Multi-Gate Mixture-of-Experts Stacked Autoencoders for Quality Prediction in Blast Furnace Ironmaking

Hongyu Zhu, Bocun He, and Xinmin Zhang*

ABSTRACT: The blast furnace is an energy-intensive and extremely complex reactor in the ironmaking process. To reduce energy consumption, improve product quality, and ensure the stability of blast furnace operation, it is very important to predict the quality indicators of molten iron accurately and in real time. However, most of the existing product quality prediction models, such as the stacked autoencoder (SAE) model, use a single-channel stack structure. For such models, when the working conditions of the blast furnace ironmaking process change, a large prediction error will occur. To solve this issue, this paper develops a novel deep learning model, called the multi-gate mixture-of-experts stacked autoencoder (MMoE-SAE), for predicting the quality variable in the blast furnace ironmaking processes. The proposed MMoE-SAE model is constructed based on a multi-gate hybrid expert structure, in which a series of SAE networks are selected as experts. The MMoE-SAE model inherits the advantages of MMoE and SAE, which can not only extract the deep features of the data but also have better adaptability to the changes of working conditions in the blast furnace ironmaking process. To verify the effectiveness and practicability of the proposed MMoE-SAE model, it was applied to predict the silicon content of molten iron in the blast furnace ironmaking process. The experimental results demonstrate that the proposed MMoE-SAE model outperforms other prediction models in prediction accuracy.

1. INTRODUCTION

The steel industry is one of the most important basic industries in the world, in which a blast furnace is a key reactor of ironmaking. Figure 1 shows the schematic of a typical blast furnace. The furnace body is generally divided into five operation areas: throat, stack, belly, bosh, and hearth from the top to the bottom of the blast furnace. When the blast furnace ironmaking system is running, the solid raw material composed of iron ore and coke is loaded in layers from the top according to a certain amount. In the meantime, pulverized coal and compressed hot air are sprayed at the lower part through the tuyere. The hot air in the furnace reacts with the carbon in the form of coke, producing carbon monoxide and a lot of heat. Carbon monoxide is considered a reducing agent for iron ore, reacting with iron oxide to generate carbon dioxide and molten iron. The produced molten iron is collected in the hearth and then cast through the tap hole for subsequent steelmaking processes.1−3

For blast furnaces, the goal is to realize the smooth operation of ironmaking and to produce high-quality molten iron with lower production costs.1−7 To realize this goal, it is necessary to monitor and control the blast furnace ironmaking process in real time. Usually, the control of the blast furnace system means controlling the final molten iron quality (MIQ). The generally used MIQ indicators are the silicon content, sulfur content, phosphorus content, and molten iron temperature. Silicon content is an important indicator to reflect the chemical heat of molten iron. In practice, blast furnace ironmaking expects low silicon content in the molten iron but avoids the risk of cooling the hearth that may lead to a chilled hearth.7 However, due to the harsh smelting environment, it is difficult to directly measure silicon content using existing conventional sensors. Therefore, it is imperative to investigate the prediction modeling of the silicon content in the blast furnace ironmaking process.

The prediction models for blast furnace ironmaking mainly include mechanism-based models and data-driven models. Mechanism-based models are generally based on physical and chemical principles, kinetics, and thermodynamics.8 However, the blast furnace smelting environment is extremely harsh, with high pressure and high temperature. Furthermore, the operation of blast furnace systems is complicated due to...
multiphase flow, multiphase coupling, and complex physicochemical reactions. These challenges make it difficult to obtain first-principles models that describe complex phenomena and quality metrics within the furnace. Therefore, in order to address this complex quality modeling problem, there is a need to study data-driven models. In the past few years, a large number of data-driven prediction algorithms and models have been developed and used. Compared to the mechanism-based models, data-driven models do not require detailed physical and chemical principles and are simply developed from the collected data. For example, Östermark and Saxén proposed a vector autoregressive moving average model to predict the molten iron silicon content and temperature of pig iron.9 Fontes et al. proposed a fuzzy c-means exogenous nonlinear autoregressive (FCMNARX) model to predict the molten iron temperature and silicon content. The FCMNARX model enables predictive models to cluster datasets based on optimization criteria.10 Shi et al. used partial least-square (PLS) regression and principal component analysis to predict the molten iron silicon content.11 Zhang et al. used a fast locally weighted partial least-squares (FLW-PLS)12 model to predict the molten iron silicon content and phosphorus content. FLW-PLS can process large-scale process data more efficiently and can save significant computational effort. In addition, a series of support vector machine models and their variants have been developed for the prediction of molten iron quality. For example, a least square support vector machine,13 binary coding support vector machine,14 multiple kernel support vector machine,15 and just-in-time-learning recursive multi-output least squares support vector regression16 have been proposed. Furthermore, neural networks17 and extreme learning machines18 have also been adopted to predict the molten iron quality. Moreover, the multi-input/multi-output models have also been widely used for molten iron quality prediction. For example, Li et al. proposed a multi-input/multi-output Takagi-Sugeno (MIMO-TS)19 fuzzy model for the prediction of molten iron quality in the blast furnace. Although these models have played an important role in improving blast furnace modeling techniques and predicting molten iron quality indicators because blast furnace ironmaking itself is a dynamic process, changes in working conditions during the process often reduce the prediction effect of prediction models.

In order to make the prediction model have a better ability to adjust to the changes of working conditions, this paper proposes a multi-gate mixture-of-experts stacked autoencoder (MMoE-SAE) model. The MMoE-SAE model is designed based on a multi-gate mixture-of-experts (MMoE) structure and a stacked autoencoder structure. Based on the traditional deep learning model, the MMoE structure can make the model adapt to changes in working conditions and adjust itself when the working conditions change. Therefore, for the dynamic blast furnace system, the MMoE-SAE model has better self-adaptation ability and quality variable prediction ability. In the application test using real industrial data, by comparing the MMoE-SAE model with the SAE model and bootstrapping aggregate stacked autoencoder (Bagging-SAE) model, the results show that the MMoE-SAE model has better working condition adjustment adaptability and comprehensive prediction effect in the blast furnace system. The main contributions of this work are summarized as follows.

- A novel deep learning model called MMoE-SAE is proposed and applied to predict the product quality of the blast furnace ironmaking process.
- The MMoE-SAE model is designed based on a MMoE structure, in which the SAE network is employed as an expert.
- By integrating a series of SAE networks into the MMoE structure, the constructed deep learning model has better adaptability to the working conditions of the blast furnace system.
- The proposed method was verified through real industrial blast furnace ironmaking data, and the application results demonstrate its superiority over other methods.

The remainder of this paper is organized as follows. In Section 2, the structure of the stacked autoencoders is described. Then, Section 3 gives the introduction of the MMoE structure. A detailed description of the proposed MMoE-SAE model is presented in Section 4. In addition, Section 5 gives comparative application results using real blast furnace data. Finally, conclusions are made in Section 6.

2. STACKED AUTOENCODERS

Deep learning has been widely used for various tasks due to its high flexibility in processing complex nonlinear data. SAEs are deep neural networks consisting of multiple autoencoders. The autoencoder (AE)20 is a type of neural network that learns hidden features of the data in an unsupervised manner. Figure 2 shows the diagram of an AE.

The main structure of an AE consists of three parts. The first part is called the input layer, and the function of the input layer is to receive the data input to the encoder. The second part is called the hidden layer, also known as the encoding layer. The function of the hidden layer is to extract the feature information in the data. The feature of the information extracted from the data is also called encoding. The dimension of encoding is determined by the number of nodes in the hidden layer. The third part is called the output layer. The function of the output layer is to reconstruct the information of the hidden layer into the original input data through the decoding process so as to minimize the reconstruction error of the reconstructed data. The process from the hidden layer to the output layer is the process of decoding the information through the decoder.
To describe the whole process in detail, we can assume that the input data of the AE can be expressed as \( x_D \), where \( D \) represents the dimension of the input data. We use \( h_H \) to represent the hidden layer from the encoder structure in an AE, and the calculation process can be described as follows:

\[
h = f_{\text{encoder}}(W_e x + b_e)
\]

(1)

where \( f_{\text{encoder}}(\cdot) \) represents the activation function of the encoder. \( W_e \) and \( b_e \) represent the weight and bias of the encoder, respectively. Let \( \hat{x} \in \mathbb{R}^D \) denote the reconstructed input from the decoder, which is computed as

\[
\hat{x} = f_{\text{decoder}}(W_d h + b_d)
\]

(2)

where \( f_{\text{decoder}}(\cdot) \) represents the activation function of the decoder. \( W_d \) and \( b_d \) represent the weight and bias of the decoder, respectively. The commonly used activation functions in an AE are the tanh function, the sigmoid function, and the rectified linear unit function. Let \( \theta = \{ W_e, W_d, b_e, b_d \} \) denote the parameter set of the AE. Given the training dataset \( X = \{ x_1, x_2, \ldots, x_N \} \), \( N \) denotes the number of training samples. The loss function of AE is denoted as

\[
J_{AE}(\theta) = \min \frac{1}{2N} \sum_{i=1}^{N} ||x_i - \hat{x}_i||^2
\]

(3)

The backpropagation algorithm is generally used to train an AE.

In practice, to learn more powerful feature representations, SAEs are usually constructed, which are obtained by stacking multiple autoencoders. Figure 3 shows the basic structure of an SAE. In an SAE, an AE is trained using the output of the last encoder, and when the training process is finished, the encoder part of the trained AE is stacked to the last encoder. This training process is also known as the unsupervised layer-wise pre-training process. Once a stack of AEs has been built, a series of weights of stacked AEs can be used in the initialization of the corresponding global learning. Then, the parameters of all layers can be fine-tuned using the stochastic gradient descent algorithm.\(^{21}\)

### 3. MULTI-GATE MIXTURE-OF-EXPERTS

Traditional machine learning methods generally only learn one task at a time. In the face of complex problems, the task is usually divided into multiple independent subtasks, and the results are integrated to get the final output. This kind of method is difficult to consider the correlation between subtasks. Multi-task learning can obtain the auxiliary information of other related subtasks while processing subtasks through parameter sharing so as to improve the application effect in multi-subtask scenarios. Traditional multi-task learning usually adopts the mechanism of parameter hard sharing, and all subtasks use the same feature extraction layer.
for feature extraction. Such methods ignore the differences between subtasks and limit the expression ability of the model.

To solve this problem, Google proposed an MMoE structure. Its core mechanism is different from the traditional hard sharing mechanism, MMoE divides the feature extraction layer into multiple experts according to certain rules instead of forcing all subtasks to share the same feature extraction layer. In addition, a gating mechanism is introduced to control the weight of each expert to each subtask. Each subtask has an independent and trainable gating network, which can make each subtask share relevant information and retain their differences through training. The structure of MMoE is shown in Figure 4.

Figure 4. Structure of MMoE.

For subtasks \( k \in \mathbb{Z} \), the output is

\[
y^k = f^k(f^k(x))
\]  

(4)

\[
f^k(x) = \sum_{i=1}^{n} g_{i,k}(x) f_{i}(x)
\]  

(5)

\[
g^k(x) = \text{softmax}(W_{g,k}(x))
\]  

(6)

where \( f^k(\cdot) \) is the \( k \)th tower network of subtasks, \( f_i(\cdot) \) is the output of the \( i \)th expert, and the shared layer of the network is composed of \( n \) experts. \( W_{g,k}(x) \) is the matrix that needs to be trained through the network, \( g_{i,k}(x) \) is the gating network output of the corresponding subtask, which converts the input features into weight vectors through a softmax layer. \( f^k(x) \) is a combination of an expert network and a gating network.

4. MMoE-SAE MODEL

The blast furnace system has the problem of multiple working conditions in the operation process, and the existing data-driven prediction models have poor ability to describe the changes of working conditions. The problem of multiple working conditions in the blast furnace operation can be regarded as the different task requirements in the description of system characteristics during the dynamic operation of the blast furnace. Therefore, the MMoE structure can be introduced to make the model better adapt to changes in working conditions. This section proposes an MMoE-SAE model that can better predict product quality variables under blast furnace dynamic conditions and introduces the application process of the model.

4.1. Structure of MMoE-SAE Model. The MMoE-SAE model integrates the SAE network into the MMoE structure and addresses the change in blast furnace process conditions. The structure diagram of the proposed model is shown in Figure 5. The model replaces the layer stacking substructure with a hybrid multi-expert structure, and each expert is designed as an SAE. Assuming that the total number of experts is \( R \), the input data \( x \) are sent to each expert, and the hidden layer in the \( r \)th \((r = 1, \ldots, R)\) expert can be expressed as

\[
h^r_i = f^i_r(W^r_{g,k}x + b^r_i)
\]  

(7)

\[
h^r_{l-1} = f^i_r(W^r_{h,k}h^r_{l-1} + b^r_i)
\]  

(8)

where \( L \) represents the number of hidden layers in the expert, \( f^i_r(\cdot) \) represents the function of the encoder between the \( l \)th hidden layer and the previous layer. \( W^r_{g,k} \) and \( b^r_i \) represent the encoder weight matrix and bias vector, respectively. The output information of the top hidden layer of all experts is \( h_l = [h_1,l, h_2,l, \ldots, h_{R,l}] \). After the output information passes through the refresher, the output value for each expert can be obtained. This process can be described as

\[
y^r_{e,r} = f_{reg,r}(W_{reg,r}h^r_{l-1} + b_{reg,r})
\]  

(9)

where \( y^r_{e,r} \) denotes the output of the \( r \)th expert, \( f_{reg,r}(\cdot) \) is the function of the regressor, and \( W_{reg,r} \) and \( b_{reg,r} \) are the regressor weight matrix and bias vector, respectively. So far, experts that describe the characteristics of the system under different working conditions have their corresponding predicted values. Then, a weight value that conforms to the expert’s contribution to the overall prediction effect needs to be provided. This process is mainly completed through the Gate Module in the model, which extracts information consisting of the output information of each expert \( h_l = [h_1,l, h_2,l, \ldots, h_{R,l}] \) and the input data \( x \). By reducing the dimension and softmax logical regression, the Gate module outputs the prediction result and obtains the output weight of each expert \( W_{e,r} = [W_{e,1}, W_{e,2}, \ldots, W_{e,R}] \). This process can be described as follows

\[
[W_{e,1}, W_{e,2}, \ldots, W_{e,R}] = \text{softmax}(	ext{Gate}(x, h_{1,l}, h_{2,l}, \ldots, h_{R,l}))
\]  

(10)

\[
\sum_{r=1}^{R} W_{e,r} = 1
\]  

(11)

Finally, in order to get the total output of the model, a weighted fusion process is required as follows

\[
y = \sum_{r=1}^{R} (W_{e,r}y^r_{e,r})
\]  

(12)

4.2. MMoE-SAE Model Training. Compared with the traditional end-to-end model training process, each expert is designed as an SAE, and thus the training process of the MMoE-SAE model is different. The parameters of the model mainly have two parts: the SAE parameters of each expert and
the parameters of the gated fusion module. The training process of the MMoE-SAE model is shown in Figure 6.

![Figure 5. Structure of the MMoE-SAE model.](image)

The overall training process starts with unsupervised layer-wise pretraining, where the SAE parameters of each expert can be obtained. The model then freezes all parameters of the SAE and enters a supervised fine-tuning process in which the parameters of the gated fusion module are trained. After two parts of parameter training, the entire model training process is completed. To ensure the specificity of experts in the training process, in the process of model design, the SAE of different experts adopts different loss functions. The mean square error (MSE) loss function and the mean absolute error (MAE) loss function are introduced, which are expressed as follows

\[
MSE = \frac{1}{N} \sum_{i=1}^{N} (\hat{y}_i - \tilde{y}_i)^2
\]

\[
MAE = \frac{1}{N} \sum_{i=1}^{N} |\hat{y}_i - \tilde{y}_i|
\]

where \(N\) represents the number of training samples; \(y_i\) and \(\hat{y}_i\) represent the measured and predicted values of the quality variable, respectively. For each expert, the corresponding design has different weight values of MSE and MAE in the loss function. MSE is more sensitive to outliers, so it is greatly affected by outliers; MAE is relatively low in sensitivity, so it is relatively stable. Assuming that the weight values are \(W_S\) and \(W_A\), the values of the two weights are \([0,1]\), and the sum is 1. The total loss function is described as follows

\[
Loss = W_S \frac{1}{N} \sum_{i=1}^{N} (\hat{y}_i - \tilde{y}_i)^2 + W_A \frac{1}{N} \sum_{i=1}^{N} |\hat{y}_i - \tilde{y}_i|
\]  

\[
W_S + W_A = 1, W_S \in [0,1], W_A \in [0,1]
\]

5. INDUSTRIAL APPLICATION

In this section, the proposed MMoE-SAE model is applied to the actual industrial process of a blast furnace. In the comparative experiments, the MMoE-SAE model was compared with the SAE model and the Bagging-SAE model. Bagging-SAE refers to SAE being processed based on the bagging method.

The experimental data collected are from an industrial blast furnace ironmaking plant. The dataset contains 9000 data samples and 110 process variables. The process variables include blast furnace air temperature, atmospheric humidity,

| model          | RMSE  | \(R^2\) |
|----------------|-------|---------|
| SAE            | 0.0517| 0.8556  |
| bagging-SAE    | 0.0499| 0.8680  |
| MMoE-SAE       | 0.0299| 0.9546  |
Figure 7. Silicon content prediction results using (a) SAE, (b) Bagging-SAE, and (c) MMoE-SAE.

Figure 8. Weight variation curve of each expert.
oxygen enrichment rate, furnace top temperature, furnace top pressure, desulfurization rate, heat flow ratio, and so on. The quality variable is the silicon content. To build the predictive model, the collected data are split into three parts: 6000 samples as the training set, 2000 samples as the validation set, and 1000 samples as the test set. To evaluate the prediction performance of each model, two evaluation criteria are used in the experiment. The first is the root mean squared error (RMSE), which can be calculated as follows

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (y_{i,t} - \hat{y}_{i,t})^2}$$  \hspace{1cm} (17)

The second is the coefficient of determination, which is abbreviated as $R^2$ and is defined as follows

$$R^2 = 1 - \frac{\sum_{i=1}^{N} (y_{i,t} - \hat{y}_{i,t})^2}{\sum_{i=1}^{N} (y_{i,t} - \bar{y}_{i})^2}$$ \hspace{1cm} (18)

where $N$ denotes the number of test data; $y_{i,t}$ and $\hat{y}_{i,t}$ are the measured and predicted values of the quality variable during the testing phase, respectively.

Table 1 shows the prediction results of each model on the test dataset. In the experiment, the SAE structures used are all three layers, and the activation function used in the SAE is the sigmoid function. For the MMOE-SAE model, the number of experts is set to be 3. For each expert, the loss function weights are designed as $[W_c, W_A] = \{[1.0], [0.5,0.5], [0.1]\}$. From Table 1, it can be seen that the Bagging-SAE model achieves better prediction performance with larger $R^2$ and smaller RMSE compared to the SAE. The proposed MMOE-SAE model obtains the best prediction performance among all methods, with the largest $R^2$ and the smallest RMSE. The detailed prediction results between the predicted and real values of the quality variable can be seen in Figure 7. As can be seen from Figure 7, the SAE model and Bagging-SAE model have limited ability to track the frequent changes of the quality variable. This change is the dynamic change process of the blast furnace operating conditions. Compared with the SAE model and Bagging-SAE model, the MMOE-SAE model has good ability to track and describe system characteristics during the dynamic change of the operating conditions. The predicted results of MMOE-SAE are in good agreement with the actual values. This is because the MMOE-SAE model can adjust the weights of experts accordingly when the operating conditions change. This dynamic and natural adjustment enables the model to have better self-adaptation ability, and thus have better comprehensive prediction ability for blast furnace product quality. This is shown in Figure 8, which shows the change in the weight of each expert.

6. CONCLUSIONS

In this paper, a new deep learning model called MMOE-SAE is proposed and applied to predict the quality variable of the blast furnace ironmaking process. The MMOE-SAE model is designed based on an MMOE structure, in which the SAE network is selected as an expert. By integrating a series of SAE networks into the MMOE structure, the constructed deep learning model has better adaptability to the working conditions of the blast furnace system. The effectiveness of the proposed MMOE-SAE model is verified through the real blast furnace ironmaking data. The application results prove that the proposed MMoE-SAE model is superior to the other predictive models in prediction accuracy.

■ AUTHOR INFORMATION

Corresponding Author

Xinmin Zhang — State Key Laboratory of Industrial Control Technology, College of Control Science and Engineering, Zhejiang University, Hangzhou 310027, China; orcid.org/0000-0002-4761-3969; Email: xinminzhang@zju.edu.cn

Authors

Hongyu Zhu — State Key Laboratory of Industrial Control Technology, College of Control Science and Engineering, Zhejiang University, Hangzhou 310027, China
Bocun He — State Key Laboratory of Industrial Control Technology, College of Control Science and Engineering, Zhejiang University, Hangzhou 310027, China

Complete contact information is available at: https://pubs.acs.org/10.1021/acsomega.2c05029

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported in part by the National Natural Science Foundation of China (grant no. 62003301) and in part by the Major Science and Technology Infrastructure Project of Zhejiang Laboratory (large-scale experimental device for information security of new generation industrial control system) (grant no. 226-2022-00001).

■ REFERENCES

(1) Geerdes, M.; Chaigneau, R.; Lingiardi, O. Modern Blast Furnace Ironmaking: An Introduction; Ios Press, 20202020.
(2) Kuang, S.; Li, Z.; Yan, D.; Ji, Y.; Yu, A. Numerical study of hot charge operation in ironmaking blast furnace. Miner. Eng. 2014, 63, 45–56.
(3) Zankl, D.; Schuster, S.; Feger, R.; Stelzer, A.; Scheibelhofer, S.; Schmid, C. M.; Ossberger, G.; Stegellner, L.; Lengauer, G.; Feilmayr, C.; Lackner, B.; Burgler, T. BLASTDAR-A Large Radar Sensor Array System for Blast Furnace Burden Surface Imaging. IEEE Sens. J. 2015, 15, 5893–5909.
(4) Gao, C.; Jian, L.; Luo, S. Modeling of the Thermal State Change of Blast Furnace Hearth With Support Vector Machines. IEEE Trans. Ind. Electron. 2012, 59, 1134–1145.
(5) Zhou, P.; Yuan, M.; Wang, H.; Wang, Z.; Chai, T.-Y. Multivariable dynamic modeling for molten iron quality using online sequential random vector functional-link networks with self-feedback connections. Inf. Sci. 2015, 325, 237–255.
(6) Zhao, J.; Wang, W.; Liu, Y.; Pedrycz, W. A Two-Stage Online Prediction Method for a Blast Furnace Gas System and Its Application. IEEE Trans. Control Syst. Technol. 2011, 19, 507–520.
(7) Zhang, X.; Kano, M.; Matsuzaki, S. A comparative study of deep and shallow predictive techniques for hot metal temperature prediction in blast furnace ironmaking. Comput. Chem. Eng. 2019, 130, 106575.
(8) Naito, M.; Takeda, K.; Matsui, Y. Ironmaking Technology for the Last 100 Years: Deployment to Advanced Technologies from Introduction of Technological Know-how, and Evolution to Next-generation Process. ISIJ Int. 2015, 55, 7–35.
(9) Ostermark, R.; Saxén, H. VARMAX-modelling of blast furnace process variables. Eur. J. Oper. Res. 1996, 90, 85–101.
(10) Fontes, D. O. L.; Vasconcelos, L. G. S.; Brito, R. P. Blast furnace hot metal temperature and silicon content prediction using
soft sensor based on fuzzy C-means and exogenous nonlinear autoregressive models. *Comput. Chem. Eng.* 2020, 141, 107028.

(11) Shi, L.; Li, Z.; Yu, T.; Li, J. Model of Hot Metal Silicon Content in Blast Furnace Based on Principal Component Analysis Application and Partial Least Square. *J. Iron Steel Res. Int.* 2011, 18, 13–16.

(12) Zhang, X.; Wei, C.; Song, Z. Fast locally weighted PLS modeling for large-scale industrial processes. *Ind. Eng. Chem. Res.* 2020, 59, 20779–20786.

(13) Jian, L.; Gao, C.; Li, L.; Zeng, J. Application of Least Squares Support Vector Machines to Predict the Silicon Content in Blast Furnace Hot Metal. *ISIJ Int.* 2008, 48, 1659–1661.

(14) Jian, L.; Gao, C. Binary Coding SVMs for the Multiclass Problem of Blast Furnace System. *IEEE Trans. Ind. Electron.* 2013, 60, 3846–3856.

(15) Jian, L.; Gao, C.; Xia, Z. Constructing Multiple Kernel Learning Framework for Blast Furnace Automation. *IEEE Trans. Autom. Sci. Eng.* 2012, 9, 763–777.

(16) Zhou, P.; Chen, W.; Yi, C.; Jiang, Z.; Yang, T.; Chai, T. Fast just-in-time-learning recursive multi-output LSSVR for quality prediction and control of multivariable dynamic systems. *Eng. Appl. Artif. Intell.* 2021, 100, 104168.

(17) Yuan, M.; Zhou, P.; Li, M.; Li, R.; Wang, H.; Chai, T. Intelligent Multivariable Modeling of Blast Furnace Molten Iron Quality Based on Dynamic AGA-ANN and PCA. *J. Iron Steel Res. Int.* 2015, 22, 487–495.

(18) Zhang, H.; Yin, Y.; Zhang, S. An improved ELM algorithm for the measurement of hot metal temperature in blast furnace. *Neurocomputing* 2016, 174, 232–237.

(19) Li, J.; Hua, C.; Yang, Y.; Guan, X. A Novel MIMO T-S Fuzzy Modeling for Prediction of Blast Furnace Molten Iron Quality With Missing Outputs. *IEEE Trans. Fuzzy Syst.* 2021, 29, 1654–1666.

(20) Charte, D.; Charte, F.; García, S.; del Jesus, M. J.; Herrera, F. A practical tutorial on autoencoders for nonlinear feature fusion: Taxonomy, models, software and guidelines. *Inf. Fusion* 2018, 44, 78–96.

(21) Bengio, Y.; Lamblin, P.; Popovici, D.; Larochelle, H. Greedy layer-wise training of deep networks. *Proceedings of the 19th International Conference on Neural Information Processing Systems*, 2006, pp. 153–160.

(22) Ma, J.; Zhao, Z.; Yi, X.; Chen, J.; Hong, L.; Chi, E. H. Modeling Task Relationships in Multi-Task Learning with Multi-Gate Mixture-of-Experts. *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*: New York, NY, USA, 2018, pp 1930–1939.

(23) Shazeer, N. M.; Mirhoseini, A.; Maziarz, K.; Davis, A.; Le, Q. V.; Hinton, G. E.; Dean, J. Outrageously Large Neural Networks: The Sparsely-Gated Mixture-of-Experts Layer. 2017, *ArXiv:1701.06538*.

(24) Qin, Z.; Cheng, Y.; Zhao, Z.; Chen, Z.; Metzler, D.; Qin, J. Multitask Mixture of Sequential Experts for User Activity Streams. *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*: New York, NY, USA, 2020, pp 3083–3091.