Machine Learning Modeling of Gas Utilization Rate in Blast Furnace

DEWEN JIANG,[1] ZHENYANG WANG [1,5] JIANLIANG ZHANG,[1,2]
DEJUN JIANG,[3] KEJIANG LI,[1] and FULONG LIU[4]

1.—School of Metallurgical and Ecological Engineering, University of Science and Technology
Beijing, Beijing 100083, China. 2.—School of Chemical Engineering, The University of
Queensland, St Lucia, QLD 4072, Australia. 3.—Dadu River Hydropower Development Co.,
Ltd., Chengdu 610000, China. 4.—He Steel Group Co. Ltd., Shijiazhuang 050024, China.
5.—e-mail: wangzhenyang@ustb.edu.cn

In the field of ironmaking, it is crucial to predict the gas utilization rate
(GUR), which is strongly related to the product quality and the smooth
operation of a blast furnace (BF). The present work proposes a model based on
the multi-layer perceptron (MLP) algorithm and data-driven to predict GUR
after 1 h, 2 h, and 3 h, respectively. First, the collected data are preprocessed
using the 3σ criterion. A maximal information coefficient (MIC) is then used
for feature selection. Meanwhile, a grid search is used to select best hyper-
parameters for the MLP and an extreme learning machine (ELM) algorithm.
Based on the above steps, the MLP and ELM models are built separately.
Finally, the prediction performance of the two models is compared in several
dimensions. The results show that the prediction result of both models is
excellent when the chosen output parameter is the GUR after 1 h. In addition,
when different output parameters are selected, the prediction accuracy of the
MLP model is always higher than that of the ELM model.

INTRODUCTION

The aim of blast furnace (BF) ironmaking is to
product high-quality iron with low production costs
and high energy efficiency,[1] where the BF is a major
equipment for converting iron ore into hot metal.
The quality of the hot metal and other products is
directly influenced by the smelting condition of the
BF.[2,3] The hot blast is blown into a BF from the
bottom to form an upward gas flow in BF ironmak-
ing.[4] At the same time, the iron ore reacts with the
gas flow to yield molten iron, BF slag, and gas. BF
gas is an important product and energy substance in
the ironmaking process, and it can promote the
internal reaction of the BF and provide heat for the
BF reaction. At the same time, the waste heat of BF
gas can be used in the air heater and coke oven. The
gas utilization rate (GUR) is defined as the ratio of
the carbon dioxide content to the total content of
carbon monoxide and carbon dioxide in the top gas
flow, and it reflects the efficiency of energy utiliza-
tion and the distribution of gas flow in the BF.[5]
Improving GUR is of great importance to improve
the energy efficiency and to reduce smelting costs.[6]
Thus, it is extremely important to build a model
that accurately predicts GUR.

In the past few decades, data-driven and first-
principle methods have been two key methods for
predicting the GUR or other important parameters
of a BF.[7,8] The first-principle model in metallurgy is
built based on metallurgical transport, metallurgi-
cal dynamics, and thermodynamics. Therefore, an
accurate and basic theory of metallurgy is needed to
build a reliable model.[9] However, it is difficult to
master the condition of a BF due to the intricate
transport phenomena and extreme environ-
ments.[2,10] In contrast, the data-driven model is
built on the basis of the underlying principles of
huge amounts of data and statistical theory. Mean-
while, with the increased power of computers, the
data-driven model is in the spotlight in BF iron-
making. For example, Bhattacharya used partial

least squares to predict the hot metal silicon content in a BF.[11] Tang tried predicting the silicon content of hot metal based a chaos particle swarm optimization.[12] The hot metal temperature and silicon content were forecast on the basis of partial least squares by Lin.[13] Nurkkala proposed multiple autoregressive models to predict the hot metal silicon content.[14] Gao predicted the thermal state change of a BF based on a support vector regression modeling method.[15] In order to improve the GUR and adjust burden distribution, Zhang constructed a decision-making strategy.[16] In addition, some scholars have tried to predict the GUR based on neural networks.[17,18]

In summary, data-driven models have yielded excellent results with low computational complexity compared to the first-principle model. In particular, extensive data-driven models have been developed in the past for BFs. In addition, many academics have focused solely on the prediction methods used in constructing a model, but have neglected to consider when gas utilization should be predicted using the current BF state parameters. In this study, we propose a data-driven model and a MLP algorithm to predict GUR after 1, 2, and 3 h in a BF, respectively. MLP is a parallel combination of many identical simple processing units, and, although the function of each unit is simple, the parallel activity of a large number of simple units makes it incredibly powerful and effective in processing information. In addition, MLP can be trained to acquire the weights and structure of the network, presenting a strong self-learning ability and adaptive capability to the environment. These features of MLP are well suited for use in the complex and extreme environments of a BF. The predicted results have demonstrated that the constructed MLP model performs better than the ELM model. The use of this method to predict GUR has not previously been reported. The remaining parts of this paper are structured as follows.

"Data pre-processing and feature selection" section gives a detailed description of the method of data pre-processing and feature selection, then, a modeling method is presented in "Construction of the model" section. In "Analysis and comparison using actual run data" section, the comparison of the predictions of the two models are given, while "Conclusion" conclusion sets out the conclusions.

DATA PRE-PROCESSING AND FEATURE SELECTION

Data Pre-Processing

Real production data including 35,198 continuous samples were collected from a 4150-m^3 BF. The interval time between the samples was 1 h. The collected parameters of each data sample are shown in Table I.

Due to the complex conditions, such as high temperature and pressure, multiphase fluid flows, and mass and heat transfer, there are missing values and outliers in the detection and collection of related parameters. Previous studies have shown that pre-processing of raw data followed by correlation modeling and prediction is more effective than using raw data directly for prediction modeling. Therefore, pre-processing of the data is essential and a 3σ criterion has been used here to judge the outliers and extreme outliers. Extreme outliers have been replaced with missing values. In order to ensure the continuity of time, extreme outliers and vacancy data are usually filled, and linear interpolation is selected to replace missing values.

Feature Selection

In order to improve the accuracy of the prediction, we have selected the parameters in the current state of the BF as inputs of the constructed models. Meanwhile, GUR after 1 h (GUR-1h), GUR after 2 h (GUR-2h), and GUR after 3 h (GUR-3h) have been selected as the output parameters.

As the BF ironmaking is a systematic process, the GUR is greatly influenced by many factors. Feature selection selects important influencing factors as input variables in order to reduce the complexity of the model during computation, and to improve the accuracy of the prediction of the model. A maximal information coefficient (MIC) was used for feature selection and to measure the correlation between the feature parameters. The calculation process of MIC is divided into the following three steps[19]:

(1) The scatter plot composed of X and Y is gridded with m columns and n rows, and the value of MIC is obtained for a given m and n.
(2) Normalization of the largest mutual information value.
(3) Selecting the maximum value of mutual information at different scales as the MIC value.

Based on the above steps, the MIC values between the characteristic parameters and the output parameters are shown in Fig. 1.

Finally, the characteristic parameters with MIC values greater than 0.15 are selected as input variables based on the above calculation results. Therefore, 16 input parameters have been finally selected and are shown in Table II.

CONSTRUCTION OF THE MODEL

MLP Algorithm

Multi-layer perceptron (MLP) is also known as artificial neural networks. MLP takes the eigenmatrix x_m and obtains a predictor variable $\tilde{y}_m$ by combining linear and nonlinear combinations for the sample set $D = \{(x_m, y_m)\}$. In addition to the input and output layers, there are multiple hidden layers within the multilayer perceptron. The simplest MLP requires only an input layer, a single hidden layer, and an output layer, which at this

Table I. The related parameters involved in the research

| Actual physical meaning	Parameter name	Actual physical meaning	Parameter name
Blast volume	BV	Wind pressure	WP
Pop pressure	TP	Pressure differential	PD
Air permeability resistance coefficient	ARC	Wind temperature	WT
Oxygen content	OC	wind velocity	WV
Coal injection quantity	CIQ	Ventilating index	VI
Theoretical combustion temperature	TCT	Center strength	Z
Edge strength	W	Intensity ratio (IR)	Z/W
Temperature of cross-temperature measuring	TOCTM	Marginal mean	MM
Content of carbon monoxide	COCM	Content of carbon dioxide	COCD
Bosh gas index	BGI	Kinetic energy	KE
Inlet water temperature	IWT	Outlet water temperature	OWT
Gas utilization rate	GUR		
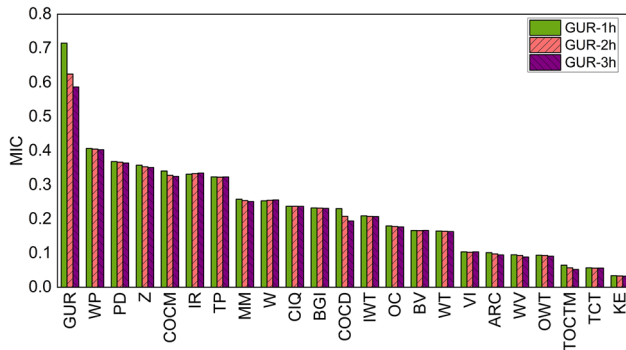
Fig. 1. The MIC of the selected feature parameters with GUR (the definitions of the abbreviations for each feature parameter can be found in Table I).

Table II. Final characteristic parameters of GUR

BV	IWT	Z	BGI
TP	CIQ	GUR	OC
WP	W	MM	IR
WT	COCD	COCM	PD

Definitions of the abbreviations for each feature parameter can be found in Table I.

point is also referred to as a single-layer feed-forward neural network. For a dataset (x_i, t_i) containing N samples, the mathematical model of the single-layer feed-forward neural network is shown in Eq. 1:

$$\sum_{i=1}^{\tilde{N}} \beta_i g_i(x_k) = \sum_{i=1}^{\tilde{N}} \beta_i g(w_i, b_i, x_k) = t_k, k = 1, 2, \ldots, N \tag{1}$$

where $x_i = [x_{i1}, x_{i2}, \ldots, x_{in}]$ which denotes the N-dimensional features of the ith sample, $t_i = [t_{i1}, t_{i2}, \ldots, t_{in}]$ and that the corresponding target vector β_i is the output weight matrix for connecting the ith hidden and output nodes, $g(w_i, b_i, x_k)$ is a nonlinear segmented continuous function, and w_i and b_i are determined model parameters. Each neuron in the hidden layer is composed of a linear combination of input features x. However, if it is just a linear combination, the result will be linearly related to the feature no matter how many layers are in this neural network. After each neuron results in a linear operation, an activation function is added that changes the linearity rule to handle this situation.

ELM Algorithm

ELM is a machine-learning method for single-layer feed-forward neural network, which has three layers of neurons, namely, the input, hidden, and output. For a dataset (x_i, t_i) containing N samples, the mathematical model of the single-layer feed-forward neural network is shown in Eq. 1. It is worth noting that $g(w_i, b_i, x_k)$ is a nonlinear segmented continuous function, and that w_i and b_i are randomly determined model parameters. Thus, Eq. 1 can be written in the form of an implicit output matrix, as shown in Eqs. 2 and 3:

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

$$H(W,B,X) = \begin{pmatrix} g(w_1 \cdot x_1 + b_1) & \ldots & g(w_{\tilde{N}} \cdot x_1 + b_{\tilde{N}}) \\ \vdots & \ddots & \vdots \\ g(w_1 \cdot x_N + b_1) & \ldots & g(w_{\tilde{N}} \cdot x_N + b_{\tilde{N}}) \end{pmatrix}_{N \times \tilde{N}} \tag{3}$$

According to the least squares method, combined with the singular value decomposition method, the solution of the ELM can be expressed as Eq. 4:

$$\overset{\wedge}{\beta} = (H^T H)^{-1} H^T T \tag{4}$$

The raw data is pre-processed and all data are normalized. The MLP and ELM algorithms were used for modeling. The dataset used in this paper is from 35,198 sets of data collected by the online detection system of BFs in China. In addition, 75% of the dataset is used for the training model and the remaining data are used for the testing model.

ANALYSIS AND COMPARISON USING ACTUAL RUN DATA

Comparison of the Predictions of the Two Models

In order to achieve accurate prediction, the ELM and MLP models were established. A total of 26,398 samples were used for training the two models. Meanwhile, the remaining 8800 samples were used for testing the two models. The best hyperparameters for the two models were obtained by a grid search. When the output parameter of the two models is the GUR-1h, the first hidden layer of the ELM model has 140 nodes and the activation function is a linear function, while the second hidden layer has 40 nodes and the activation function is a tanh function. In addition, the MLP model has 50 hidden layers and the activation function is the tanh function.

With the predicted parameter of GUR-1h, the results are expressed in Fig. 2a and b. In order to show the comparison results clearly, 100 data points were randomly selected for comparison. Figure 2 indicates that the fit of both models is excellent and that the fitting of the predicted and measured values of MLP is better than those of ELM.

Figure 3a and b provides the scatter plots of the two methods in terms of the predicted and measured GUR. It can be seen that ELM and MLP fit the data very well, since the data are distributed compactly along the diagonal direction. In comparison, MLP fitted the data better than the ELM method, since the data are distributed more closely along the diagonal direction, as shown in Fig. 3a and b. In addition, the probability density of the predicted error of the MLP model is much smaller than that of the ELM method, as shown in Fig. 3c and d.

When the output parameter of the two models is the GUR-2h, the first hidden layer of the ELM model has 140 nodes and the activation function is a linear function, while the second hidden layer has 10 nodes and the activation function is the tanh function. In addition, the MLP model has 250 hidden layers and the activation function is the relu function. For the predicted parameter of GUR-2h, the results are expressed in Fig. 4a and b.

Both Figs. 4 and 5 demonstrate that the MLP model is more accurate for predicting the GUR-2h than the ELM model.

When the output parameter of the two models is the GUR-3h, the first hidden layer of the ELM model has 220 nodes and the activation function is a linear function, while the second hidden layer has 80 nodes and the activation function is a tanh function. In addition, the MLP model has 50 hidden layers and the activation function is the tanh function. For the predicted parameter of GUR-3h, the results are expressed in Fig. 6a and b

A full comparison of the predicted results of the two methods is shown in Fig. 7a and b, where the predicted GUR values were plotted against the observed GUR values. As shown in the figure, the data are distributed more closely along the diagonal direction, indicating that the MLP model fits the data better than the ELM model. The same results can also be observed in Fig. 7c and d. Compared with the ELM model, the predicted GUR values of the MLP model are in better agreement with the actual observed GUR values.

All of the above analysis results demonstrate that the proposed MLP model is more accurate for GUR prediction compared with the ELM model.

Evaluation of Forecast Results

In order to characterize the accuracy of the two models in predicting GUR from several aspects, the root mean squared error (RMSE) and hit rate (HR) were used. RMSE and HR are defined as follows:

$$\text{RMSE} = \sqrt{\frac{1}{n} \cdot \sum_{i=1}^{n} (h(x_i) - y_i)^2} \tag{5}$$
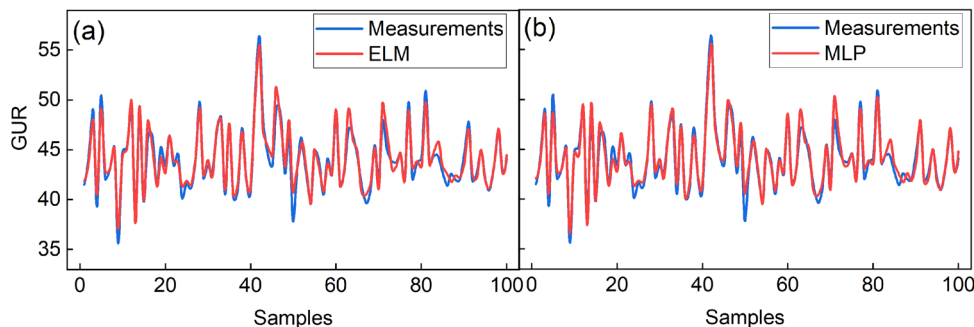

Fig. 2. Comparison of the predictions of the ELM (a) and MLP (b) model when the output parameter is GUR-1h
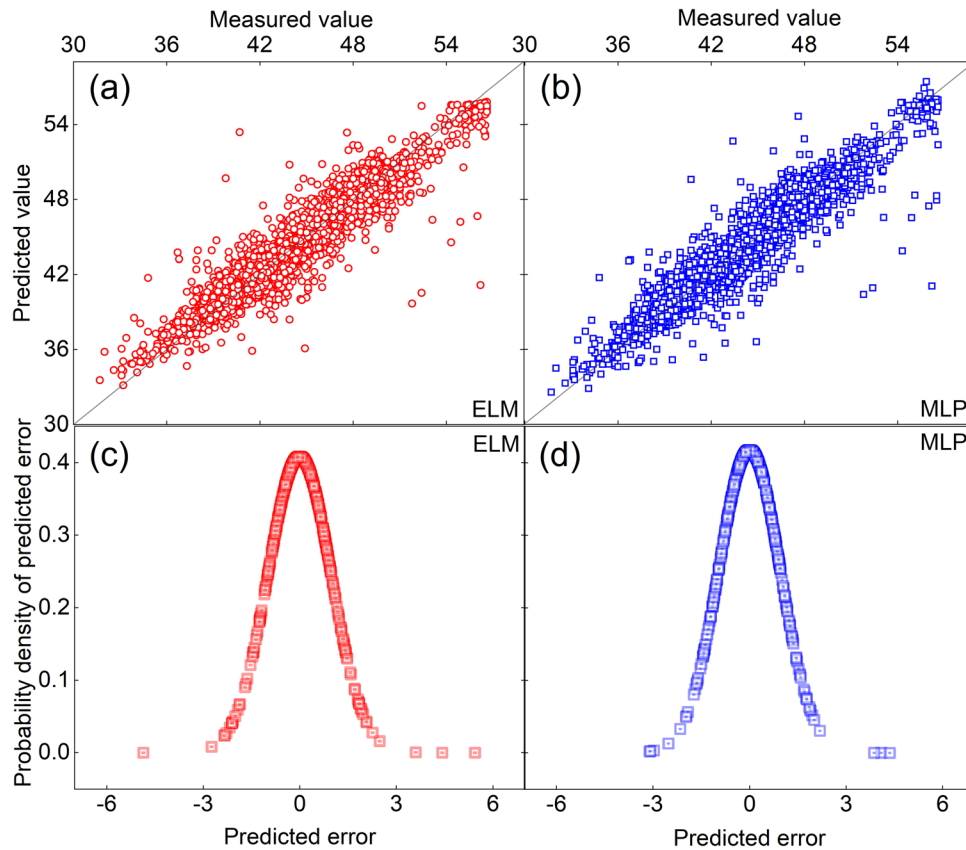
Fig. 3. (a–d) Comparison of the prediction errors of the two models when the output parameter is GUR-1h
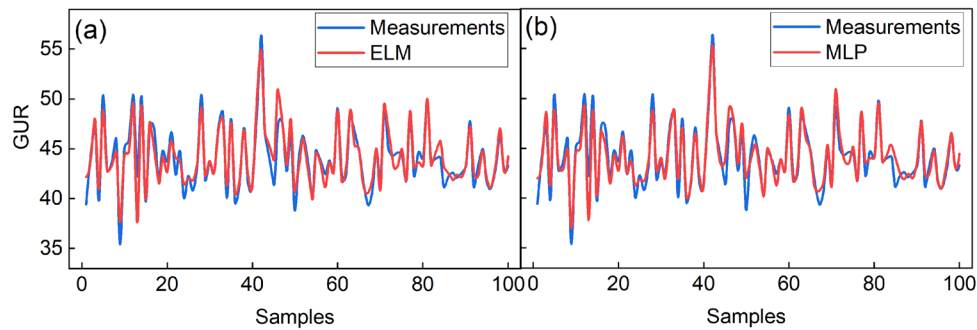

Fig. 4. Comparison of the predictions of the ELM (a) and MLP (b) model when the output parameter is GUR-2h

$$
\begin{cases}
\mathrm{HR} = \dfrac{1}{n} \cdot \displaystyle\sum_{i=1}^{n} \mathrm{HR}_i \times 100\% \\[2mm]
\mathrm{HR}_i = \begin{cases} 1, & |h(x_i) - y_i| \le c \\ 0, & |h(x_i) - y_i| > c \end{cases}
\end{cases}
\tag{6}
$$

where n denotes the total number of samples in the test set. y_i the measured value, $h(x_i)$ the predicted value, while c is the boundary value of the hit rate. In this paper, the value of c was selected as 2%.

The MLP model obtained the highest prediction accuracy between the two approaches, as shown in Table III. When GUR-1h, GUR-2h, and GUR-3h are predicted by the two models, the RMSE values of the MLP model are 0.022, 0.063, and 0.071 lower than the RMSE values of the ELM model, respectively. Consequently, compared with the ELM method, the predicted GUR values of the MLP model are in better agreement with the actual observed GUR values. All of the above analysis results demonstrate that the proposed MLP model
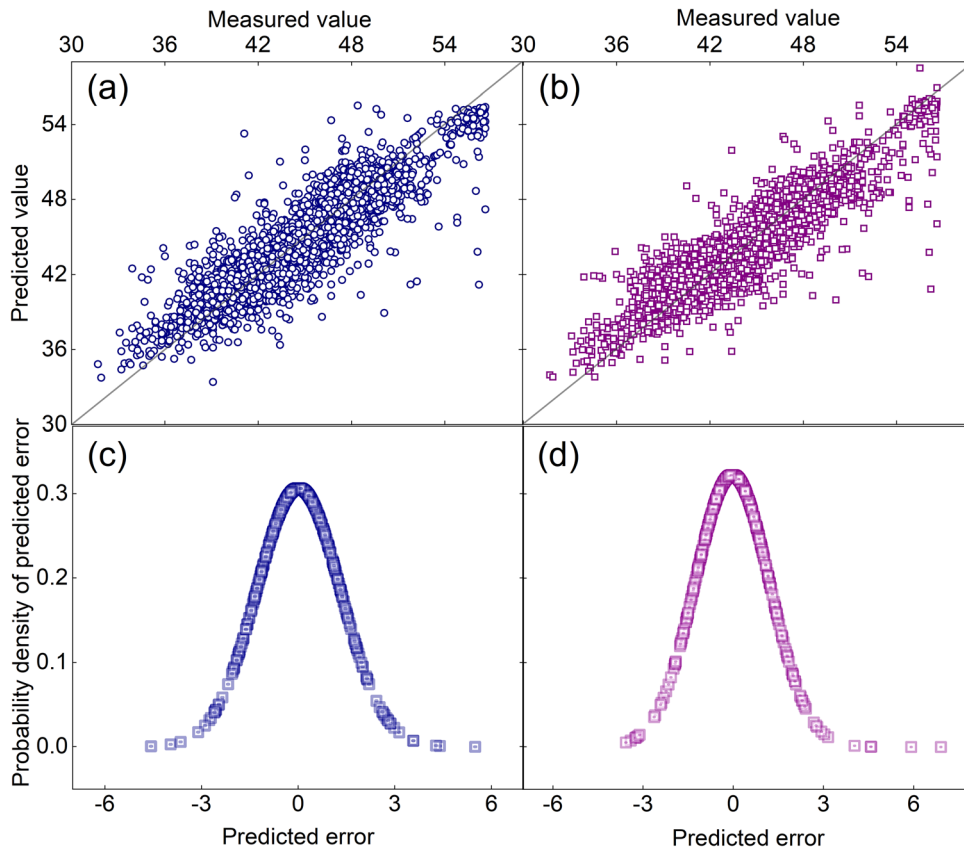
Fig. 5. (a–d) Comparison of the prediction errors of the two models when the output parameter is GUR-2h
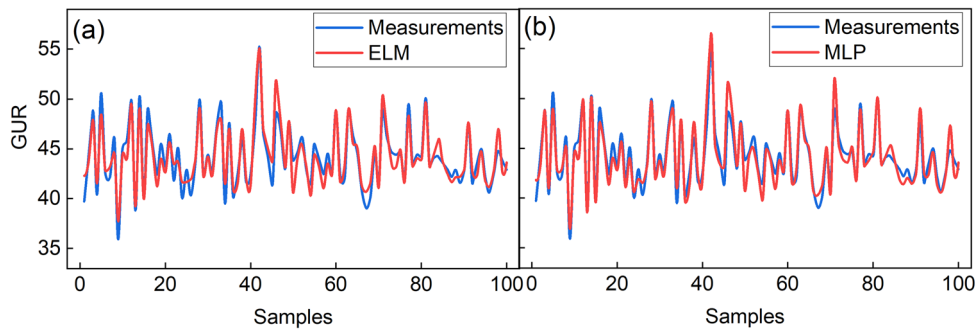

Fig. 6. Comparison of the predictions of the ELM (a) and MLP (b) model when the output parameter is GUR-3h

is more accurate for GUR prediction compared with the ELM method. Furthermore, ELM and MLP are highly suitable for predicting the GUR after 1 h, while ELM and MLP proved to be highly unsuitable for predicting GUR after 3 h. Finally, we chose to predict the gas utilization after 1, 2, and 3 h, and there was a decreasing trend in the accuracy of the predictions. It is worth noting that Yu analyzed the residence time of the gas in a BF through numerical simulation, and that his study showed that the mean residence time and space time of gas fluids were predicted as 13.5 s and 16.3 s, respectively.[20] This phenomenon suggests that the impact of the time that the gas is retained in a BF does not exceed 1 h.
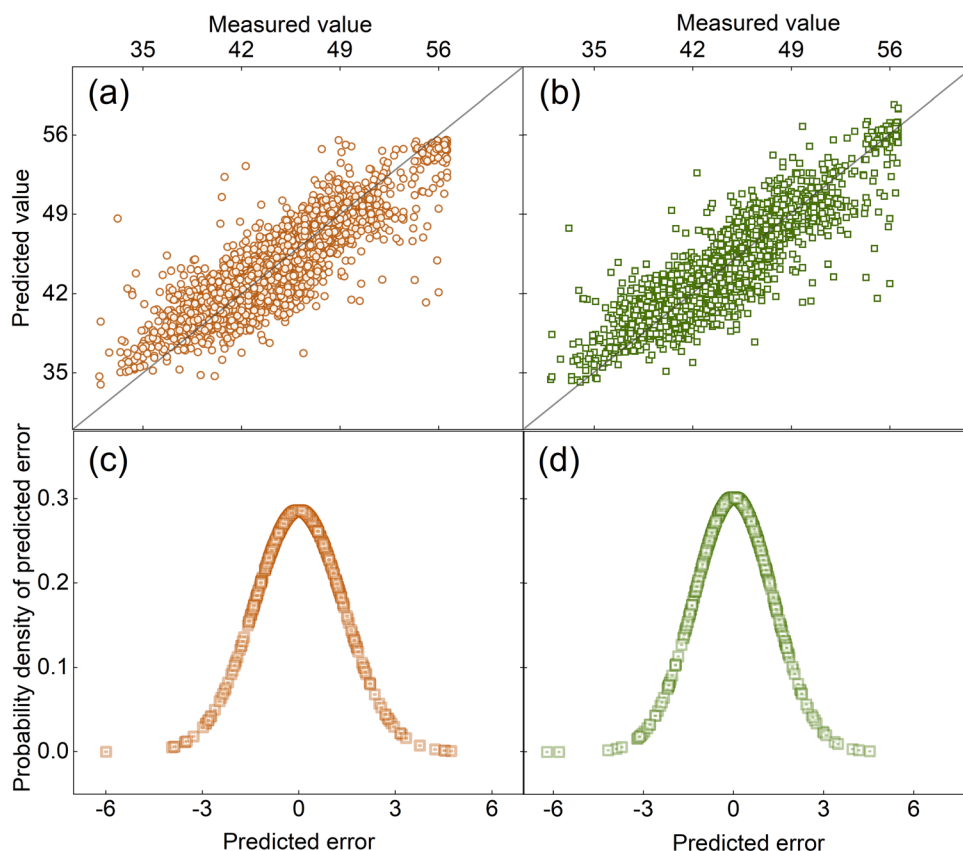
Fig. 7. (a–d) Comparison of the prediction errors of the two models when the output parameter is GUR-3h

Table III. Comparison of evaluation indicators of the two models

Output parameter	RMSE (MLP)	RMSE (ELM)	HR (MLP)	HR (ELM)
GUR-1h	0.955	0.977	0.964	0.961
GUR-2h	1.237	1.300	0.921	0.911
GUR-3h	1.321	1.392	0.902	0.893

CONCLUSION

The gas utilization rate of a BF is an important indicator of its energy consumption and smooth operation. Constructing a prediction model for the gas utilization rate is an important step in achieving highly efficient production. In this paper, two data-driven models are proposed for predicting gas utilization rate in a BF-based multi-layer perceptron and an extreme learning machine algorithm, which accurately predicted the gas utilization rates after 1, 2, and 3 h, according to the real production data for a BF. The simulation results show that both the multi-layer perceptron model and the extreme learning machine model achieved accurate predictions for BF gas utilization rate after 1 h based on the collected data. When the predicted parameter is the gas utilization rate after 1 h, the prediction accuracy of the multi-layer perceptron model reaches 96.4%. In addition, the multi-layer perceptron model is more accurate for the prediction of the BF gas utilization rate than the extreme learning machine model, while the predicted parameters are the gas utilization rates after 1, 2, and 3 h. Overall, a better prediction of the gas utilization rate after 1 h can be achieved using the multi-layer perceptron model based on the data and modeling approach used. More BF production data for different volumes and different production states will be used to verify these conclusions the future work.

CONFLICT OF INTEREST

There is no conflict of interest.

REFERENCES

1. M. Geerdes, R. Chaigneau, and I. Kurunov, *Modern Blast Furnace Ironmaking: An Introduction*, 3rd edn (IOS Press, Amsterdam, 2015).
2. Z. Ge, Z. Song, S.X. Ding, and B. Huang, *IEEE Access* 5, 20590. (2017).
3. P. Zhou, H. Song, H. Wang, and T. Chai, *IEEE Trans. Contr. Sys. Technol.* 25, 1761. (2017).
4. L. Jian, C. Gao, and Z. Xia, *IEEE Trans. Automat. Sci. Eng.* 9, 763. (2012).
5. L. Zhang, C. Hua, J. Li, and X. Guan, *IEEE Trans. Contr. Sys. Technol.* 25, 262. (2017).
6. Q. An, L. Shen, M. Wu, and H. She, *Contr. Eng. Pract.* 92, 104120. (2019).
7. Q. Ge, *J. Process Contr.* 65, 107. (2018).
8. W. Liu, Q. Ge, J. Chen, and H. Song, *Contr. Eng. Pract.* 72, 19. (2018).
9. M. Zhang, M. Kano, and Y. Li, *Comput. Chem. Eng.* 104, 164. (2017).
10. B. Zhou, H. Ye, F. Zhang, and L. Li, *Contr. Eng. Pract.* 47, 1. (2016).
11. T. Bhattacharya, *ISIJ Int.* 45, 1943. (2005).
12. L. Tang, L. Zhuang, and J. Jiang, *Exp. Sys. Appl.* 36, 11853. (2009).
13. L. Shi, L. Li, T. Yu, and P. Li, *J. Iron Steel Res. Int.* 18, 13. (2011).
14. A. Nurkkala, F. Pettersson, and H. Saxen, *ISIJ Int.* 52, 1764. (2012).
15. C. Gao, L. Jian, and S. Luo, *IEEE Trans. Ind. Electron.* 59, 1134. (2012).
16. M. Wu, X. Zhang, Q. An, H. She, and Z. Liu, *Contr. Eng. Pract.* 78, 186. (2018).
17. D. Xiao, C. Yang, and Z. Song, *J Zhejiang Univ.* 46, 2103. (2012).
18. J. Zhang, Q. An, M. Wu, and W. Cao. In *Proceedings of the 5th international workshop on advanced computational intelligence and intelligent informatics*, 2017, pp. 2-5.
19. D.N. Reshef, Y.A. Reshef, H.K. Finucane, S.R. Grossman, G. McVean, P.J. Turnbaugh, E.S. Lander, M. Mitzenmacher, and P.C. Sabeti, *Science* 334, 1518. (2011).
20. X. Yu, and Y. Shen, *Chem. Eng. Sci.* 199, 50. (2019).