount in the charged lime. Phosphorus combines with CaO to form $3CaO \cdot P_2O_5$ in the slag phase. As a result, 2 mol P consumes 3 mol CaO. Based on the mass balance, the lime utilization ratio is calculated using Eq. (1), where M_{CaO} and M_P refer to the molar masses of CaO and phosphorus (g·mol^{-1}), respectively; W_m and W_{lime} denote the weights of hot metal and lime (t), respectively; $[\%P]_i$ and $[\%P]_{end}$ indicate the initial P content of the hot metal and endpoint P contents of molten steel (wt%), respectively. The value of 0.88 in Eq. (1) means that the CaO content of the lime is 88wt%. η is the target output variable of prediction models and labeled y.

$$\eta = \frac{3 \times M_{CaO} W_m ([\%P]_i - [\%P]_{end})}{2 \times 0.88 \times M_P W_{lime}} \qquad (1)$$

Variable selection is conducted to remove unnecessary variables from the dataset. Among the original 28 variables, 17 key variables are selected manually as input variables, as listed in Table 1. The 17 variables are labeled x_1–x_{17}.

Given the harsh environment of steelmaking plants, abnormal data are inevitably generated because of sampling failure, data loss during data transformation and storage, and so on. In order to eliminate the adverse impact of abnormal data on the models, data cleaning is conducted. Through manual elimination of abnormal samples, 9575 samples are retained in this study. In addition, the Pauta criterion is applied for further data cleaning [29].

A total of 5526 samples are retained after data cleaning. Compared with the original dataset, less than half of the samples can be used for machine learning model establishment in this study. Table 1 lists the statistical results of the variables, including the output variable of the lime utilization ratio and 17 input variables. Fig. 1 illustrates the Pearson correlation coefficients of the variables, which display a linear relationship between them. The red and blue circles indicate positive and negative correlations, respectively. The sizes of the circles are related to the absolute values of correlation coefficients. Among the 17 input variables, scrap weight, initial P content, and lime and limestone weights have significant linear relationships with the lime utilization ratio, with the Pearson correlation coefficient values of −0.56, 0.59, −0.86,

Table 1.   Statistical results on the variables after data cleaning

Variable	Label	Minimum	Maximum	Mean	Standard deviation
η / %	y	3.82	23.41	11.84	3.88
Scrap weight / t	x_1	12.0	50.0	27.8	7.6
Hot metal weight / t	x_2	108.0	140.0	124.0	5.4
[%Si]$_i$ / wt%	x_3	0.1175	0.5750	0.3401	0.0785
[%S]$_i$ / wt%	x_4	0.0008	0.0607	0.0282	0.0109
[%P]$_i$ / wt%	x_5	0.0802	0.1506	0.1148	0.0120
[%Mn]$_i$ / wt%	x_6	0.0876	0.2845	0.1856	0.0331
Initial temperature / °C	x_7	1227.0	1449.0	1339.8	37.8
Lime weight / t	x_8	1.415	6.095	3.469	0.879
Dolomite weight / t	x_9	0	3.920	1.642	0.762
Limestone weight / t	x_{10}	0	3.410	0.770	0.887
Sintered ore weight / t	x_{11}	0	4.035	0.801	1.086
Coke weight / t	x_{12}	0	1.295	0.239	0.353
Main blowing time / s	x_{13}	701.0	977.0	838.7	46.2
Second blowing time / s	x_{14}	0	200.0	95.0	35.2
Oxygen consumption amount / m³	x_{15}	4939.0	7073.0	5976.5	365.6
Argon consumption amount / m³	x_{16}	0	194.0	36.5	53.2
Nitrogen consumption amount / m³	x_{17}	0	665.0	192.1	157.8

Note: [%Si]$_i$, [%S]$_i$, and [%Mn]$_i$ are the initial Si, S, and Mn content of the hot metal, respectively.
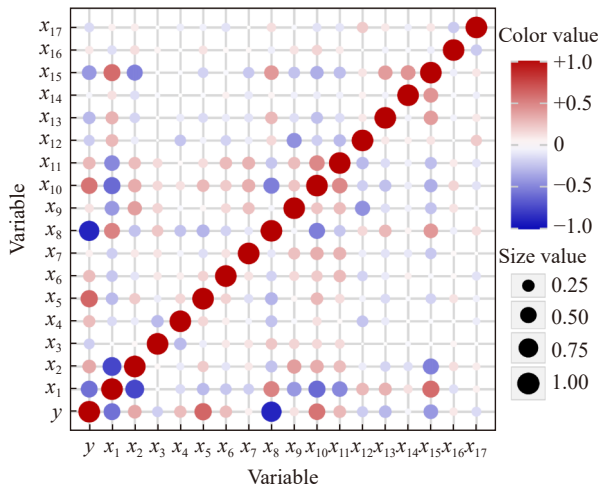
Fig. 1.   Pearson correlation coefficients of the variables.

and 0.54, respectively.

Table 1 reveals that the variables differ in scale and magnitude. Variables with large magnitudes are assigned high weights and become important in model training. Meanwhile, differences in the magnitude of variables hinder the convergence rate of iterative algorithms and increase the training time. Data normalization is applied to normalize the variable values to the range of [0, 1] [28].

3. Description of the proposed models

3.1. ELM

The ELM is a single hidden layer feedforward network established by Huang *et al.* in 2004; it randomly selects input weights and hidden layer biases [30–31]. For a dataset with N samples, the mathematical model of ELM is displayed in Eq. (2).

$$y_j = \sum_{i=1}^{\tilde{N}} \beta_i g\left(\omega_i^{\mathrm{T}} x_j + b_i\right), j = 1, 2, \ldots, N \quad (2)$$

where $x_j = \left[x_{j1}, x_{j2}, \ldots, x_{jn}\right]^{\mathrm{T}}$ and $y_j = \left[y_{j1}, y_{j2}, \ldots, y_{jm}\right]^{\mathrm{T}}$, and n and m indicate the dimensions of input and output data, respectively; $\tilde{N}$ refers to the number of hidden nodes; β corresponds to the weight connecting the hidden node to the output node; $g\left(\omega_i^{\mathrm{T}} x_j + b_i\right)$ denotes the output of the hidden node. Eq. (2) can be converted into the form of matrix, as shown in Eq. (3).

$$H\beta = Y \quad (3)$$

where

$$H = \begin{bmatrix} g\left(\omega_1^{\mathrm{T}} x_1 + b_1\right) & \cdots & g\left(\omega_{\tilde{N}}^{\mathrm{T}} x_1 + b_{\tilde{N}}\right) \\ \vdots & \ddots & \vdots \\ g\left(\omega_1^{\mathrm{T}} x_N + b_1\right) & \cdots & g\left(\omega_{\tilde{N}}^{\mathrm{T}} x_N + b_N\right) \end{bmatrix}_{N \times \tilde{N}} \quad (4)$$

$$\beta = \begin{bmatrix} \beta_1^{\mathrm{T}} \\ \vdots \\ \beta_{\tilde{N}}^{\mathrm{T}} \end{bmatrix}_{\tilde{N} \times m} \quad (5)$$

$$Y = \begin{bmatrix} y_1^{\mathrm{T}} \\ \vdots \\ y_N^{\mathrm{T}} \end{bmatrix}_{N \times m} \quad (6)$$

H is called the hidden layer output matrix. Fig. 2 shows the schematic of the ELM.

Among all the least-squares solutions of $H\beta = Y$, a unique solution of $\widehat{\beta}$ with the smallest norm is obtained, as displayed in Eq. (7).

$$\widehat{\beta} = \left(H^{\mathrm{T}} H\right)^{-1} H^{\mathrm{T}} Y \quad (7)$$

ELM exhibits an extremely higher learning speed compared with other algorithms, such as BPNN. This property is attributed to the following reasons. First, the input weights
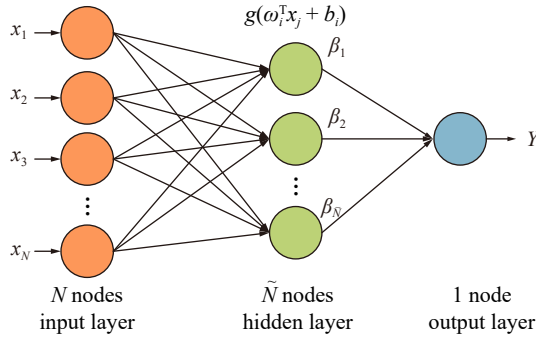
Fig. 2.   Schematic of the extreme learning machine.

and hidden layer biases in ELM are randomly assigned, which diminishes the time needed for the iterative learning process. Second, ELM can process data in batches. Thus, parallel and matrix calculations can efficiently hasten the training process.

3.2. OS-ELM

In steelmaking plants, new data are constantly generated via uninterrupted industrial production. The model needs to be retrained using past and new data, which consumes a considerable amount of time, especially when the dataset increasingly becomes larger. In 2006, Liang *et al.* [32] established the OS-ELM. The online sequential learning algorithm allows the model to learn new data one by one or chunk by chunk without retraining the whole dataset. According to Eqs. (3) and (7), the new ELM model can be expressed as Eqs. (8)–(11). $\beta^{(0)}$ and $\beta^{(1)}$ indicate the weights of trained and new models, respectively.

$$\begin{bmatrix} H_0 \\ H_1 \end{bmatrix} \beta = \begin{bmatrix} Y_0 \\ Y_1 \end{bmatrix} \tag{8}$$

$$\beta^{(1)} = \left(\begin{bmatrix} H_0 \\ H_1 \end{bmatrix}^{\mathrm{T}} \begin{bmatrix} H_0 \\ H_1 \end{bmatrix} \right)^{-1} \begin{bmatrix} H_0 \\ H_1 \end{bmatrix}^{\mathrm{T}} \begin{bmatrix} Y_0 \\ Y_1 \end{bmatrix} =$$
$$\beta^{(0)} + P_1 H_1^{\mathrm{T}} \left(Y_1 - H_1 \beta^{(0)} \right) \tag{9}$$

where

$$P_1 = \left(H_0^{\mathrm{T}} H_0 + H_1^{\mathrm{T}} H_1 \right)^{-1} \tag{10}$$

$$\beta^{(0)} = \left(H_0^{\mathrm{T}} H_0 \right)^{-1} H_0^{\mathrm{T}} Y_0 \tag{11}$$

Based on Eqs. (9)–(11), the recursive formula for $(k + 1)$th, $\beta^{(k+1)}$ can be obtained as displayed in Eq. (12).

$$\beta^{(k+1)} = \beta^{(k)} + P_{k+1} H_{k+1}^{\mathrm{T}} \left(Y_{k+1} - H_{k+1} \beta^{(k)} \right) \tag{12}$$

where

$$P_{k+1} = P_k - P_k H_{k+1}^{\mathrm{T}} \left(I + H_{k+1} P_k H_{k+1}^{\mathrm{T}} \right)^{-1} H_{k+1} P_k \tag{13}$$

where I is the identity matrix. As the initial ELM model has been trained with the past dataset, $\beta^{(k)}$ and P_k are known parameters. Only Eqs. (12) and (13) are needed to calculate H_{k+1} and Y_{k+1} from the new dataset.

3.3. FOS-ELM

In the steelmaking process, the production conditions may change because of process optimizations, replacement of raw

material suppliers, or other reasons. Thus, the older the data, the less relevant they are to the current production status. Outdated data should be removed in subsequent model updates. The OS-ELM can deal with newly added data but cannot delete the original ones. Zhao *et al.* [33] proposed a modified ELM model, named FOS-ELM, that can gradually expel outdated data. For data with a validity term, the term can have values of several days, weeks, or months. The s indicates the number of samples within the validity term. According to Eqs. (3) and (7), the new ELM model can be expressed as Eqs. (14) and (15).

$$\begin{bmatrix} H_{k-s+1} \\ \vdots \\ H_k \end{bmatrix} \beta = \begin{bmatrix} Y_{k-s+1} \\ \vdots \\ Y_k \end{bmatrix} \tag{14}$$

$$\beta^{(k)} = \left(\begin{bmatrix} H_{k-s+1} \\ \vdots \\ H_k \end{bmatrix}^{\mathrm{T}} \begin{bmatrix} H_{k-s+1} \\ \vdots \\ H_k \end{bmatrix} \right)^{-1} \begin{bmatrix} H_{k-s+1} \\ \vdots \\ H_k \end{bmatrix}^{\mathrm{T}} \begin{bmatrix} Y_{k-s+1} \\ \vdots \\ Y_k \end{bmatrix} =$$
$$P_k \begin{bmatrix} H_{k-s+1} \\ \vdots \\ H_k \end{bmatrix}^{\mathrm{T}} \begin{bmatrix} Y_{k-s+1} \\ \vdots \\ Y_k \end{bmatrix} \tag{15}$$

Keep the total number of samples in the training data unchanged, add a new sample, and delete one original sample simultaneously. Then, the smallest norm solution can be expressed as Eq. (16).

$$\beta^{(k+1)} = P_{k+1} \begin{bmatrix} H_{k-s+2} \\ \vdots \\ H_{k+1} \end{bmatrix}^{\mathrm{T}} \begin{bmatrix} Y_{k-s+2} \\ \vdots \\ Y_{k+1} \end{bmatrix} \tag{16}$$

Through matrix calculations, the recursive formulas for P_{k+1} and $\beta^{(k+1)}$ can be written as Eqs. (17) and (18), respectively.

$$P_{k+1} = P_k - P_k \begin{bmatrix} -H_{k-s+1} \\ H_{k+1} \end{bmatrix}^{\mathrm{T}} \left(I + \right.$$
$$\left. \begin{bmatrix} H_{k-s+1} \\ H_{k+1} \end{bmatrix} P_k \begin{bmatrix} -H_{k-s+1} \\ H_{k+1} \end{bmatrix}^{\mathrm{T}} \right)^{-1} \begin{bmatrix} H_{k-s+1} \\ H_{k+1} \end{bmatrix} P_k \tag{17}$$

$$\beta^{(k+1)} = \beta^{(k)} + P_{k+1} \begin{bmatrix} -H_{k-s+1} \\ H_{k+1} \end{bmatrix}^{\mathrm{T}} \left(\begin{bmatrix} Y_{k-s+1} \\ Y_{k+1} \end{bmatrix} - \begin{bmatrix} H_{k-s+1} \\ H_{k+1} \end{bmatrix} \beta^{(k)} \right) \tag{18}$$

When the sample number k is smaller than $s − 1$, no sample is deleted. In this case, FOS-ELM and OS-ELM yield the same results because all data are valid.

3.4. Modeling method

Through data preprocessing, 5526 samples are retained for model training and testing. Fig. 3 shows the division method of the training and testing sets, which are indicated by the green and red grids, respectively. For the comparison of the three basic machine learning algorithms, namely, MLR, SVR, and ELM, the prepared dataset is randomly divided into two parts, with 80% of the samples used as the training set and the remaining 20% as the testing set to verify the prediction accuracy of the model. The same random seed is applied to ensure the application of the same division methods in
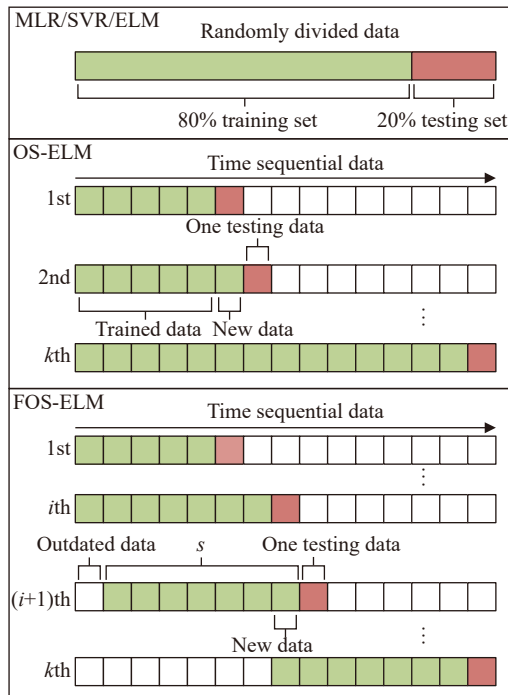
512

Int. J. Miner. Metall. Mater., Vol. 31 , No. 3 , Mar. 2024

Fig. 3.   Division method of training and testing sets.

these three models.

In OS-ELM model training, the samples are arranged in a time sequential order; that is, data are generated one by one with time in the steelmaking plant. The first step involves training the initial OS-ELM model using the first 500 samples. The initial model is used to predict the lime utilization ratio of the next heat, and the prediction accuracy can be obtained. Then, the 501st sample is added to the training set as new data. The OS-ELM model is retrained based on Eq. (12). Afterward, the retrained model is used to predict the result for the 502nd sample. The training process is performed repeatedly until the last sample becomes a part of the testing data.

In the FOS-ELM model, the samples are also arranged in a time sequential order. The initial FOS-ELM model is established using a training set containing 500 samples and retrained when new samples are input. Given the validity term for the data, when the sample number in the trained model is larger than the value of s, outdated data are removed to retain the total number of samples in the training data. In the present work, the s is set at a fixed value. During production, outdated data are deleted from the training set, and new ones are added to establish a continuously updated model. It can be seen that the operation modes of OS-ELM and FOS-ELM are appropriate for steelmaking production.

All the models in this study are developed using the Python programming language. The Scikit-learn [34] package is used to establish the MLR and SVR models. ELM and OS-ELM are implemented using a Python package named pyoselm developed by Ferrado [35]. Given the lack of a ready-made FOS-ELM algorithm for Python, we develop the FOS-ELM model ourselves based on pyoselm. Modeling is conducted on a Windows personal computer with Intel(R) Core(TM) i5-8400 CPU@2.80 GHz, NVIDIA(R) GeForce GTX(TM) 1050Ti GPU, and 16 GB RAM. Grid searching is performed to obtain the optimal hyperparameters of the models [36].

4. Results and discussion

4.1. Performance of the different models

To quantify the performance of the three basic machine learning models of MLR, SVR, and ELM in predicting the lime utilization ratio, three criterions are implemented. The three criterions of coefficient of determination R-squared (R^2), mean absolute relative error (MARE), and root mean square error (RMSE) can be calculated using Eqs. (19)–(21), respectively.

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(M_i - P_i)^2}{\sum_{i=1}^{N}\left(M_i - \bar{M}\right)^2} \tag{19}$$

$$\text{MARE} = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{P_i - M_i}{M_i}\right| \tag{20}$$

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

where M_i and P_i are the measured and predicted values, respectively. $\bar{M}$ denotes the mean value of the measured values. R^2 indicates the proportion of variation in the dependent variable that is predictable from the independent variable (range: 0–1). When the value of R^2 is close to 1, the model can explain the variability of data well. MARE and RMSE reflect the prediction errors from the point of view of relative and absolute errors, respectively. A model with good prediction accuracy has small values of MARE and RMSE. The prediction performances of MLR, SVR, and ELM models are compared, as shown in Table 2. Among the three models, the ELM model exhibits the best performance, with the largest R^2 value of 0.9486 and the smallest MARE and RMSE values of 0.06221 and 0.8759, respectively. Regarding the training time, MLR has the fastest training time of 0.003 s because of its simple calculation method. The training time of the SVR model (0.235 s) is about nine times longer than that of the ELM model (0.026 s).

Fig. 4 presents the comparisons between the measured and predicted lime utilization ratios obtained using the three basic models. Fig. 4(a) shows the arc-shaped distribution of the predicted η, indicating that the MLR has a low prediction accuracy when the measured value of η deviates from the average value. It can be seen that the measured and predicted lime utilization ratios of the ELM model have the best fitting degree, with the R^2 of 0.9486.

Fig. 5 illustrates the box plots of the predicted errors for

Table 2.   Comparison of prediction performances of MLR, SVR, and ELM models

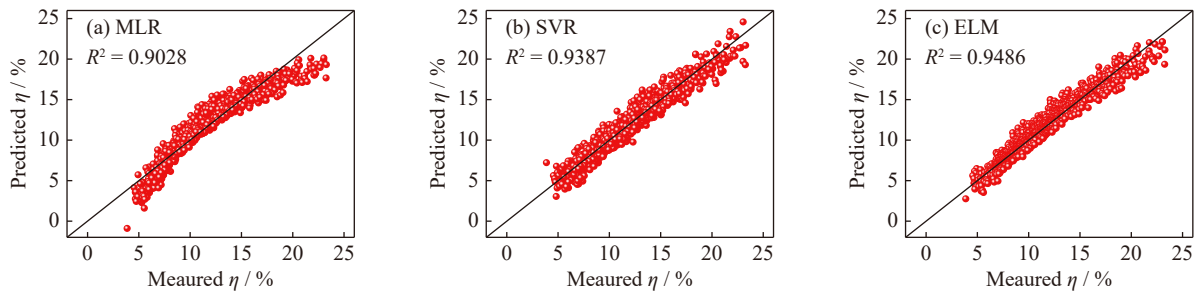| Model	R^2	MARE	RMSE	Training time / s
MLR	0.9028	0.09958	1.2045	0.003
SVR	0.9387	0.07212	0.9567	0.235
ELM	0.9486	0.06221	0.8759	0.026

Fig. 4.   Comparisons of the measured and predicted lime utilization ratios obtained using the (a) MLR, (b) SVR, and (c) ELM models..

the three models. The fluctuation range obtained for the ELM model is narrower than those of the MLR and SVR models, indicating that the ELM model achieves the smallest predicted error.
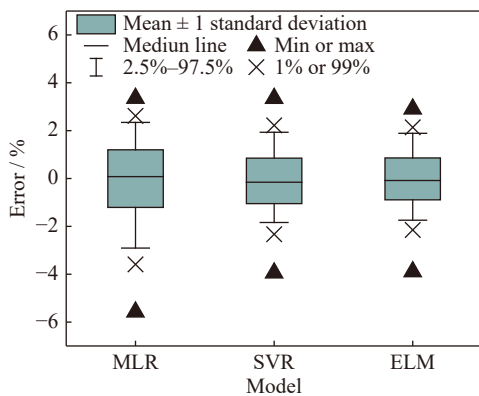

Fig. 5.   Box plots of the predicted errors of the MLR, SVR, and ELM models.

Fig. 6 displays the distributions of the predicted errors of the three models. The step size of the predicted error is set to 0.5, and the samples in each numerical range are counted. It can be seen that all the predicted error distributions obey a Gaussian distribution. Among the three models, the SVR model shows a biased error; that is, the predicted error interval with the largest proportion is outside the range of [−0.25, 0.25), and the MLR and ELM models obtain unbiased estimations.
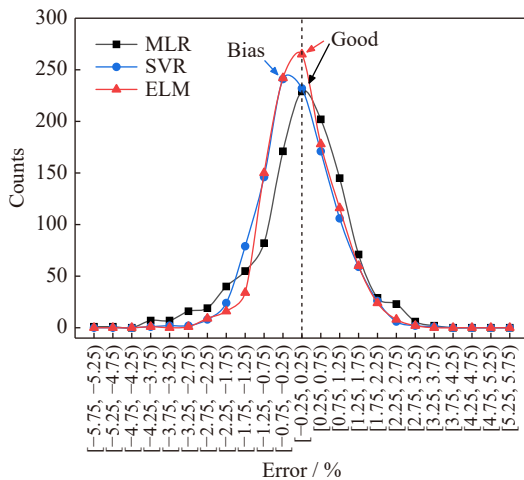

Fig. 6.   Distributions of the predicted errors of the MLR, SVR, and ELM models.

Compared with MLR and SVR models, it is obvious that ELM exhibits the best fitting degree of measured and predicted values, the narrowest fluctuation range in predicted errors, and no biased error.

The ELM model with 89 hidden nodes and the active function of sigmoid attains the best prediction performance through hyperparameter optimization of grid search. These hyperparameters are also used for the OS-ELM and FOS-ELM models. In the OS-ELM model, the samples are arranged in a time sequential order. The first 500 samples are used for the initial model training, and the remaining 5026 samples are utilized for sequential learning and model updating. The modeling method of OS-ELM can refer to Fig. 3. Every time the OS-ELM model makes a prediction, the predicted values obtained previously are collected for the calculation of MARE. Fig. 7 shows the changes in the MARE with the number of updates. An unstable zone is obtained when the number of updates is less than 1500. MARE easily fluctuates when the sample number is small. As a data-driven model, the OS-ELM model needs large amounts of data for model training to achieve a high prediction accuracy. With the increase in the sample number in the training set, the MARE value decreases and tends to be stable. The final calculated MARE, namely population MARE, of the OS-ELM model is 0.059424.

The FOS-ELM model has a validity term for data. In this work, s samples are assumed to be included in the validity term. The s is set to a fixed value. When a new sample is added to the training set, one outdated sample will be removed from the training set to retain the size of the training set. The range value of s is set to 500 to 5500, with a step size of 500, to detect its effect on the population MARE. The result is displayed in Table 3. With the increase in s, the population MARE value decreases first and then increases. The smallest population MARE value of 0.058226 is obtained when s is 1500, indicating that the validity term containing 1500 samples is optimum for the present dataset to predict the lime utilization ratio in the BOF steelmaking process. When s is 500 or 1000, the training set is considerably small for an adequate fitting of the FOS-ELM model. When s is equal to or larger than 2000, the data with poor validity can still be used for model training, but it will result in a low prediction accuracy. When s is larger than 5526, all the samples are valid, and no data are removed from the training set. At this point, the population MARE value of the FOS-ELM model is the same as that of the OS-ELM model, both of which are 0.059424.
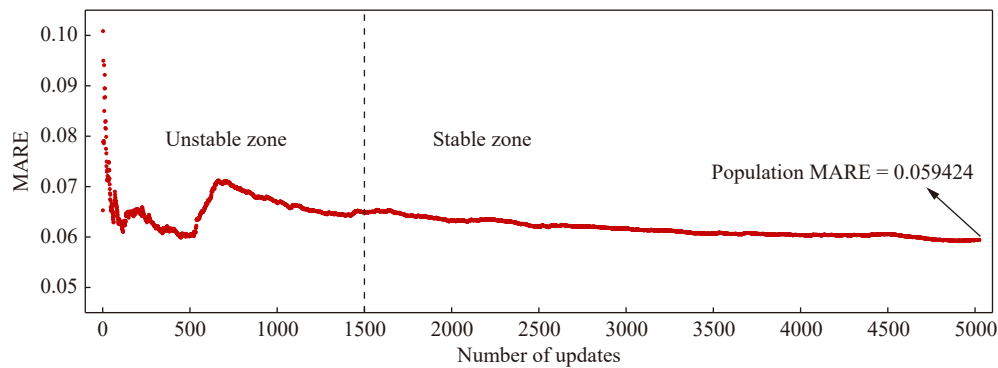
Fig. 7. Changes in MARE value with the number of updates obtained by the OS-ELM model.

Table 3. Population MARE values of the FOS-ELM model with various values of s

Value of s	Population MARE
500	0.061728
1000	0.058680
1500	0.058226
2000	0.058237
2500	0.058522
3000	0.058705
3500	0.058955
4000	0.059157
4500	0.059256
5000	0.059363
5500	0.059423
$\geq$5526	0.059424

Fig. 8. Values of MIV of various variables calculated using the established FOS-ELM model.

Based on the established FOS-ELM model, the variable importance is calculated to determine the process parameter with the greatest effect on lime utilization ratio prediction. The method of MIV is employed [6]. When the value of MIV is larger than 0, the variable has a positive impact on the lime utilization ratio. On the contrary, the variable with the value of MIV smaller than 0 has a negative impact. Fig. 8 shows the variable importance analysis results for various variables. The absolute values of MIV, which are sorted by the degree of importance, are obtained for comparison. The red and blue bars represent the values of MIV larger and smaller than 0, respectively. The three most important variables are the lime weight, initial P content, and hot metal weight. The lime weight shows a negative impact, which means that the lime utilization ratio increases with the decrease in the lime weight. The lime utilization ratio increases when the initial P content and hot metal weight are increased. This result dovetails well with the lime utilization ratio calculation formula of Eq. (1).

The variable importance analysis reveals that the most effective operation to increase the lime utilization ratio is reducing the amount of lime added per ton of hot metal, which means decreasing the lime weight and increasing the hot metal weight. When excess lime is charged into the converter, a portion of it is excluded from the dephosphorization reaction, resulting in wastage. As shown in Fig. 8, oxygen consumption ranks fifth in terms of the MIV. The increase in oxygen consumption enhances the stirring of the furnace bath. The
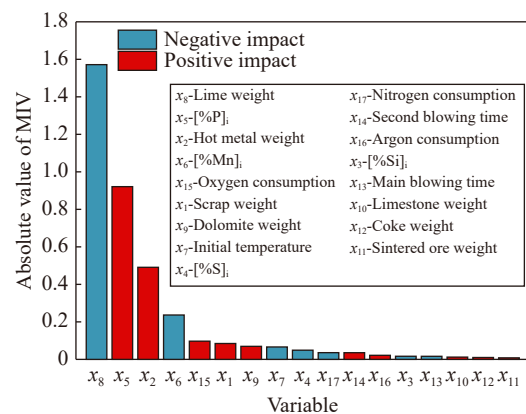
kinetic conditions can be improved to promote mass transfer between the hot metal and slag, which benefits the lime utilization efficiency. In recent years, the double slag converter steelmaking process (DSP) has been increasingly adopted by steelmaking plants [37]. In DSP, the decarburization slag is left in the converter and reused in the next heat; thus, lime consumption can be greatly reduced, and the lime utilization ratio can be enhanced [38].

4.2. Development of a prediction system based on the FOS-ELM model

The good performance of the FOS-ELM model in predicting the lime utilization ratio indicates its viability for implementation in actual industrial production. In this work, a lime utilization ratio prediction system for BOFs based on FOS-ELM is developed. Fig. 9 illustrates the flow chart of the prediction system. The current heat data of BOFs are read using a programmable logic controller and stored in the database as a new sample. Then, whether all variables in the current sample fall within the setting range is evaluated. The range here refers to the range between the minimum and maximum values in Table 1. If one or more variables are outside the bounds of the setting ranges, the current sample is used just for lime utilization prediction after data normalization. On the other hand, when all variables in the current sample fall within the setting ranges, the sample is stored as a valid heat, and an outdated sample is removed from the valid dataset to retain the original quantity. After data normalization, the valid sample is used not only to predict the lime utilization ratio
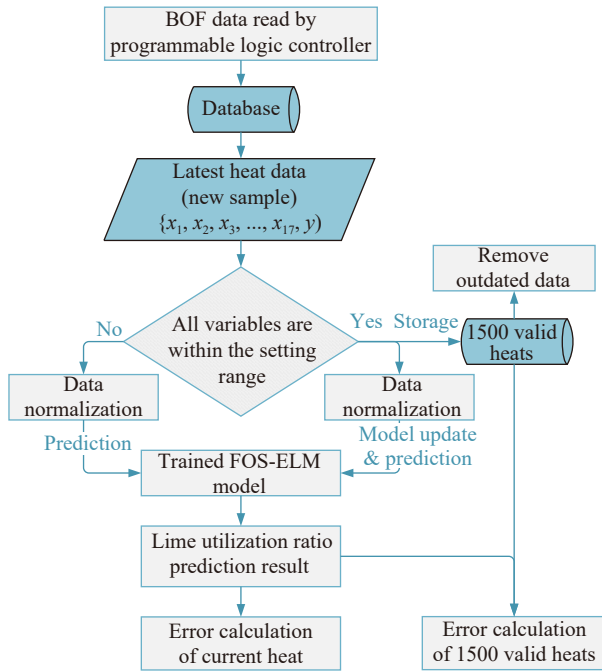
Fig. 9.   Flow chart of the prediction system based on FOS-ELM.

but also to update the FOS-ELM model. Finally, calculations of error of the current and 1500 valid heats are conducted. With the progress of BOF production, the above process is repeated when a new sample is generated.

A desktop application used in Windows operating system is developed by Microsoft Visual Studio. The desktop application provides a graphical user interface (GUI) as shown in Fig. 10, that allows operators to easily access real-time and historical data of a steelmaking plant. Once a new sample is obtained, the data are automatically collected and shown on the GUI. Clicking the "Prediction" button can be used to obtain the predicted lime utilization ratio, errors of the current heat, and the latest 1500 valid heats. The application also allows operators to optimize steelmaking operations before or during the processes to enhance the lime utilization ratios based on prediction results.

The lime utilization ratio prediction system is applied in actual industrial production for one month. All the 986 heats obtained in this month are used to verify the performance of the present system, as shown in Fig. 11. The hit ratios of 61.16%, 90.63%, and 94.11% of the predicted lime utilization ratio are acquired by the system in the error ranges of
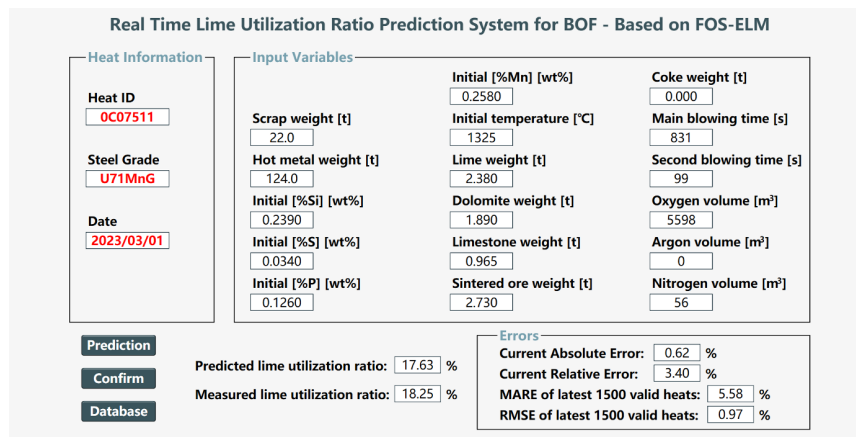

Fig. 10.   Graphical user interface of the real-time prediction system developed based on FOS-ELM.

±1%, ±3%, and ±5%, respectively. The values of R^2, MARE, and RMSE are 0.8670, 0.06823, and 1.4265, respectively. The system exhibits desirable performance for applications in actual industrial production.
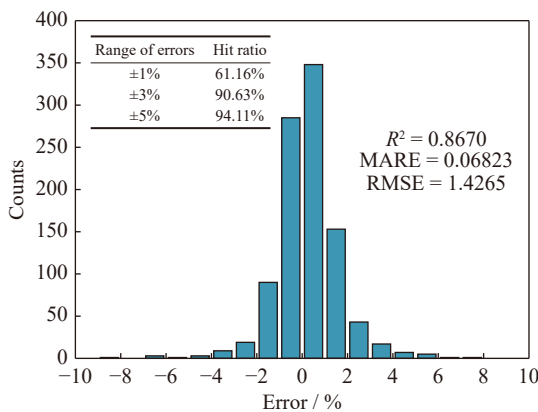

Fig. 11.   Performance of the prediction system using the production data for one month.

5. Conclusions

In the present work, the prediction of the lime utilization ratio of dephosphorization in the BOF steelmaking process is conducted using machine learning models, including MLR, SVR, and ELM models and the proposed OS-ELM and FOS-ELM models. The conclusions are obtained as follows.

(1) Among the three basic machine learning models, ELM performs the best with the largest R^2 value of 0.9486 and the smallest MARE and RMSE values of 0.06221 and 0.8759, respectively. The ELM model also attains the best fitting degree of the measured and predicted values, the narrowest fluctuation range in predicted errors, and no biased error.

(2) For the FOS-ELM model, with the increase in the number of samples in the validity term, the population MARE decreases first and then increases. The validity term containing 1500 samples is the optimum for the present dataset, with the smallest population MARE value of 0.058226.

(3) According to the variable importance evaluated using

516

Int. J. Miner. Metall. Mater., Vol. 31, No. 3, Mar. 2024

the MIV method, the three most important variables for the lime utilization ratio are the lime weight, initial P content, and hot metal weight. The lime utilization ratio increases with the decrease in lime weight, and the lime utilization ratio increases when the initial P content and hot metal weight are increased.

(4) A prediction system for the BOF steelmaking process based on FOS-ELM is developed and applied in actual industrial production for one month. The hit ratios of the predicted lime utilization ratio in the error ranges of ±1%, ±3%, and ±5% are 61.16%, 90.63%, and 94.11%, respectively. The values of R^2, MARE, and RMSE are 0.8670, 0.06823, and 1.4265, respectively. The system exhibits desirable performance for applications in actual industrial production.

Acknowledgements

Conflict of Interest

Jian Yang is an editorial board member for this journal and is not involved in the editorial review or the decision to publish this article. All authors state that there is no conflict of interest.

References

[1] X.H. Huang, *Principles of Iron and Steel Metallurgy*, 4th ed., Publishing house of Metallurgical Industry, Beijing, 2013.

[2] S.Y. Kitamura, H. Shibata, K.I. Shimauchi, and S.Y. Saito, The importance of dicalcium-silicate on hot metal dephosphorization reaction, *Rev. Met. Paris*, 105(2008), No. 5, p. 263.

[3] W.K. Yang, J. Yang, Y.Q. Shi, *et al.*, Effect of temperature on dephosphorization of hot metal in double-slag converter steelmaking process by high-temperature laboratorial experiments, *Steel Res. Int.*, 92(2021), No. 3, art. No. 2000438.

[4] H. Sun, J. Yang, X.W. Lu, *et al.*, Dephosphorization in double slag converter steelmaking process at different temperatures by industrial experiments, *Metals*, 11(2021), No. 7, art. No. 1030.

[5] J. Yang, M. Kuwabara, T. Asano, A. Chuma, and J. Du, Effect of lime particle size on melting behavior of lime-containing flux, *ISIJ Int.*, 47(2007), No. 10, p. 1401.

[6] R.H. Zhang, J. Yang, S.W. Wu, H. Sun, and W.K. Yang, Comparison of the prediction of BOF end-point phosphorus content among machine learning models and metallurgical mechanism model, *Steel Res. Int.*, 94(2023), No. 5, art. No. 2200682.

[7] Z.C. Xin, J.S. Zhang, Y. Jin, J. Zheng, and Q. Liu, Predicting the alloying element yield in a ladle furnace using principal component analysis and deep neural network, *Int. J. Miner. Metall. Mater.*, 30(2023), No. 2, p. 335.

[8] K. Feng, A.J. Xu, D.F. He, and L.Z. Yang, Case-based reasoning method based on mechanistic model correction for predicting endpoint sulphur content of molten iron in KR desulphurization, *Ironmaking Steelmaking*, 47(2020), p. 799.

[9] L. Qi, H. Liu, Q. Xiong, and Z.X. Chen, Just-in-time-learning based prediction model of BOF endpoint carbon content and temperature via vMF mixture model and weighted extreme learning machine, *Comput. Chem. Eng.*, 154(2021), art. No. 107488.

[10] Z.L. Wang, Y.P. Bao, and C. Gu, Convolutional neural network-based method for predicting oxygen content at the end point of converter, *Steel Res. Int.*, 94(2023), No. 1, art. No. 2200342.

[11] S.L. Jiang, X.Y. Shen, and Z. Zheng, Gaussian process-based hybrid model for predicting oxygen consumption in the converter steelmaking process, *Processes*, 7(2019), No. 6, art. No. 352.

[12] L.S. Carlsson, P.B. Samuelsson, and P.G. Jönsson, Interpretable machine learning—Tools to interpret the predictions of a machine learning model predicting the electrical energy consumption of an electric arc furnace, *Steel Res. Int.*, 91(2020), No. 11, art. No. 2000053.

[13] G.S. Wei, R. Zhu, L.Z. Yang, and T.P. Tang, Hybrid modeling for endpoint carbon content prediction in EAF steelmaking, [in] *Materials Processing Fundamentals 2018*, Springer International Publishing, Switzerland, 2018, p. 211.

[14] L.Z. Yang, B. Li, Y.F. Guo, S. Wang, B.T. Xue, and S.Y. Hu, Influence factor analysis and prediction model of end-point carbon content based on artificial neural network in electric arc furnace steelmaking process, *Coatings*, 12(2022), No. 10, art. No. 1508.

[15] Q.D. 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.*, 83(2019), art. No. 105675.

[16] Y.P. Bao, X. Li, and M. Wang, A novel method for endpoint temperature prediction in RH, *Ironmaking Steelmaking*, 46(2019), No. 4, p. 343.

[17] X.J. Wang, M.S. You, Z.Z. Mao, and P. Yuan, Tree-structure ensemble general regression neural networks applied to predict the molten steel temperature in ladle furnace, *Adv. Eng. Inform.*, 30(2016), No. 3, p. 368.

[18] Y.H. Liu, H.B. Lu, H.Q. Zhang, X. Wu, Y.B. Zhong, and Z.S. Lei, Quality prediction of continuous casting slabs based on weighted extreme learning machine, *IEEE Access*, 10(2022), p. 78231.

[19] D. Cemernek, S. Cemernek, H. Gursch, *et al.*, Machine learning in continuous casting of steel: A state-of-the-art survey, *J. Intell. Manuf.*, 33(2022), No. 6, p. 1561.

[20] Z. Chen, J.G. Wang, G.Q. Zhao, Y. Yao, and C. Xu, Endpoint temperature prediction of molten steel in VD furnace based on AdaBoost.RT-ELM, [in] *2020 IEEE 9th Data Driven Control and Learning Systems Conference (DDCLS)*, Liuzhou, 2020, p. 789.

[21] K. Feng, A.J. Xu, D.F. He, and H.B. Wang, An improved CBR model based on mechanistic model similarity for predicting end phosphorus content in dephosphorization converter, *Steel Res. Int.*, 89(2018), No. 6, art. No. 1800063.

[22] S. Pal and C. Halder, Optimization of phosphorous in steel produced by basic oxygen steel making process using multi-objective evolutionary and genetic algorithms, *Steel Res. Int.*, 88(2017), No. 3, art. No. 1600193.

[23] F. He and L.Y. Zhang, Prediction model of end-point phosphorus content in BOF steelmaking process based on PCA and BP neural network, *J. Process. Contr.*, 66(2018), p. 51.

[24] H.B. Wang, J. Cai, and K. Feng, Predicting the endpoint phosphorus content of molten steel in BOF by two-stage hybrid method, *J. Iron Steel Res. Int.*, 21(2014), p. 65.

[25] Z. Liu, S.S. Cheng, and P.B. Liu, Prediction model of BOF endpoint P and O contents based on PCA–GA–BP neural network, *High Temp. Mater. Process.*, 41(2022), No. 1, p. 505.

[26] K.X. Zhou, W.H. Lin, J.K. Sun, *et al.*, Prediction model of endpoint phosphorus content for BOF based on monotone-constrained BP neural network, *J. Iron Steel Res. Int.*, 29(2022), No. 5, p. 751.

[27] S.M. Acosta, A.L. Amoroso, Â.M.O. Sant'Anna, and O.C. Junior, Predictive modeling in a steelmaking process using optim-

ized relevance vector regression and support vector regression, *Ann. Oper. Res.*, 316(2022), No. 2, p. 905.

[28] R. Zhang and J. Yang, State of the art in applications of machine learning in steelmaking process modeling, *Int. J. Miner. Metall. Mater.*, 30(2023), No. 11, p. 2055.

[29] S.W. Wu, J. Yang, R.H. Zhang, and H. Ono, Prediction of endpoint sulfur content in KR desulfurization based on the hybrid algorithm combining artificial neural network with SAPSO, *IEEE Access*, 8(2020), p. 33778.

[30] G.B. Huang, Q.Y. Zhu, and C.K. Siew, Extreme learning machine: Theory and applications, *Neurocomputing*, 70(2006), No. 1-3, p. 489.

[31] G.B. Huang, Q.Y. Zhu, and C.K. Siew, Extreme learning machine: A new learning scheme of feedforward neural networks, [in] *2004 IEEE International Joint Conference on Neural Networks*, Budapest, 2005, p. 985.

[32] N.Y. Liang, G.B. Huang, P. Saratchandran, and N. Sundararajan, A fast and accurate online sequential learning algorithm for feedforward networks, *IEEE Trans. Neural Netw.*, 17(2006), No. 6, p. 1411.

[33] J.W. Zhao, Z.H. Wang, and D.S. Park, Online sequential extreme learning machine with forgetting mechanism, *Neurocomputing*, 87(2012), p. 79.

[34] F. Pedregosa, G. Varoquaux, A. Gramfort, *et al*., Scikit-learn: Machine learning in python, *J. Mach. Learn. Res.*, 12(2011), p. 2825.

[35] L. Ferrado, Pyoselm. A Python implementation of Online Sequential Extreme Machine Learning (OS-ELM) for Online Machine Learning, 2021 [2023–03–02]. https://github.com/leferrad/pyoselm

[36] L.J. Feng, C.H. Zhao, Y.L. Li, M. Zhou, H.L. Qiao, and C. Fu, Multichannel diffusion graph convolutional network for the prediction of endpoint composition in the converter steelmaking process, *IEEE Trans. Instrum. Meas.*, 70(2021), art. No. 3000413.

[37] M. Iwasaki and M. Matsuo, Change and development of steelmaking technology, *Nippon Steel Tech. Rep.*, 391(2011), p. 88.

[38] H. Sun, J. Yang, W.K. Yang, and R.H. Zhang, Comprehensive evaluation of phosphorus enrichment capacity for decarburization slag at different temperatures based on industrial experiments, mineral phase analysis and ion–molecule coexistence theory, *Metall. Mater. Trans. B*, 54(2023), No. 1, p. 115.