

Deep Learning Based Silicon Content Estimation in Ironmaking Process

Heng Zhou* Haifeng Zhang** Chunjie Yang***
 Youxian Sun****

* 1. College of Control Science and Engineering, Zhejiang University, Hangzhou 310027 China (e-mail: zhouheng@zju.edu.cn).

2. Department of Chemical and Biomolecular Engineering, University of California, Los Angeles, CA 90024 USA (e-mail: zhouheng309@ucla.edu).

** Guangxi Liuzhou Iron and Steel Group Co. Ltd., Liuzhou 545002 China (e-mail: 13627728767@163.com).

*** College of Control Science and Engineering, Zhejiang University, Hangzhou 310027 China (e-mail: cjyang999@zju.edu.cn).

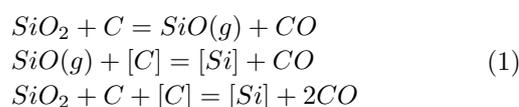
**** College of Control Science and Engineering, Zhejiang University, Hangzhou 310027 China (e-mail: yxsun@ipc.zju.edu.cn).

Abstract: Given the complexity and isolation of the blast furnace (BF), field engineers generally operate the system upon their former experience and the operating manual. Harsh environment and equipment shortage have made the testing of silicon content a prevailing method for the detection of temperature within BF. As the silicon content is a comprehensive performance of internal thermal state, knowing the exact value in advance can be very helpful for operators to keep the furnace temperature at a reasonable extent. Thus, an improved gated recurrent unit recurrent neural network (GRU-RNN) is proposed to predict the silicon content of hot metal, indicating a competitive performance at 92.4% hit rate among several deep learning methods.

Keywords: Deep Learning, Silicon Content, Blast Furnace.

1. INTRODUCTION

Iron making process is essentially a nonlinear and dynamic process due to the complexity of heat and mass transfer, chemical reactions, and phase change. Take extreme conditions in the sealed BF (Fig. 1) as an example, it is difficult to control the system stably and safely, resulting in an extra waste of fuel and energy. As the BF accounts for 70% of energy consumption during the whole production chain, little improvement in temperature control strategy can save huge money and make great contributions to environmental protection. It is widely accepted that the percentage of silicon content in hot metal can represent the comprehensive thermal state due to the shortage of detection and analysis methods for temperature within the BF by Chen and Gao (2019); Li et al. (2017b); Jiang et al. (2020). There are three kinds of materials including silica, carbon, and oxygen that participate in the silicon reduction reactions. As Peacey and Davenport (2016) has written, the chemical concept of silica reduction reactions are illustrated in Eq. 1.



* This work is supported by the National Science Foundation of China (61933015). This work is also funded by China Scholarship Council (CSC).

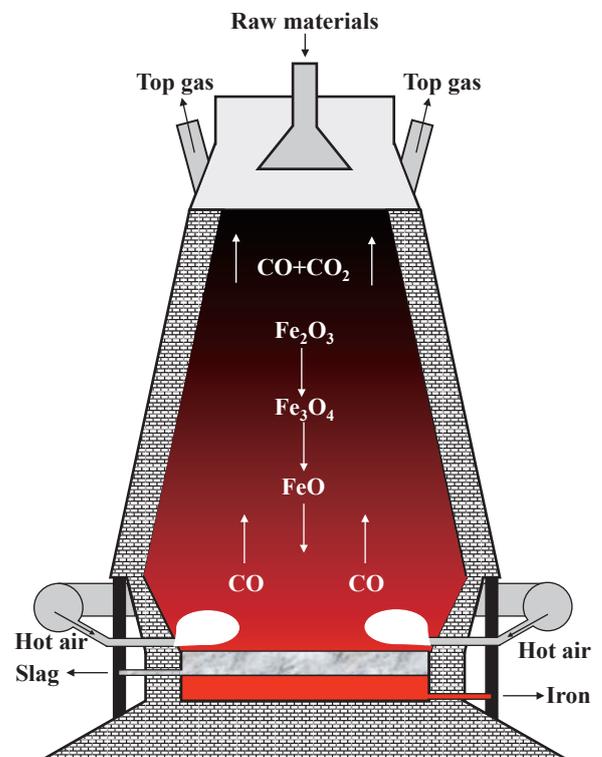


Fig. 1. The schematic structure of blast furnace.

The heat, mass and momentum transferring in BF are strongly influenced by deeply correlated external factors such as air temperature, pressure differential and O_2 enrichment. With materials descending in BF, the reduction of silica starts from the belly and ends with the tuyere level, reaching the maximum of silicon content. After that, partial silicon is oxidized again at the metal-slag interface. With those being said, the silicon content is an integrated expression of the thermal status inside the BF, leading to a variety of mechanism and data-driven prediction models. The pruning algorithms in Nurkkala et al. (2011); Xie and Zhou (2020) were based on feedforward neural networks, as a start, have been used to find the relevant inputs, time lags and network connectivity. Besides, some other intelligent algorithms like evolutionary algorithms, principal component analysis, and fuzzy neural network have also been applied to silicon content modeling by Zhou et al. (2020a,b); Zhou et al. (2017a); Yin et al. (2020). Even though much attention has been drawn to deep learning amongst all machine learning methods with the help of TensorFlow, seldom have they been utilized to solve the exiting problems in the process industry.

In recent years, LeCun et al. (2015); Ma et al. (2015); Le (2013) have propelled deep learning structures such as deep belief network, deep neural network and recurrent neural network to the forefront of automatic speech recognition, natural language processing, and man-machine games. In 2017, the world No.1 ranked Go player Jie Ke is defeated by AlphaGo, a program that is developed upon deep neural network and tree search algorithms under TensorFlow framework proposed by Silver et al. (2016). TensorFlow, an open source project, serves as math library for data-flow computation across a variety of tasks. With no doubts, the application of deep neural networks under the TensorFlow framework has great potential in estimation of silicon content in hot metal.

2. METHODOLOGY

The proposed model to learn the mapping rules between input and output arguments is based on deep neural network, more accurately, the modified recurrent neural network used by Guo et al. (2020); Liu et al. (2019); Alemany et al. (2019). The deep learning method consists of input layer, output layer, and hidden layer that is modified by disposition gated recurrent unit (dGRU). Before applying the dGRU-RNN to learn the physical characteristics of BF, it is necessary to determine the input variables as well as their time lags by a feature selection approach. In this paper, we depend on mutual information (MI) to prepare the input arguments of proposed deep learning method.

2.1 MI Feature Selection

In the field of machine learning, it is widely accepted that algorithms can only approach the upper bound of performance determined by feature selection and data characteristics. Therefore, it is of vital importance to select proper input arguments before modeling the industrial process by machine learning methods. To reduce the dimensionality and complexity of modeling process, a feature selection approach based on MI is used to select

typical input variables of dGRU-RNN. The MI, used by Cakir et al. (2019); Cai and Verdú (2019), calculates the degree of nonlinear relationship between variables based on entropy, which is a measurement of uncertainty for the random variables. The entropy $H(X)$ of a single random variable X is defined as:

$$H(X) = - \sum_{x \in \mathcal{X}} p(x) \log p(x) \quad (2)$$

where X belongs to \mathcal{X} and $p(x)$ is the probability density function.

Based on the uni-variable entropy in Eq. 2, we can get the conditional entropy $H(Y|X)$ of two random variables (X, Y) in Eq. 3.

$$\begin{aligned} H(Y|X) &= \sum_{x \in \mathcal{X}} p(x) H(Y|X = x) \\ &= - \sum_{x \in \mathcal{X}} p(x) \sum_{y \in \mathcal{Y}} p(y|x) \log p(y|x) \\ &= - \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p(x, y) \log p(y|x) \end{aligned} \quad (3)$$

where $p(x, y)$ is the joint probability distribution function.

Thus, the definition of MI is illustrated in Eq. 4, where the $I(X; Y)$ represents the uncertainty reduction of random variable X under given Y .

$$\begin{aligned} I(X; Y) &= \sum_{x, y} p(x, y) \log \frac{p(x, y)}{p(x)p(y)} \\ &= \sum_{x, y} p(x, y) \log \frac{p(x|y)}{p(x)} \\ &= - \sum_{x, y} p(x, y) \log p(x) + \sum_{x, y} p(x, y) \log p(x|y) \\ &= - \sum_x p(x) \log p(x) - \left(- \sum_{x, y} p(x, y) \log p(x|y) \right) \\ &= H(X) - H(X|Y) \end{aligned} \quad (4)$$

On the whole, when MI equals to zero, two variables have no connection at all, corresponding that when MI is equal to one, they totally depend on each other.

2.2 Improved GRU-RNN

With numerous variations, the hidden layers are efficient and powerful due to distributed representations and compositionality. The RNN plays a significant role in the time series analysis as the hidden layer nodes form a directional cycle. By adding internal memory units, RNN is able to store history information and process sequence inputs. Detailed graphical and mathematical explanations are listed in Fig. 2 and Eq. 5.

$$h_t = \phi(W_{xh} \cdot x_t + W_{hh} \cdot h_{t-1} + b) \quad (5)$$

where h_t is the hidden state at this moment, h_{t-1} is the hidden state at last moment, x_t is the input variables at this moment, b is the threshold, ϕ is the activation function, W_{xh} is the weight of input variables, and W_{hh} is the weight of hidden state. The sharing of W_{xh} and W_{hh} enables RNN to deal with arbitrary length of time series.

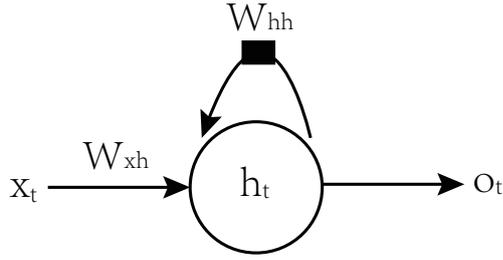


Fig. 2. The schematic structure of RNN.

When ignoring the nonlinear part in Eq. 5, the equation can be simplified as Eq. 6. The eigenvalue of W_{hh}^t can easily lead to zero or infinity, leading to the obscurity of the information contained in h_0 .

$$h_t = W_{hh} \cdot h_{t-1} = W_{hh}^t \cdot h_0 \quad (6)$$

In order to avoid the gradient vanishing and explosion, researchers develop leaky units whose linear self-connections weight is manual setting. Furthermore, the weight in gated RNNs is able to update under the supervision of gates in each computational step. There are two major kinds of gates in RNN, one is long-short term memory (LSTM) like Li et al. (2019), and the other is gated recurrent unit (GRU) like Yan et al. (2019). The GRU depends on reset gate and update gate to control the magnitude of data that flow by. Different from conventional GRU, the dGRU uses disposition gate to replace the reset and update gates, which is explained in Fig. 3 and Eq. 7.

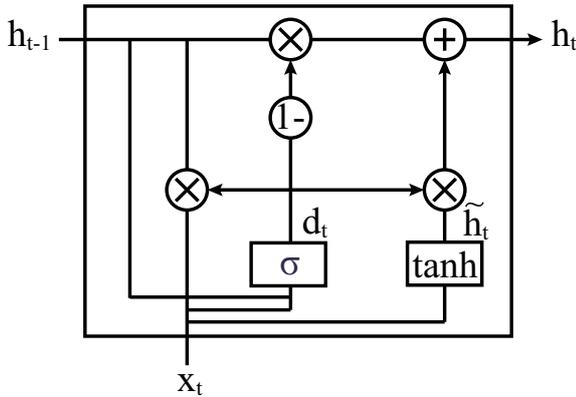


Fig. 3. The internal structure of modified dGRU.

$$\begin{aligned} h_t &= (1 - d_t) \circ h_{t-1} + d_t \circ \tilde{h}_t \\ \tilde{h}_t &= \tanh(Wx_t + d_t U h_{t-1}) \\ d_t &= \sigma(W_d x_t + U_d h_{t-1}) \end{aligned} \quad (7)$$

where x_t is the input vector, h_t is the output of hidden layer, d_t is the disposition gate. Besides, the σ represents the sigmoid activation function.

3. EXPERIMENTS

3.1 Feature Selection

To verify the effectiveness of proposed algorithm, we apply it to estimate the silicon content of hot metal in ironmaking process. The datasets sampled from multiple

Table 1. Process variables in the BF

Designation	Variable	Unit
x_1	Top blast pressure	kPa
x_2	Gas permeability	$m^3/min \cdot kPa$
x_3	Coal rate	t/h
x_4	O_2 enrichment ratio	vol%
x_5	Top blast temperature	$^{\circ}C$
x_6	Total pressure differential	kPa
x_7	Hot air pressure	kPa
x_8	Hot air temperature	$^{\circ}C$
x_9	Hot air volume rate	m^3/min
x_{10}	Cool air humidity	vol%
x_{11}	Former silicon content	wt%
y	Silicon content	wt%

sources are regularized to the same time scale according to the chemical test of silicon content. Therefore, we obtained 12 variables with 1000 samples, as illustrated in Table 1. The results of MI calculation between operation variables and silicon content are shown in Fig. 4. It is easy to notice that x_{11} is most related to y , while other decision variables, including x_1 , x_3 , and x_5 , cannot be ignored. Thus, the input variables of dGRU-RNN chosen by feature selection is $[x_1, x_3, x_5, x_{11}]$.

It is reasonable that the former silicon content has the greatest influence as the ironmaking is a continuous process with large inertia. In view of the silica reduction procedures, other parameters can be analyzed by chemical principle such as kinetic equation Eq. 8 and Arrhenius equation Eq. 9. It is easy to notice that the reactant concentration c_A and reaction rate constant k are the two critical factors that influence chemical reaction rate in Eq. 8. From Eq. 9, the temperature characterized by top blast temperature and top blast pressure has a great influence on the reaction rate constant. Besides, silicon content in the pig iron, illustrated by Zhou et al. (2017b), mostly comes from the coal and coke rather than the iron ore.

$$r = kc_A^n \quad (8)$$

$$\ln k = \ln A - \frac{Ea}{RT} \quad (9)$$

Furthermore, the time lags need to be specified due to the high relation between former and current silicon content. The partial autocorrelation function (PACF) is able to estimate the optimal order in time series regression analysis by measuring the relationship of time series at t and $t - k$. From Fig. 5, it is easy to find that the value of PACF has a significant decrease within 2 lags. As a result, the time lag of silicon content is set as 2, i.e., the input vector includes previous silicon content two steps backwards.

3.2 Simulation Results

1000 continuous samples are obtained from the BF, of which 80% are set as training data and 20% are test data. The prediction results from different methods are exhibited in Fig. 6. Obviously, dGRU-RNN has the most competitive performance among these RNN variants, not only in loss function, but also in accuracy. In industrial process, the accuracy in prediction of silicon content is defined by hit rate (HR) Li et al. (2017a), representing the percentage of predicted value whose absolute difference

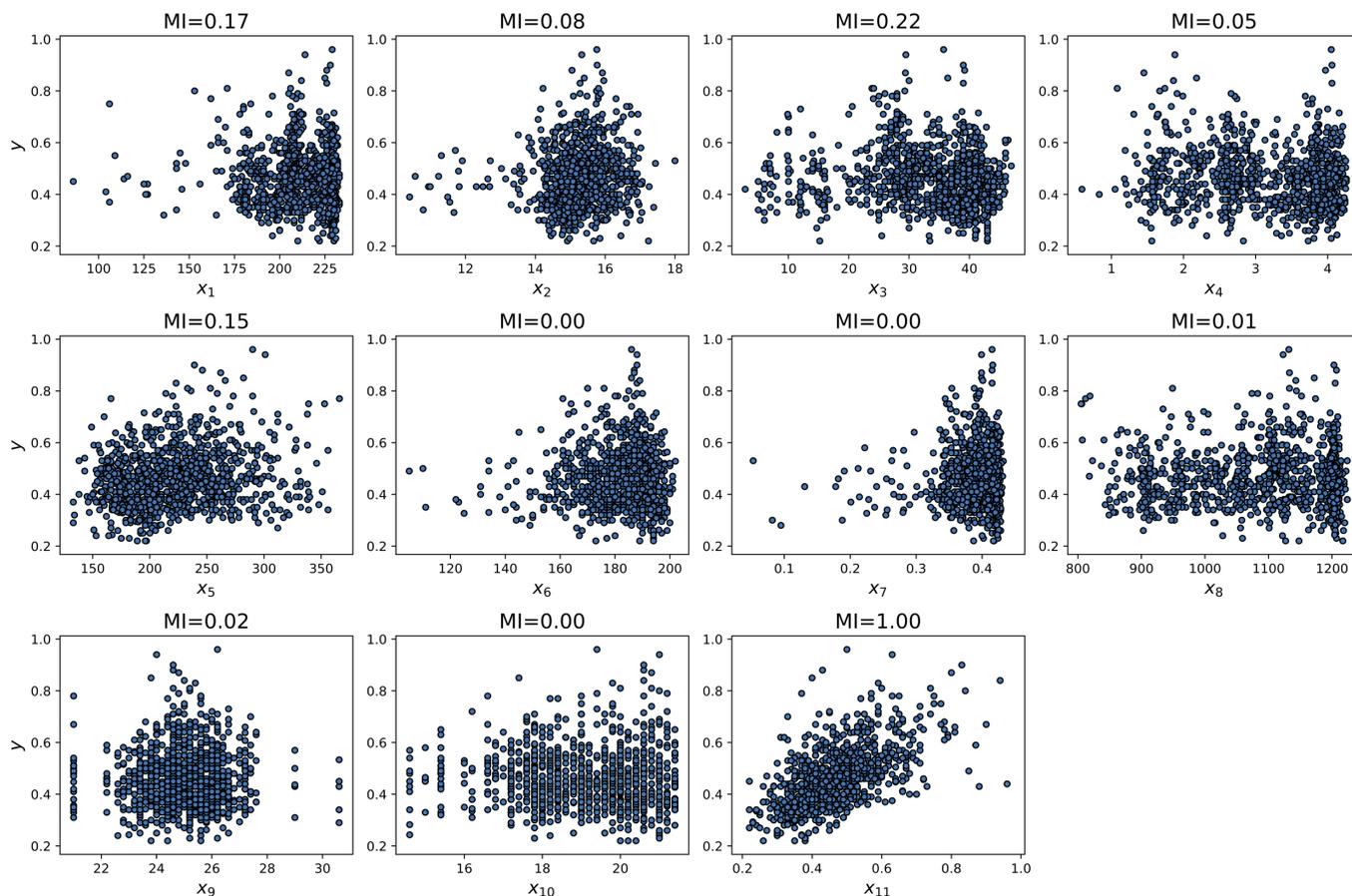


Fig. 4. MI relation between input and output arguments.

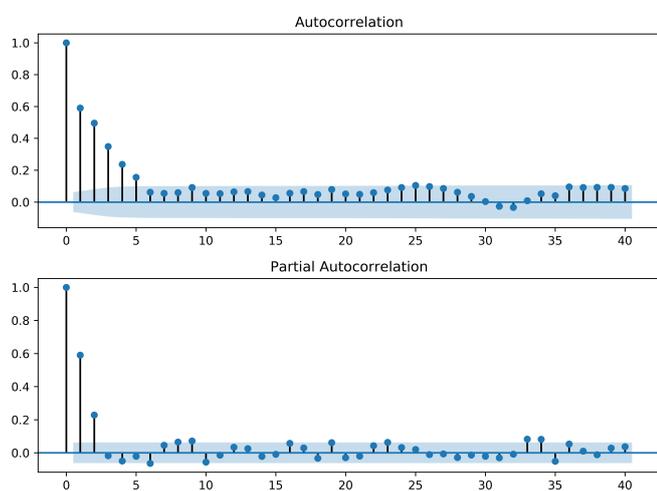


Fig. 5. Correlation of silicon content series.

with real data is less than one. Eq. 10 provide a detailed mathematical definition for HR.

$$\begin{aligned}
 H_i &= \begin{cases} 1, & |e(i)| \leq 0.1 \\ 0, & \text{others} \end{cases} \\
 HR &= \frac{1}{n} \left(\sum_{i=1}^n H_i \right) \cdot 100\%
 \end{aligned}
 \tag{10}$$

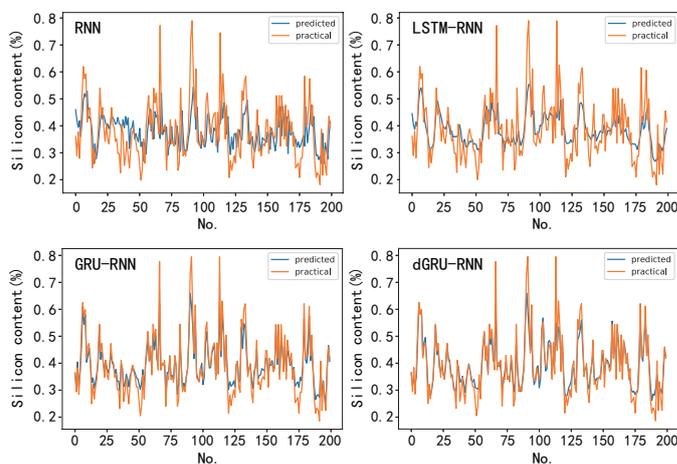


Fig. 6. Prediction results of RNN variants.

Table 2. Comparison among RNN variants

Model	Loss function		HR/%
	train	test	
RNN	1.01	1.46	75.1
LSTM-RNN	0.54	0.73	82.1
GRU-RNN	0.42	0.58	87.2
dGRU-RNN	0.22	0.31	92.4

From Table 2, the deep dGRU-RNN model shows great advantages at both loss function and HR in prediction of silicon content. The 92.4% HR of dGRU-RNN has a

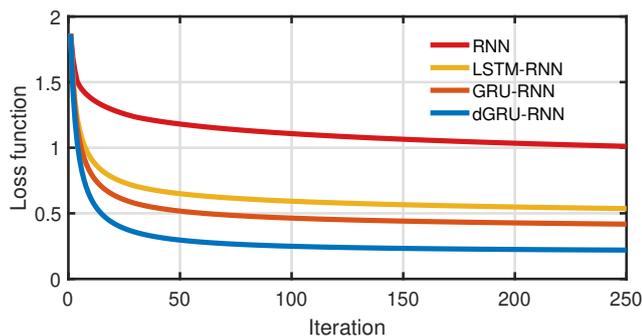


Fig. 7. The loss function of RNN variants during testing process.

huge advantage over the other three RNNs. Besides, the training loss of standard RNN is 1.01, much higher than 0.22 of dGRU-RNN. The Fig. 7 provides a detailed change trend of the final test loss, in which the dGRU-RNN can reach 0.31, much lower than 1.46 of deep RNN after 200 epochs, showing an obtuse fluctuation of prediction error. In consideration of the complex environment of manufacturing process, 90% HR is high enough to help the field engineers to control the temperature of pig iron in advance. The high accuracy and stability of dGRU-RNN algorithm indicates great potential in prediction of BF internal temperature.

4. CONCLUSIONS

Previous studies have promoted the superiority of gated RNN in prediction of time series. When combined with proper input process variables, as well as typical time lags, the deep dGRU-RNN algorithm can learn the physics of ironmaking process well and give out proper suggestions to help engineers operate the BF, indicating a bright future in online silicon content prediction. However, the proposed deep learning framework solely pertains to silicon content in hot metal without concerning other iron quality indexes such as temperature and sulfur content, which deserves further investigations.

REFERENCES

Alemanly, S., Beltran, J., Perez, A., and Ganzfried, S. (2019). Predicting hurricane trajectories using a recurrent neural network. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, 468–475.

Cai, C. and Verdú, S. (2019). Conditional rényi divergence saddlepoint and the maximization of α -mutual information. *Entropy*, 21(10), 969.

Cakir, F., He, K., Bargal, S.A., and Sclaroff, S. (2019). Hashing with mutual information. *IEEE transactions on pattern analysis and machine intelligence*, 41(10), 2424–2437.

Chen, S. and Gao, C. (2019). Linear priors minded and integrated for transparency of blast furnace black-box svm model. *IEEE Transactions on Industrial Informatics*, 1–1.

Guo, K., Hu, Y., Qian, Z., Liu, H., Zhang, K., Sun, Y., Gao, J., and Yin, B. (2020). Optimized graph convolution recurrent neural network for traffic prediction.

IEEE Transactions on Intelligent Transportation Systems, 1–12.

Jiang, K., Jiang, Z., Xie, Y., Chen, Z., Pan, D., and Gui, W. (2020). Classification of silicon content variation trend based on fusion of multilevel features in blast furnace ironmaking. *Information Sciences*, 521, 32 – 45.

Le, Q.V. (2013). Building high-level features using large scale unsupervised learning. In *2013 IEEE international conference on acoustics, speech and signal processing*, 8595–8598. IEEE.

LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *nature*, 521(7553), 436.

Li, C., Wang, Z., Rao, M., Belkin, D., Song, W., Jiang, H., Yan, P., Li, Y., Lin, P., Hu, M., et al. (2019). Long short-term memory networks in memristor crossbar arrays. *Nature Machine Intelligence*, 1(1), 49–57.

Li, J., Hua, C., Yang, Y., and Guan, X. (2017a). Bayesian block structure sparse based t-s fuzzy modeling for dynamic prediction of hot metal silicon content in the blast furnace. *IEEE Transactions on Industrial Electronics*, 65(6), 4933–4942.

Li, J., Hua, C., Yang, Y., and Guan, X. (2017b). Fuzzy classifier design for development tendency of hot metal silicon content in blast furnace. *IEEE Transactions on Industrial Informatics*, 14(3), 1115–1123.

Liu, Y., Zheng, Y., Lu, J., Cao, J., and Rutkowski, L. (2019). Constrained quaternion-variable convex optimization: A quaternion-valued recurrent neural network approach. *IEEE Transactions on Neural Networks and Learning Systems*, 1–14.

Ma, J., Sheridan, R.P., Liaw, A., Dahl, G.E., and Svetnik, V. (2015). Deep neural nets as a method for quantitative structure–activity relationships. *Journal of chemical information and modeling*, 55(2), 263–274.

Nurkkala, A., Pettersson, F., and Saxén, H. (2011). Non-linear modeling method applied to prediction of hot metal silicon in the ironmaking blast furnace. *Industrial & Engineering Chemistry Research*, 50(15), 9236–9248.

Peacey, J.G. and Davenport, W.G. (2016). *The iron blast furnace: theory and practice*. Elsevier.

Silver, D., Huang, A., Maddison, C.J., Guez, A., Sifre, L., Van Den Driessche, G., Schrittwieser, J., Antonoglou, I., Panneershelvam, V., Lanctot, M., et al. (2016). Mastering the game of go with deep neural networks and tree search. *nature*, 529(7587), 484.

Xie, J. and Zhou, P. (2020). Robust stochastic configuration network multi-output modeling of molten iron quality in blast furnace ironmaking. *Neurocomputing*.

Yan, Z., Yang, K., Wang, Z., Yang, B., Kaizuka, T., and Nakano, K. (2019). Time to lane change and completion prediction based on gated recurrent unit network. In *2019 IEEE Intelligent Vehicles Symposium (IV)*, 102–107. IEEE.

Yin, X., Niu, Z., He, Z., Li, Z., and Lee, D. (2020). An integrated computational intelligence technique based operating parameters optimization scheme for quality improvement oriented process-manufacturing system. *Computers & Industrial Engineering*, 140, 106284.

Zhou, H., Zhang, H., and Yang, C. (2020a). Hybrid-model-based intelligent optimization of ironmaking process. *IEEE Transactions on Industrial Electronics*, 67(3), 2469–2479.

- Zhou, H., Yang, C., Liu, W., and Zhuang, T. (2017a). A sliding-window ts fuzzy neural network model for prediction of silicon content in hot metal. *IFAC-PapersOnLine*, 50(1), 14988–14991.
- Zhou, H., Yang, C., Zhuang, T., Li, Z., Li, Y., and Wang, L. (2017b). Multi-objective optimization of operating parameters based on neural network and genetic algorithm in the blast furnace. In *2017 36th Chinese Control Conference (CCC)*, 2607–2610. IEEE.
- Zhou, P., Zhang, R., Xie, J., Liu, J., Wang, H., and Chai, T. (2020b). Data-driven monitoring and diagnosing of abnormal furnace conditions in blast furnace ironmaking: an integrated pca-ica method. *IEEE Transactions on Industrial Electronics*, 1–1.