A Hybrid Modeling Method Based on Expert Control and Deep Neural Network for Temperature Prediction of Molten Steel in LF

Zi-cheng XIN,¹⁾ Jiang-shan ZHANG,¹⁾ Jin ZHENG,²⁾ Yu JIN²⁾ and Qing LIU^{1)*}

1) State Key Laboratory of Advanced Metallurgy, University of Science and Technology Beijing, Beijing, 100083 China.

2) Tangshan Branch, Hebei Iron and Steel Co., Ltd., Tangshan, Hebei, 063000 China.

(Received on June 2, 2021; accepted on November 16, 2021; J-STAGE Advance published date: January 8, 2022)

The temperature control of molten steel in ladle furnace (LF) has a critical impact on steelmaking production. In this work, production data were collected from a steelmaking plant and a hybrid model based on expert control and deep neural network (DNN) was established to predict the molten steel temperature in LF. In order to obtain the optimal DNN model, the trial and error method was used to determine the hyperparameters. And the optimal architecture of DNN model corresponds to the hidden layers of 4, hidden layer neurons of 35, iterations of 3 000, and learning rate of 0.2. Compared with the multiple linear regression model and the shallow neural network model, the DNN model exhibits stronger generalisation performance and higher accuracy. The coefficient of determination (R^2), correlation coefficient (r), mean square error (MSE), and root-mean-square error ($RMSE$) of the optimal DNN model reached 0.897, 0.947, 2.924, 1.710, respectively. Meanwhile, in the error scope of temperature from -5 to 5°C , the hit ratio of the hybrid model acquired 99.4%. The results demonstrate that the proposed model is effective to predict temperature of molten steel in LF.

KEY WORDS: ladle furnace; temperature prediction; DNN; expert control; metallurgical mechanism.

1. Introduction

With the development of industrial intelligence and the increasing pressure of environmental protection, the intelligence, greenization and low carbonization have been paid more attention to the transformation and upgrading of steel-making plants. Ladle furnace (LF) has been widely used in steelmaking plants for its low equipment investment and outstanding refining performance. The functions of LF refining include deoxidation, desulfurization, inclusion removal, adjustment of temperature and composition of molten steel, and homogenization of molten steel temperature and composition. The other main function is to ensure the smooth running of continuous casting process by controlling the temperature of molten steel. Therefore, it is of great significance to conduct accurate prediction of molten steel temperature, which is of benefit to the intelligent control of LF refining process.^{1,2)}

Many researchers have studied the temperature prediction of molten steel in LF. There are currently three strategies to calculate or predict this temperature in the literature. One is to develop mechanism models and then obtain the correlation between temperature and the influencing parameters by using

energy conservation equation, heat transfer equation, and mass conservation equation. Wu *et al.*³⁾ established the mechanistic model of the heating rate of molten steel by using the law of energy conservation. Nath *et al.*⁴⁾ developed a LF on-line reckoner based on simplified physics, material and heat balance, and statistical analysis of plant data to predict temperature and composition of steel. However, due to the complexity of LF refining process, many key technological parameters cannot be obtained directly, such as heat transfer coefficient of ladle furnace shell, temperature drop coefficient of argon stirring, heat loss coefficient of slag surface, which makes it difficult for the mechanism model to achieve accurate prediction. The second strategy is based on the intelligent algorithm to establish the nonlinear correlation between temperature and the influencing parameters. Tian *et al.*⁵⁻⁸⁾ predicted the temperature of molten steel in LF by using extreme learning machine (ELM), back propagation (BP) neural network, and modified adaptive boosting (AdaBoost) algorithm. Lü *et al.*⁹⁾ predicted the temperature of molten steel in LF based on optimally pruned bagging combined with partial linear extreme learning machine (PLELM), indicating the prediction accuracy of the model is higher by comparing with genetic algorithm-back propagation (GA-BP) model, partial least squares-support vector machine (PLS-SVM) model, and AdaBoost. RT-ELM model. However, the intelligent

* Corresponding author: E-mail: qliu@metall.ustb.edu.cn

model excessively relies on data, which makes the model sensitive to abnormal data and vulnerable to the influence of imbalance data. The third strategy is based on combining the mechanism model and the intelligent algorithm to establish the hybrid model. Tang *et al.*¹⁰⁾ and Fu *et al.*¹¹⁾ established a gray box model to predict the end-point temperature of molten steel in LF based on the mechanistic model and the black box model, indicating the prediction accuracy of the model is higher, which provides guidance for the LF refining process. However, the BP, SVM and ELM algorithm belong to shallow neural networks (SNN), and their limitations are mainly reflected in that not only the expression ability of complex functions in finite samples is limited, but also the generalization ability is restricted to some extent.¹²⁾

With the rapid development of artificial intelligence technology, another widely used burgeoning algorithm is deep neural networks (DNN).¹³⁾ Kwon *et al.*¹⁴⁾ predicted the hot ductility of steels based on elemental composition and thermal history by using the DNN indicating superior to the previous studies. Myers *et al.*¹⁵⁾ predicted the nucleation lag time based on the elemental composition and the temperature for iron and steelmaking slags by using DNN. The DNN algorithm highlights the importance of feature learning. Through multi-layer nonlinear feature transformation, the feature representation of the sample in the original space is transformed into a new feature space. Meanwhile, each hidden layer can learn part of the data features, which can make the network store more information. The DNN algorithm can solve more complex data mapping problem and has stronger expression ability of complex function and generalization ability, which is of great benefit to the accurate prediction of molten steel temperature in LF refining process.

In this study, a hybrid model based on the expert control and the DNN algorithm was established to predict the temperature of molten steel in LF. The original production data collected from a steelmaking plant in China were pre-processed to delete outliers, and then the correlation analysis between the input and output variables was carried out. Moreover, the hybrid model was evaluated and compared with other models by using different evaluation indexes. It is expected to establish a more accurate and applicatory hybrid model to predict the temperature of molten steel in LF.

2. Analysis of LF Refining Process

2.1. Description of LF Refining Process

The steelmaking process includes BOF→LF→CC. The specific flow chart of LF refining process is shown in Fig. 1. When the ladle reached the LF station, the starting temperature of molten steel was measured, and a high-basicity reducing slag was obtained by adding slag making materials and deoxidizer. During the LF refining process, slag making materials and alloy are added to adjust the final composition of molten steel. When the composition of molten steel is qualified, the calcium packed wire is fed to realize the inclusion modification. Finally, the end temperature of molten steel is measured before the ladle leaves the LF station.

2.2. Analysis Conservation of Energy

According to the law of conservation of energy, the energy can be divided into heat income and heat output in

Fig. 1. The specific flow chart of LF refining process.¹⁶⁾ (Online version in color.)

Fig. 2. The heat budget of LF refining. (Online version in color.)

LF refining. The heat income comes from the heat of electric arc, the chemical reaction heat of additions. The heat output includes the heat exchanges between ladle furnace and surroundings, the heat storage of ladle shell, the heat loss from the slag surface, the heat loss by argon stirring, and the heat required for heating molten steel, as shown in Fig. 2. The Q_{arc} is the heat of electric arc; the $Q_{addition}$ is the chemical reaction heat of additions; the Q_{shell} is the heat exchanges between ladle furnace and surroundings; the $Q_{storage}$ is the heat storage of ladle; the $Q_{surface}$ is the heat loss from the slag surface; the Q_{argon} is the heat loss by argon stirring; the Q_{steel} is the heat change of molten steel. The LF can be seen as a system, and the heat income and the heat output of LF refining can reach a balance, as shown in Eq. (1).

$$Q_{arc} + Q_{addition} = Q_{surface} + Q_{argon} + Q_{shell} + Q_{storage} + Q_{steel} \dots (1)$$

According to the relationship between input and output of energy, the heat change of molten steel can be expressed as Eq. (2).

$$Q_{steel} = Q_{arc} + Q_{addition} - Q_{surface} - Q_{argon} - Q_{shell} - Q_{storage} \dots (2)$$

2.3. Analysis of Main Factors

Based on the actual operating conditions and the analysis of energy conservation of LF refining process, the main factors affecting the temperature prediction of molten steel

were put forward. As shown in **Table 1**, the main factors include the addition amount of alloy, the addition amount of slag making materials, the turnover cycle of ladle, the weight of molten steel, the starting temperature of molten steel, the refining time, the heating duration time, and the argon consumption.

The quantity of heat can be classified to the heat incomes and the heat outgoings. Based on the actual production, the relationship between the quantity of heat and the main factors affecting the temperature prediction of molten steel is analyzed as follows.

- The heat incomes
 - Heat of electric arc (Q_{arc}): In the heating phase the voltage and the current are set to 200 V and 35 kA, respectively. The Q_{arc} is mainly related to electric-arc heating duration.
 - Chemical reaction heat of additions (Q_{addition}): The weight and composition of molten steel are different in different heats, which lead to the different addition of alloy and slag making materials. Based on the temperature effect coefficients of additions, the Q_{addition} can be calculated according to the amount of alloy and slag making materials.
- The heat outgoings
 - Heat loss from the slag surface (Q_{surface}): The Q_{surface} is mainly influenced by the thickness of slag layer. The thin slag usually promotes the heat loss from the slag surface.¹⁷⁾ Actually, the casting residue will be recycled to the ladle furnace in the refining process, and the slag thickness is larger. Therefore, the effect of different slag thickness on heat loss is ignored. The Q_{surface} is mainly reflected by the refining time.
 - Heat loss by argon stirring (Q_{argon}): The bottom argon blowing will induce the exposure of the melt and result in heat loss of molten steel, and the argon consumption is proportional to the temperature drop of molten steel. Therefore, the Q_{argon} can be reflected by the argon consumption.
 - Heat storage of ladle (Q_{storage}): The ladle itself will absorb and store the heat of the molten steel in the refining process, and the Q_{storage} is mainly influenced by the thermal status of ladle. The thermal status of ladle is reflected by the turnover cycle of ladle in steelmaking process.
 - Heat exchanges between ladle furnace and sur-

roundings (Q_{shell}): The heat is inevitably lost to the surroundings from the ladle furnace due to the temperature difference. The Q_{shell} will increase with time, which can be reflected by the refining time.

Here, the starting temperature is the temperature measured for the first time after the ladle enters the LF station. The end temperature is the temperature measured for the last time before the ladle leaves the LF station.

3. Temperature Calculation of Molten Steel Using Hybrid Model

3.1. Modelling in Hybrid Model

A hybrid model based on metallurgical mechanism, expert control and DNN is established by analyzing the LF refining process of against the background of 150 t LF of a steelmaking plant in China. In this study, 2 000 groups of production data were divided into two categories, of which the 1 500 groups were used to train the hybrid training model, and the left 500 groups were used to test the hybrid model when the error between the predicted value ($T_{\text{predicted}}$) of training model and the T_1 reached the minimum. The T_1 was the end temperature (T_{measured}) of molten steel minus the temperature change ($\Delta T_{\text{addition}}$) caused by addition of alloy and slag making materials. The flowchart of construction for hybrid model is illustrated in **Fig. 3**.

3.1.1. Introduction of Expert Control

The concept of expert control proposed by K. J. Astrom, is to combine the theory and method of expert system with the control theory, and realize the control of the system by imitating the experience of experts under the unknown environment.^{18,19)} Through field investigation, it can be seen that there are many kinds of slag making materials and alloy, and there are complex physical and chemical reactions in LF refining process. Meanwhile, the temperature effect coefficients of various additions were evaluated only by the experience of operators, which led to a large calculation error. Thus, a more effective control, that could give a reasonable quantity of heat generated by slag making materials and alloy when the end temperature calculated, was needed. The idea of expert control can solve the above problems well. In this study, the temperature effect coefficients of additions to molten steel, which was used to calculate the heat generated

Table 1. Main factors affecting the temperature prediction of molten steel.

Variables	Description of variables	Units
X_1	Addition amount of alloy	kg
X_2	Addition amount of slag making materials	kg
X_3	Turnover cycle of ladle	min
X_4	Weight of molten steel	kg
X_5	Starting temperature of molten steel	°C
X_6	Refining time	min
X_7	Heating duration time	s
X_8	Argon consumption	NL

Fig. 3. A flowchart of construction for hybrid model. (Online version in color.)

by slag-making material and alloy, was obtained by using the statistical analysis as shown in **Table 2**. And then, the temperature change of molten steel of various additions to molten steel was then calculated by using the idea of expert control, which combines the strengths of statistical analysis (obtaining the temperature effect coefficients), metallurgical mechanism (thermal equilibrium) and the experience of operators. The temperature change of molten steel owing to additions can be calculated using Eq. (3).

$$\Delta T_{\text{addition}} = \sum_i G_i q_i \quad \dots\dots\dots (3)$$

Where, i designates a specific addition (alloy or slag making materials); G_i is the weight of addition i (kg); q_i is the temperature effect coefficient of i ($^{\circ}\text{C}/\text{kg}$).

3.1.2. Introduction of DNN

Hinton *et al.*²⁰⁾ proposed the concept of deep learning (DL). The network structure of DNN is divided into input layer, hidden layer (multi-layer), output layer. Meanwhile, the layer and layer are directly fully connected.²¹⁾ The DNN is used to predict temperature, the structure diagram of which is shown in **Fig. 4**. The input layer consists of six components, including turnover cycle of ladle, weight of molten steel, starting temperature of molten steel, refining time, heating duration time, and argon consumption. The output layer consists one neuron. The hyperparameter setting of DNN has an important effect on the prediction results, and the role of each hyperparameter will be described in detail below.

(1) Activation Function and Learning Rate

The activation function is a function that runs on the neurons of a neural network and is used to map the neural input to its output.²²⁾ Activation functions commonly used in DNN model include sigmoid (Sig), hyperbolic tangent (Tanh), and rectified linear unit (ReLU). For the DNN, the Sig and Tanh are prone to gradient disappearance, while the derivative of ReLU does not have this problem. Meanwhile, both Sig and Tanh need to calculate the exponential from the perspective of calculation, which is of high computational complexity. The ReLU only needs a threshold to get the activation value, and the rate of learning convergence is six times faster than the Sig or Tanh.²³⁾ However, ReLU units are fragile. When the gradient of input function is too large during training process, the neurons of ReLU no longer have the function to activate after the network parameters are updated, which results in the gradient will forever be zero.²⁴⁾ The leaky rectified linear unit (LReLU) don't have this problem.¹⁴⁾ Thus, the LReLU was selected as the activation function, as shown in Eq. (4). Learning rate is an important hyperparameter in DNN, which determines

the change of weight generated in each training cycle.^{25,26)} When the learning rate is large, the weight changes greatly, which easily leads to the instability of model. When the learning rate is small, the training time is increased and the convergence speed of model is slow.

$$f_{\text{irelu}} = \begin{cases} x & x > 0 \\ 0.01x & x \leq 0 \end{cases} \quad \dots\dots\dots (4)$$

(2) Number of Hidden Layers and Nodes

Both the number of hidden layers and nodes in hidden layer have great influence on the performance of DNN. The number of hidden layers mainly determines the complexity of network structure. The number of hidden layers of DNN commonly has 3 hidden layer (or >3 hidden layers).²⁷⁾ The number of hidden layer nodes mainly determines the learning ability and learning rate of DNN. When the number of hidden layer nodes is small, the learning ability and information processing ability of neural network are weak. When the number of hidden layer nodes is large, the learning rate of network will be slow, and it is easy to make the network fall into the local minimum value, which even leads to overfitting. The optimal number of hidden layer nodes and hidden layers was determined by using trial and error method.

(3) Optimization Algorithms

Gradient descent is one of the most popular optimization algorithms in DNN. The stochastic gradient descent (SGD) with momentum and L_2 regularization was used to optimize the DNN, which not only solves the problems of slow convergence speed and easily getting into local optimum, but also prevents overfitting.^{28–30)} The algorithm flow chart is shown in **Table 3**.

3.2. Analyzing of Database and Data

3.2.1. Data Cleansing

The production data of LF refining process were collected

Fig. 4. The structure diagram of DNN. (Online version in color.)

Table 2. Temperature effect coefficients of various additions to molten steel in 150 t LF.

Addition	Temperature effect coefficient of addition to molten steel $\times 10^{-2}$ ($^{\circ}\text{C}/\text{kg}$)	Addition	Temperature effect coefficient of addition to molten steel $\times 10^{-2}$ ($^{\circ}\text{C}/\text{kg}$)
C	-4.30	FeSi	+1.10
HCFeMn	-1.30	Slag making materials	-2.00
LCFeMn	-1.15	FeNb	-0.32
CaSi	-1.10	FeTi	-0.38

from a steelmaking plant in China. The original data were processed by using boxplot.^{31,32)} The boxplot consists of five statistics, including lower quartile (Q_1), median (Q_2), upper quartile (Q_3), $Q_1-1.5IQR$, and $Q_3+1.5IQR$. The interquartile range (IQR) is defined as Q_3-Q_1 . In boxplot, the outlier data took values below $Q_1-1.5IQR$ or above $Q_3+1.5IQR$, as shown in Fig. 5.

Boxplot is a widely used method to detect outliers in sample data. The advantage of boxplot is that it is not influenced by outliers, so it can describe the discrete distribution of data in a relatively stable way, and it is also conducive to data cleansing. In this study, the original data were processed using boxplot, and the outliers below $Q_1-1.5IQR$ and beyond $Q_3+1.5IQR$ were eliminated to realize the data cleaning. The outliers distribution of original data were shown after the original data was normalized, as shown in Fig. 6.

There are 2 000 groups of production data in total, as the data were obtained based on the boxplot and production operation rules. The data distribution of a single variable and the descriptive statistics for all input and output vari-

Table 3. Algorithm flow chart.

Algorithm 1: SGD with momentum and L_2 regularization ^{29,30)}	
Require: α : learning rate, β : momentum factor (Parameter defaults to 0.9), λ : L_2 regularization factor (Parameter defaults to 1.0 e^{-4}), $f(\theta)$: objective function with parameter θ	
Require: $\theta_{t=0}$	(Initialize parameter vector)
$t \leftarrow 0$	(Initialize timestep)
$m_{t=0}$	(Initialize moment vector)
$\eta_{t=0}$	(Initialize schedule multiplier)
while θ not converged do	
$t \leftarrow t+1$	
$\nabla f_t(\theta_{t-1}) \leftarrow \theta_{t-1}$	(The gradient of θ_{t-1} against loss)
$g_t \leftarrow \nabla f_t(\theta_{t-1}) + \lambda \theta_{t-1}$	(Get gradients objective at timestep t)
$\eta_t \leftarrow$ Set schedule multiplier (t)	
$m_t \leftarrow \beta m_{t-1} + \eta_t g_t$	(Update biased moment estimate)
$\theta_t \leftarrow \theta_{t-1} - m_t - \eta_t \lambda \theta_{t-1}$	(Update parameters)
end while <i>stopping criterion is met</i>	
return θ_t	(Optimized parameters)

Fig. 5. Outlier detection based on boxplot. (Online version in color.)

ables of the prediction models are showed in Fig. 7 and Table 4. The range of the turnover cycle of ladle and the refining time was 28–90 min and 18–55 min, respectively. The value of weight for molten steel ranged from 150 000 to 158 260 kg. The value of starting temperature ranged from 1 518 to 1 603°C. The range of the heating duration time was 155–1 256 s. The range of the argon consumption was from 15 000 to 48 000 NL. The value of the T_1 ranged from 1 576 to 1 612°C.

3.2.2. Correlation Analysis and Data Normalization

The correlation between two random variables was analyzed by using the Pearson correlation coefficient and the Student's t-Test in the statistics analysis, as shown in Eqs. (5) and (6). In Fig. 8, the different impacts of input variables on the T_1 of molten steel are shown in an order from strong to weak as: heating duration time, refining time, argon consumption, starting temperature of molten steel, turnover cycle of ladle and weight of molten steel. The Student's t-Test was mainly used to test the significance of correlation coefficient. When the p -value is less than 0.01, it means that the correlation of variables is very significant. When the p -value of significance probability is less than 0.05, it means that the correlation of variables is significant. When the p -value is greater than 0.05, it means that the correlation of variables is not significant.³³⁾

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} \quad \text{..... (5)}$$

Where, $\bar{x}$ is the mean of variable x ; $\bar{y}$ is the mean of variable y ; x_i is the i th value of variable x ; y_i is the i th value of variable y .

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} \quad \text{..... (6)}$$

Where, r is the correlation coefficient; n is the sample size of variable; $n-2$ is the degree of freedom.

Fig. 6. Outlier detection of boxplot based on original data. (Online version in color.)

 Fig. 7. Scatter plot matrix visualization of X_3 , X_4 , X_5 , X_6 , X_7 , X_8 and T_1 . (Online version in color.)

Table 4. Descriptive statistics of the variables.

Variables	Mean	Minimum	Maximum	Range
Turnover cycle of ladle/min	63	28	90	62
Weight of molten steel $\times 10^3$ /kg	1.5255	1.5000	1.5826	0.08260
Starting temperature $\times 10^3$ /°C	1.555	1.518	1.603	0.0850
Refining time/min	35	18	55	37
Heating duration time/s	693	155	1 256	1 101
Argon consumption $\times 10^4$ /NL	3.0	1.5	4.8	3.3
$T_1 \times 10^3$ /°C	1.593	1.576	1.612	0.0360

In this study, the p -value method was used to test the significance of the correlation between two variables. The results of correlation analysis are shown in Table 5. In Table 5, it can be seen that the p -value is less than 0.01 between T_1 and starting temperature of molten steel, refining time, heating duration time, argon consumption, respectively; the p -value is less than 0.05 between T_1 and turnover cycle of ladle; the p -value is greater than 0.05 between T_1 and weight of molten steel. Taken together, these results indicated that the relationship between starting temperature of molten steel, refining time, heating duration time, argon consumption and T_1 reach very significant correlation, and the relationship between turnover cycle of ladle and T_1 reach significant correlation. However, there is no significant correlation between T_1 and weight of molten steel. Actually, the weight of molten steel is non-negligible, which not only affects the addition of alloy and slag making materials but also affects the refining heating duration and the heat loss.^{6-8,34} Meanwhile, the initial energy of molten steel is

Fig. 8. Correlation analysis results between input variables and output variables. (Online version in color.)

determined by the weight of molten steel. Thus, the weight of molten steel was considered when the hybrid model was established.

In order to improve the accuracy and accelerate the convergence of model, the data processed by using the boxplot and production operation rules were normalized. The mathematical formula of normalization processing is shown in Eq. (7).

$$x^* = \frac{x - \min}{\max - \min} \dots\dots\dots (7)$$

Where, $\max$ is the maximum value of the data; $\min$ is the minimum value of the data.

3.3. Model Evaluation

The performance of hybrid model was evaluated according to different statistical evaluation indexes, including the coefficient of determination (R^2), mean square error (MSE), and root-mean-square error ($RMSE$). The computational formulas of the above indexes are displayed in Eqs. (8)–(10):

$$R^2 = \frac{\sum_{i=1}^{N_p} (y_i^{act} - \bar{y}_m)^2 - \sum_{i=1}^{N_p} (y_i^{pre} - y_i^{act})^2}{\sum_{i=1}^{N_p} (y_i^{act} - \bar{y}_m)^2} \dots\dots\dots (8)$$

$$MSE = \sum_{i=1}^{N_p} (y_i^{pre} - y_i^{act})^2 / N_p \dots\dots\dots (9)$$

$$RMSE = \sqrt{\sum_{i=1}^{N_p} (y_i^{pre} - y_i^{act})^2 / N_p} \dots\dots\dots (10)$$

Where, N_p is the number of the total data set; y is the value of the temperature; $\bar{y}_m$ is the average value of the temperature; the “act” means the actual values; the “pre” means the predicted values.

4. Results and Discussion

4.1. Hyperparameter Optimization of DNN

In order to obtain the optimal DNN model, the trial and error method was used to determine the hyperparameters.³⁵⁾ To construct the optimal architectures of DNN model, the models were trained with varying the number of hidden layers. Considering the problem of training time, the initial hyperparameters of models were fixed as follows: the number of hidden layer neurons (20), the number of iterations (500) and learning rate (0.3). It can be seen from Fig. 9(a) that with the increase of the number of hidden layers, the

Table 5. The calculation results of p -value between T_1 and input variables.

	X_3	X_4	X_5	X_6	X_7	X_8
r	-0.053	-0.043	-0.218	0.380	0.645	0.292
p -value	$1.81 \times 10^{-2*}$	5.69×10^{-2}	$6.07 \times 10^{-23**}$	$1.59 \times 10^{-69**}$	$2.76 \times 10^{-235**}$	$1.22 \times 10^{-40**}$

Note: (*) $p < 0.05$, (**) $p < 0.01$. The sample size is 2 000.

Fig. 9. MSE with variation of (a) the number of hidden layers, (b) the number of hidden layer neurons, (c) iterations, and (d) learning rate for DNN models. Note: The arrows indicate the optimal hyperparameters. (Online version in color.)

MSE first decreased and then increased. When the number of hidden layers is 4, the *MSE* reaches the minimum. Hence, architectures consisting of 4 hidden layers were preferred for the DNN model. Figure 9(b) shows the effect of the number of hidden layer neurons on *MSE* at fixed hidden layer (4), iterations (500), and learning rate (0.3). It is evident from Fig. 9(b) that with the increase of the number of hidden layer neurons, the *MSE* decreased. When the number of hidden layer neurons exceeds 35, the *MSE* has little change. Then, at fixed hidden layer (4), hidden layer neurons (35), and learning rate (0.3), the *MSE* decreased with increasing iterations. When the iterations exceed 3 000, the *MSE* has little change, as shown in Fig. 9(c). Figure 9(d) shows the variation of *MES* with learning rates at fixed hidden layer (4), hidden layer neurons (35), and iterations (3 000). In Fig. 9(d), with the increase of the learning rate, the *MSE* first slightly decreased and then increased. The minimum *MSE* was obtained at learning rate of 0.2. To sum up, the optimal architecture of DNN model corresponds to the hidden layers of 4, hidden layer neurons of 35, iterations of 3 000, and learning rate of 0.2.

4.2. Comparison of the DNN Model with Other Models

Based on the Pearson correlation analysis, the relationship tends to be linear between the T_1 and a part of input parameters, such as turnover cycle of ladle, starting temperature of molten steel, refining time, heating duration time, argon consumption respectively. However, there is a nonlinear feature between the weight of molten steel and the T_1 , which indicated that there are complex relationships (hybrid linear and nonlinear relationship) between input variables and output variables. Meanwhile, in the actual production, the weighing system has an error of more than 10 kg or even 100 kg; for example, the hopper weighing error is ± 10 kg and the crown block weighing error is ± 600 kg. In addition, the production datasets inevitably even contain fault data, abnormal data, or missing data, which led to the potential noisy data in the production dataset even after preprocessing. These are the two most important features of the data for modelling.

In order to verify the prediction accuracy and the generalization performance of the DNN model, the optimal DNN model were compared with multiple linear regression (MLR) model, back propagation (BP) neural network model, and regularized extreme learning machine (RELM) model. The performance evaluation results and prediction results of various models are shown in **Tables 6, 7** and **Fig. 10**. In the MLR model, the R^2 and the *RMSE* are 0.554, 3.556, respectively. In the shallow neural network (SNN) model, the R^2 and the *RMSE* of the BP model and the RELM model are 0.668, 2.995 and 0.725, 2.729, respectively. These results show that the SNN model is better than the MLR model by comparing the R^2 and the *RMSE*. The reason why the SNN model outperforms the MLR model is shown as follows: (1) The MLR is a relationship model between input variables (characteristic variables) and output variables. Since output variables are linear combinations of input variables, the MLR is always linear. The MLR does not fit well when the nonlinear data is processed. (2) All the data information of the SNN is stored in the neural unit, so the SNN has strong robustness. (3) The SNN model can adequately approxi-

mate any complex nonlinear relation. Therefore, the SNN model can get more accurate results when the input data are close to the training data of SNN model. (4) In the MLR models, the original features are directly used for modelling. However, in the SNN models, the original features of the input layer are processed by the hidden layer to obtain the new features, which is similar to the process of data pre-processing. Therefore, the prediction effect of the SNN model is better than that of the MLR model.³⁶⁾

The R^2 and the *RMSE* of the DNN model are 0.897, 1.710, respectively. These results show that the DNN model is better than the SNN model by comparing the R^2 and the *RMSE*. Since DNN not only has the advantages of SNN, but also has the advantages as follows: (1) Stronger learning ability. DNN takes the original data as input and gradually performs feature transformation through multiple hidden layers. Meanwhile, the rules of data are automatically extracted and stored in the network by learning, which can realize a more complex nonlinear mapping relationship.^{37,38)} (2) Better generalization performance. As data structures become more complex and the amount of data increases, DNN has a stronger ability to process the high-dimensional data and express complex functions than SNN. Therefore, the DNN has a better generalization performance for complex problems.²²⁾ (3) Stronger anti-interference ability. In the case of large amount of data and complex data structure, the DNN will not have a great impact on the global training results when there are noisy data.

As shown in Fig. 10 and Table 7, the closer the scatter is to the 45-degree diagonal dotted line, the smaller the error between the T_1 and $T_{\text{predicted}}$ will be. The overall scatter plot of DNN model is closer to the 45-degree diagonal dotted line than that of MRL model, BP model, and RELM model. In the error scope of temperature from -3 to 3°C , the hit ratio of MLR model, BP model, RELM model and DNN model acquired 63.2%, 70.4%, 76.8%, and 91.4%, respectively. In the error scope of temperature from -5 to 5°C , the hit ratio of MLR model, BP model, RELM model and DNN model acquired 85.4%, 90.4%, 93.4%, and 99.4%, respectively. The results show that the DNN model has higher accuracy and stronger generalization performance. Although hybrid model based on expert control and DNN

Table 6. Comparison of various models.

Indices	Various models			
	MLR	BP	RELM	DNN
R^2	0.554	0.668	0.725	0.897
r	0.744	0.817	0.852	0.947
<i>MSE</i>	12.642	8.972	7.446	2.924
<i>RMSE</i>	3.556	2.995	2.729	1.710

Table 7. Hit ratio of various models.

Error scope	Various models			
	MLR	BP	RELM	DNN
From -3 to 3°C	63.2%	70.4%	76.8%	91.4%
From -5 to 5°C	85.4%	90.4%	93.4%	99.4%

Fig. 10. Comparison between T_1 and $T_{\text{predicted}}$ obtained by (a) MLR, (b) BP, (c) RELM, and (d) DNN. The 45-degree diagonal dotted line is the identity where the T_1 is equal to the $T_{\text{predicted}}$. The inset is the distribution of the error scope. (Online version in color.)

has high prediction accuracy, there are still some shortcomings as follows: (1) Lack of theoretical support. Although there are a lot of optimization techniques for DNN models, there is still no relevant theory to support how to tune the hyperparameters, and the values of the hyperparameters are mainly determined by trial and error method. Therefore, the DNN models need to be further studied theoretically.³⁹⁾ (2) Computational efficiency problem. With the increase of data volume, the DNN model has higher prediction accuracy than the SNN model, but the DNN model has many hyperparameters, complex structure, large computing and other characteristics, which often leads to longer training time. Therefore, on the premise of ensuring the computational accuracy of DNN model, the improvement of computational efficiency should also be considered.⁴⁰⁾ (3) Engineering application problem. On the one hand, LF refining is a complicated physical and chemical process at high temperature, and there is a very complex nonlinear relationship between different variables. The complex process cannot be described completely and accurately by using the DNN model. Meanwhile, the expert control should be adjusted and optimized in real time to improve the hit ratio of the hybrid model, according to the change of the production condition, such as the change of the alloy and slag making materials supplier. On the other hand, in the actual production, the original production datasets inevitably contain fault

data, abnormal data, and missing data, which will affect the calculation accuracy of the hybrid model. Therefore, the effective combination of hybrid model and big data mining technology should be considered to improve the data quality and then improve the calculation accuracy. After solving the above problems, the DNN model and the modeling strategy of the hybrid model can also be better applied in other metallurgy problems, which is of significance for the intelligent development of the iron and steel industry.

5. Conclusions

In this study, a hybrid modeling method based on expert control and deep neural network has been proposed for temperature prediction of molten steel in LF and the following conclusions can be drawn.

(1) Based on the boxplot and production operation rules, the original production data were preprocessed to delete outliers. In addition, the correlation analysis between the input and output variables was carried out. The results demonstrated that the different impacts of input variables on the T_1 of molten steel are shown in an order from strong to weak as: the heating duration time (X_7), the refining time (X_6), the argon consumption (X_8), the starting temperature of molten steel (X_5), the turnover cycle of ladle (X_3), and the weight of molten steel (X_4).

(2) In order to obtain the optimal DNN model, the trial and error method was used to determine the hyperparameters. Meanwhile, the *MSE* of DNN model were considered when the DNN model was trained. The results demonstrated that the optimal architecture of DNN model corresponds to the hidden layers of 4, hidden layer neurons of 35, iterations of 3 000, and learning rate of 0.2.

(3) The performance of the optimal DNN model was evaluated according to the statistical evaluation indexes and the distribution of error scope. The results demonstrated that the performance of the DNN model was better than that of the MLR model and SNN model; and the R^2 and *RMSE* of the DNN model were 0.897, 1.710, respectively. Meanwhile, in the error scope from -5 to 5°C , the hit ratio of DNN model was 99.4%. To sum up, the hybrid model has higher accuracy and stronger generalisation performance to predict the temperature in LF refining process.

Acknowledgements

This work was financially supported by the National Natural Science Foundation of China (No. 51974023) and the funding of State Key Laboratory of Advanced Metallurgy, University of Science and Technology Beijing (No. 41621005).

REFERENCES

- 1) K. Feng, D. F. He, A. J. Xu and H. B. Wang: *Steel Res. Int.*, **87** (2016), 79.
- 2) F. He, D. F. He, A. J. Xu, H. B. Wang and N. Y. Tian: *J. Iron Steel Res. Int.*, **21** (2014), 181.
- 3) Y. J. Wu, Z. H. Jiang and M. F. Jiang: *J. Northeast. Univ.*, **23** (2002), 247 (in Chinese).
- 4) N. K. Nath, K. Mandal, A. K. Singh, B. Basu, C. Bhanu, S. Kumar and A. Ghosh: *Ironmaking Steelmaking*, **33** (2006), 140.
- 5) H. X. Tian, Z. Z. Mao and A. N. Wang: *J. Iron Steel Res. Int.*, **16** (2009), 1.
- 6) H. X. Tian, Y. D. Liu, K. Li, R. R. Yang and B. Meng: *ISIJ Int.*, **57** (2017), 841.
- 7) H. X. Tian, Z. Z. Mao and Y. Wang: *ISIJ Int.*, **48** (2008), 58.
- 8) H. X. Tian, Z. Z. Mao and A. N. Wang: *ISIJ Int.*, **49** (2009), 58.
- 9) W. Lü, Z. Z. Mao and P. Yuan: *J. Iron Steel Res. Int.*, **19** (2012), 21.
- 10) H. Y. Tang, X. C. Guo, J. L. Wang, Y. Wang and P. F. Cheng: *Chin. J. Eng.*, **38** (2016), S139 (in Chinese).
- 11) G. Q. Fu, Q. Liu, Z. Wang, J. Chang, B. Wang, F. M. Xie, X. C. Lu and Q. P. Ju: *J. Univ. Sci. Technol. Beijing*, **35** (2013), 948 (in Chinese).
- 12) G. B. Huang, Z. Bai, L. L. C. Kasun and C. M. Vong: *IEEE Comput. Intell. Mag.*, **10** (2015), 18.
- 13) J. P. Yang, J. S. Zhang, W. D. Guo, S. Gao and Q. Liu: *ISIJ Int.*, **61** (2021), 2100.
- 14) S. H. Kwon, D. G. Hong and C. H. Yim: *Ironmaking Steelmaking*, **47** (2020), 1176.
- 15) C. A. Myers and T. Nakagaki: *ISIJ Int.*, **59** (2019), 687.
- 16) Z. C. Xin, J. S. Zhang, J. G. Zhang, Y. Jin, J. Zheng and Q. Liu: *Ironmaking Steelmaking*, **48** (2021), 1123. <http://doi.org/10.1080/03019233.2021.1935143>
- 17) J. Li: *LF Refining Technology*, Metallurgical Industry Press, Beijing, (2012), 135.
- 18) K. J. Åström, J. J. Anton and K. E. Årzén: *Automatica*, **22** (1986), 277.
- 19) S. Y. Li and Y. Li: *Intelligent Control*, Tsinghua University Press, Beijing, (2016), 122.
- 20) G. E. Hinton and R. R. Salakhutdinov: *Science*, **313** (2006), 504.
- 21) S. Koohi and S. Hessabi: *J. Parallel Distrib. Comput.*, **72** (2012), 1493.
- 22) J. Schmidhuber: *Neural Netw.*, **61** (2015), 85.
- 23) Y. D. Zhang, C. C. Pan, J. D. Sun and C. S. Tang: *J. Comput. Sci.*, **28** (2018), 1.
- 24) G. F. Lin and W. Shen: *Procedia Comput. Sci.*, **131** (2018), 977.
- 25) M. A. Ranzato, F. J. Huang, Y. L. Boureau and Y. LeCun: *Proc. 2007 IEEE Conf. on Computer Vision Pattern and Recognition (CVPR 2007)*, (Minneapolis), IEEE, Piscataway, NJ, (2007), 1.
- 26) H. X. Yang, J. H. Liu, H. W. Sun and H. G. Zhang: *IEEE Access*, **8** (2020), 112805.
- 27) S. Feng, H. Y. Zhou and H. B. Dong: *Mater. Des.*, **162** (2019), 300.
- 28) N. Qian: *Neural Netw.*, **12** (1999), 145.
- 29) I. Loshchilov and F. Hutter: *Int. Conf. on Learning Representations (ICLR 2019)*, (New Orleans), ICRL, La Jolla, CA, (2019), 1.
- 30) M. H. Zhao, S. S. Zhong, X. Y. Fu, B. P. Tang and M. Pecht: *IEEE Trans. Ind. Inform.*, **16** (2020), 4681.
- 31) D. R. Cassar, A. C. P. L. F. de Carvalho and E. D. Zanotto: *Acta Mater.*, **159** (2018), 249.
- 32) A. B. Aicha: *Procedia Comput. Sci.*, **126** (2018), 586.
- 33) Z. Sheng, S. Q. Xie and C. Y. Pan: *Probability Theory and Mathematical Statistics*, Higher Education Press, Beijing, (2008), 178.
- 34) F. He, A. J. Xu, H. B. Wang, D. F. He and N. Y. Tian: *Steel Res. Int.*, **83** (2012), 1079.
- 35) S. Samarasinghe: *Neural Networks for Applied Sciences and Engineering*, Auerbach Publications, Boca Raton, FL, (2006), 136.
- 36) Z. C. Xin, J. S. Zhang, W. H. Lin, J. G. Zhang, Y. Jin, J. Zheng, J. F. Cui and Q. Liu: *Ironmaking Steelmaking*, **48** (2021), 275.
- 37) Y. LeCun, Y. Bengio and G. Hinton: *Nature*, **521** (2015), 436.
- 38) Z. H. Zhou: *Machine Learning*, Tsinghua University Press, Beijing, (2016), 114.
- 39) R. Y. Sun: *J. Oper. Res. Soc. China*, **8** (2020), 249.
- 40) J. Nickolls and W. J. Dally: *IEEE Micro*, **30** (2010), 56.