



Attention mechanism-based deep learning for heat load prediction in blast furnace ironmaking process

Hong-Wei Xu¹ · Wei Qin¹ · Yan-Ning Sun¹ · You-Long Lv² · Jie Zhang²

Received: 24 May 2022 / Accepted: 1 March 2023 / Published online: 23 March 2023
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2023

Abstract

Heat load prediction is essential to discover blast furnace (BF) anomalies in time and take measures in advance to reduce erosion in the ironmaking process. However, owing to the redundancy of the high dimensional data and the multi-granularity features of the state monitoring data, the general prediction model is hard to accurately predict the heat load, especially the rapid change caused by physical and chemical reactions. Therefore, this paper puts forward an attention-based one-dimension convolution neural network (1DCNN) combined with a bidirectional long short-term memory (BiLSTM) network for heat load prediction. Firstly, the two-stage data pre-processing realizes dimension reduction and key variable selection. Secondly, fine-grained features are extracted by 1DCNN, and the BiLSTM extracts the coarse-grained features for prediction output. Moreover, an attention branch is added to the 1DCNN to extract the critical fine-grained features when the heat load changes rapidly. Finally, experiments are carried out with the actual industrial data from a BF ironmaking process. The efforts show that the proposed prediction model presents better performances in the result of different metrics and has higher accuracy than the traditional prediction algorithms.

Keywords Heat load · Blast furnace ironmaking · Attention mechanism · One-dimension convolutional neural network (1DCNN) · Bidirectional long short-term memory (BiLSTM) network

Introduction

Blast furnace (BF) ironmaking is a typical process manufacturing with complex mechanisms and high energy consumption (Pan et al. 2018), which is mainly composed of the BF body and the auxiliary supporting systems (Wang et al.

2015), as shown in Fig. 1. The feeding system is to feed the raw materials to the top of the BF body, including iron ore, coke, etc. The coal injection system continuously injects pulverized coal into the BF body. The hot blast system constantly supplies the BF body with high-temperature hot air above 1000 degrees. The steel tapping system produces the final iron and slag from the bottom of the BF body (Li et al. 2022). As iron ore and coke intermittently enter the top of the BF body, hot air is continuously bubbled into the bottom of the BF body, and the molten iron and slag intermittently exit the bottom of the BF body (Li et al. 2021). In addition, waste heat and gases are collected and treated by the top gas treatment system and the top gas recovery turbine (TRT) system. Therefore, it is crucial to ensure continuous and stable production by monitoring the condition of the complex BF ironmaking system in real-time and effectively.

The BF heat load refers to the heat carried away by cooling equipment per unit area of time (Zhou et al. 2017). It is a critical index for monitoring the BF body's state. The total heat load of the BF is composed of 7 parts: the furnace bottom, furnace cylinder, furnace waist, furnace belly, the lower part

✉ Wei Qin
wqin@sjtu.edu.cn
Hong-Wei Xu
hongwei_xu@sjtu.edu.cn
Yan-Ning Sun
sunyaning@sjtu.edu.cn
You-Long Lv
lvyoulong@dhu.edu.cn
Jie Zhang
mezhangjie@dhu.edu.cn

¹ School of Mechanical Engineering, Shanghai Jiao Tong University, Shanghai 200240, China

² Institute of Artificial Intelligence, Dong Hua University, Shanghai 201620, China

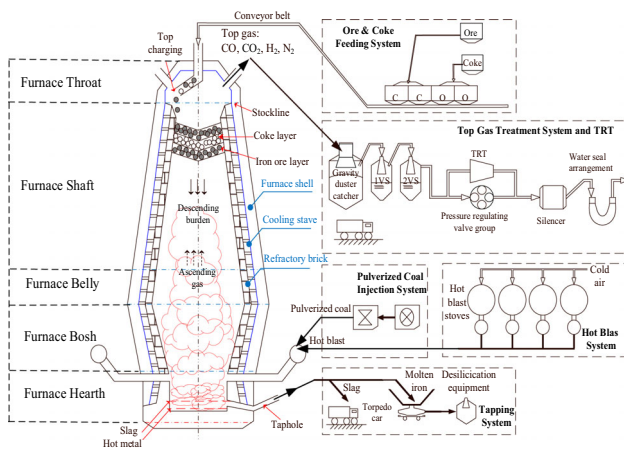


Fig. 1 Blast furnace body and its auxiliary system. The BF body can be roughly divided into five zones: throat, shaft, belly, bosh, and hearth, and the auxiliary supporting systems mainly include the ore and coke feeding system, the top gas treatment system and the top gas recovery turbine (TRT) system, the pulverized coal injection system, the hot blast system, and the steel tapping system

of the furnace body, the middle part of the furnace body, and the upper part of the furnace body, that is, the total heat load value is accumulated from the heat load of these seven parts. However, the traditional method of measuring heat load by heat transfer and water flow has significant hysteresis (Zhou et al. 2018), and it is difficult to grasp the changing trend of the heat load in time. Furthermore, in the actual ironmaking process, when the temperature and pressure in the BF body fluctuate violently, the heat load change will increase exponentially in the short temporal (Semenov et al. 2017). But the current heat load monitoring methods can not acquire the rapid change in time (Wang et al. 2018a, 2018b). That is, there are problems such as unmeasured, inaccurate, and measurement delay. Once the heat load exceeds the controllable range, it will bring in BF erosion and even cause accidents (Kurunov et al. 2006). Therefore, it is essential to predict the heat load and take cooling measures in time to ensure the stability of the BF body.

In the existing literature, there is little research on the heat load prediction of BF systems, and more attention is paid to small-scale boiler systems such as utility boilers and solar boilers (Xing et al. 2017). Traditional heat load prediction methods are mainly based on mechanism analysis. Taler et al., (2019) proposed a mathematical model of a supercritical power boiler for simulating rapid changes in boiler heat load. In addition, to solve the problem of boiler tube failure caused by the heat load difference, Xu et al., (2000) proposed a heat load monitoring model based on thermal variables, thermal deviation theory, and flow deviation theory of power plants. The above modeling and analysis methods have achieved excellent boiler analysis and optimization scenarios. However, these methods mainly rely on empirical

knowledge and assumptions. Their performance is usually demonstrated by post-inspection, which makes it challenging to obtain the state of the BF body in advance.

As an alternative method in recent years, automatic data-driven prediction methods have been used to extract hidden features from complex industrial data, mainly using machine learning and neural network models to predict the heat load. Dong et al., (2011) proposed a modified Takagi-Sugeno-type neuro-fuzzy system, realizing the online identification of thermal processes. In order to further improve the accuracy of heat load prediction in time-varying and uncertain environments, Xie, (2017) proposed a Back Propagation Neural Netp-Markov-based prediction model. In addition, scholars are also concerned with the simultaneous prediction of multiple indicators. Yukun et al., (2011) proposed a hybrid prediction model based on the artificial neural network (ANN), which realized the prediction of furnace temperature, NOx emissions, and other variables simultaneously. For the modeling and optimization of the heat load prediction, Zhao and Wang, (2009) proposed a hybrid model based on support vector regression (SVR) and improved the center particle swarm optimization (ICPSO) algorithm, which can improve the accuracy of heat load prediction. Moreover, to enhance the stability and robustness of heat load prediction, Hu et al., (2017) and Li et al., (2013) proposed ANN-based heat load prediction models, which are based on generalized regression neural network (GRNN) and fruit fly optimization algorithm (FOA).

In conclusion, the data-driven methods represented by the neural network models have been widely used in the prediction model. Optimization methods are also employed to improve the prediction model's accuracy and robustness. However, most of them can not accurately predict the heat load due to the complicated and changeable characteristics of the BF ironmaking process. The main reasons are as follows. First of all, there are many variables related to the heat load, including both temporal and non-temporal series variables, which results in high dimensionality and redundancy in the multiple data. Secondly, the diverse multi-granularity features (Liao et al., 2018; Lv et al., 2021; Zhao et al., 2020) in the state monitoring data. It mainly includes coarse-grained features of different measurement objects, such as temperature, pressure, load, etc. It also contains the fine-grained features of the same measurement object at different positions, such as throat temperature, shaft temperature, belly temperature, and temperature of other parts. The shallow neural network model can not effectively extract these features. Last but not least, most of the existing methods ignore the characteristics of the rapid change of heat load in a short period, resulting in insufficient accuracy of rapid change heat load prediction.

Therefore, we propose a novel data-driven framework for BF heat load prediction to address these three challenges

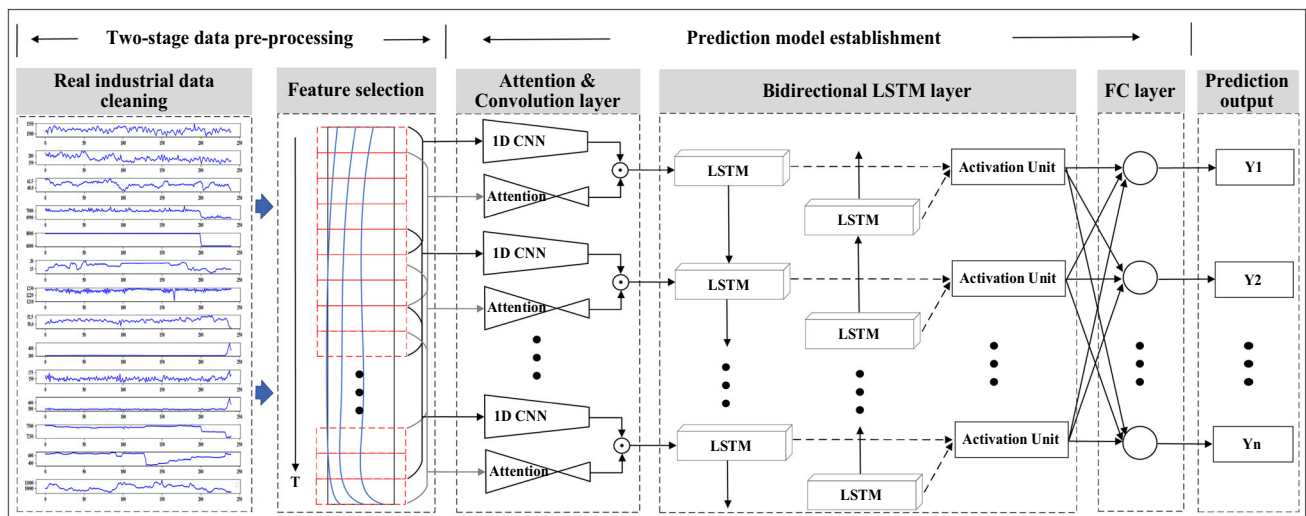


Fig. 2 The framework of attention mechanism-based deep learning for heat load prediction

mentioned above, and an attention-based one-dimension convolution neural network (1DCNN) combined with a bidirectional long short-term memory (BiLSTM) network is designed. The heat load prediction framework is shown in Fig. 2 and consists of two main parts: two-stage data pre-processing and the prediction model establishment. The main processes and contributions are shown as follows:

1. A two-stage data pre-processing model combining data cleaning and mutual information-based feature selection model is designed to solve multivariate BF ironmaking data's high dimensionality and redundancy, which realizes data dimension reduction and key variable selection.
2. We consider both coarse-grained and fine-grained features in multivariate data to realize the output prediction. We use the 1DCNN model to extract fine-grained features, and the coarse-grained features are extracted by the BiLSTM for prediction output under the designed fusion training mechanism.
3. In order to extract the rapid change characteristics of heat load caused by the violent oxidation–reduction reaction in the BF body, an attention mechanism is added to the 1DCNN model to extract the local features of the prominent fluctuation characteristics.

The rest of this article is organized as follows. Section "Related works" presents the related works, followed by the Two-stage data pre-processing in section "Two-stage data pre-processing". In Section "Prediction model", the composition and fusion strategy of the prediction model is introduced. Then the validity and accuracy experiments of the prediction model are evaluated in section "Industrial data-based experiments", and in section "Conclusion", the conclusion and future work are presented.

Related works

1DCNN model

Convolutional neural network (CNN) is a successful deep learning architecture first proposed by (Lecun et al. 1998) in artificial intelligence, which has strong feature extraction ability and high robustness. It has been widely applied in the fields of pattern recognition (Wang et al. 2020), image processing (Zhang et al. 2020), and natural language processing (Wang et al. 2021). However, the 1DCNN model is more suitable for temporal series feature extraction (Eren et al. 2019). The difference between the two models is that the 1DCNN model only convolutes on one dimension, while the two dimensions Convolutional neural network (2DCNN) model is convoluted on two dimensions simultaneously, as shown in Fig. 3. The brown box is the convolution kernel, which is the detector for feature extraction, the direction of the arrow is the direction of movement of the convolution kernel, and

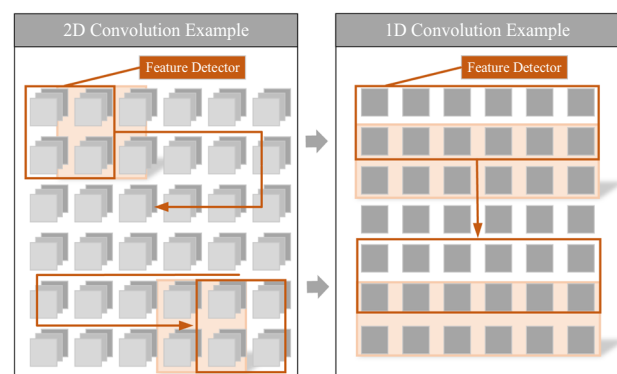


Fig. 3 Difference between 2 and 1DCNN

a one-way convolution kernel extracts the sequence features. Because the direction of movement is one-dimensional, it moves in one direction. It does not move back and forth as it does in multiple dimensions. Therefore, in the 1DCNN layer, the input data is convoluted by the activation function before flowing to the next layer refers to (1):

$$h_t = \sigma_{cnn}(W_{cnn} * x_t + b_{cnn}) \quad (1)$$

where W_{cnn} represents the weight coefficient of the filter, namely the convolution kernel; x_t represents the t input sample data information; $*$ represents the discrete convolution operation between x_t and W_{cnn} ; b_{cnn} is a bias parameter, which is obtained mainly through learning the model during training. $\sigma_{cnn}(\cdot)$ represents the activation function; h_t represents the output data at the end of the convolution operation.

The convolution operation establishes a mapping relationship between the layers. However, after convolution, the dimensionality of the data becomes higher and higher, and the feature map does not change much. After several successive convolutions, a large number of parameters will be generated, which will greatly increase the difficulty of network training and easily cause the phenomenon of overfitting. Therefore, to further reduce the computational effort. Pooling layers are used to reduce the dimensionality of the convolutional features and reduce the risk of network overfitting. Its calculation process refers to (2):

$$h_t = \text{down}(h_{t-1}) \quad (2)$$

where h_{t-1} and h_t represent the eigenvalues before and after pooling, and $\text{down}()$ represents the pooling function.

To obtain a larger receptive field, the 1DCNN model can use a wider convolution kernel to get the eigenvalues of the sequence more comprehensively and then realize feature extraction by stacking convolution and pooling layers.

Attention mechanism

The attention mechanism was proposed based on human vision research, introduced in computer vision and natural language processing to optimize the existing models and focus on the most practical information with limited resources (Li et al. 2017). The attention mechanism is mainly divided into hard attention and soft attention (Zhong et al. 2018). The hard attention mechanism filters the area of interest as input, effectively focusing on the target after removing meaningless background data in the image recognition. But the hard attention mechanisms are not fully applicable in temporal series prediction for the direct limits input (Chen et al. 2020).

Comparatively speaking, the soft attention mechanism uses the weights trained by the neural network to measure the

global input features in space or channel, which can achieve attention on specific space areas or channels (Hou et al. 2021). At the same time, this method can also perform differential operations in reverse computation. Therefore, it can use the end-to-end learning method to train the attention network directly.

In general, a key-value pair can be used to represent an input message so that N input messages can be described as $(K, V) = [(k_1, v_1), \dots, (k_i, v_i), \dots, (k_N, v_N)]$, where "keys" are used to calculate the attention distribution α_i and "values" are used to calculate the aggregated information. The attention mechanism is usually thought of as a soft addressing operation. Think of the input information as the content stored in the memory. The elements consist of the address key and the value. The value corresponding to each address key is extracted from the content and then summed. The weight of each value is first calculated using a similarity criterion, and then the values are weighted and summed. The weighted sum is then applied to the values to obtain the final value, which is the attention value.

$$\begin{aligned} \text{Att}((K, V), q) &= \sum_{i=1}^N \alpha_i v_i \\ &= \sum_{i=1}^N \frac{\exp(s(k_i, q))}{\sum_j \exp(s(k_j, q))} v_i \end{aligned} \quad (3)$$

where $s(k_i, q)$ represents the attention score calculated based on the dot product model, α_i represents the result of the numerical transformation of the attention score by the *softmax* function.

LSTM model

Long short-term memory (LSTM) model is a variant of recurrent neural networks (RNN), composed of input, hidden, and output layers (Tian et al. 2020). Based on the RNN model, the LSTM model adds three gate structures: input gate, forgetting gate, and output gate. However, the cell state is the core of the LSTM model, and the cell state transmits relevant information along with temporal series and updates the state through the three gate structures. The typical structure of the LSTM neural unit is shown in Fig. 4. The forgetting gate selectively forgets the input of the previous node refers

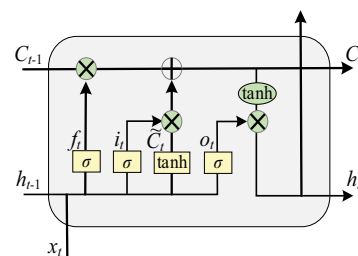


Fig. 4 The basic unit of the LSTM

to (4):

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \tag{4}$$

where h_{t-1} represents the hidden layer information of the previous moment, x_t represents the current input, $\sigma(\cdot)$ is the sigmoid function, and W_f and b_f are training parameters. The input gate selectively remembers the input at this stage, and the result of the forgetting gate determines the cell state update refers to (5):

$$\begin{cases} i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \\ \tilde{C}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \\ C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \end{cases} \tag{5}$$

where C_{t-1} represents the cell state at the last moment, \tilde{C}_t represents the new candidate value vector, C_t represents the current cell state, W_i, b_i, W_c and b_c are training parameters.

The output gate determines the output of the current state refers to (6):

$$\begin{cases} o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \\ h_t = o_t * \tanh(C_t) \end{cases} \tag{6}$$

where o_t represents the operation result of the output gate, W_o represents the weight, b_o represents the bias, $\tanh(\cdot)$ is the activation function, $*$ represents the inner vector product.

Two-stage data pre-processing

In this section, data pre-processing is carried out for the problems in the real data set. The first stage is data cleaning, and a feature selection model based on mutual information is designed to select the key variables and reduce feature dimensions in the second stage.

Data cleaning

Some collected data have missing values in the BF ironmaking process due to the abnormal collection equipment and the high ambient temperature. The next-minute values filling method (Deng & Wang 2017) fills in the missing values. In addition, considering the dimensional differences between the BF ironmaking temperature, feed rate, coal ratio, and other variables, the collected BF ironmaking state variables need to be standardized (Wang et al. 2019). Therefore, the maximum and minimum normalization method is used to standardize the related variables refers to (7):

$$x_i^* = \frac{x_i - x_{min}}{x_{max} - x_{min}} \tag{7}$$

where x_i and x_i^* are the variables before and after normalization, x_{max} and x_{min} are the maximum and minimum variables before normalization respectively. The original data are quantized to 0 and 1 through linearization.

Feature selection

BF ironmaking is a complicated industrial process with an intense change of material and energy. The collected BF ironmaking state data contain different variables, including temporal and non-temporal series-related variables, such as temperature, feed rate, coal ratio, coke ratio, etc. In addition, there are significant linear and nonlinear relationships between some variables, and some of the variables with multi-location measurement, resulting in a certain degree of redundancy (Lv et al. 2020), so it is difficult to identify the variables related to heat load accurately. Moreover, taking all variables as inputs causes the complexity of model training and reduces the model prediction’s response time. It is not conducive to the requirement of real-time monitoring of the BF body’s state. Therefore, the feature selection method (Xu et al. 2020) needs to be used to identify the key variables affecting the heat load and take these critical variables as the input of the prediction model.

Mutual information is an index to measure the degree of interdependence between two variables (Qin et al. 2020). It can measure the relationships between BF ironmaking process variables and heat load. The mutual information of the continuous random variable refers to (8) and is used to analyze the nonlinear relationship between each variable and the heat load.

$$I_c(X_i; Y) = \int_Y \int_{X_i} p(x_i, y) \log\left(\frac{p(x_i, y)}{p(x_i)p(y)}\right) dx dy \tag{8}$$

where $p(x_i, y)$ represents the joint probability density function of the current variable X_i and heat load Y , while $p(x_i)$ and $p(y)$ represents the edge probability density function of the current variable X_i and heat load Y respectively. The nonlinear relationships between each variable and the heat load are obtained through mutual information (Xu et al. 2022). Then the key variables closely related to the heat load are sorted in reverse order.

Prediction model

In this section, a heat load prediction model combining an attention-based 1DCNN with a BiLSTM network is designed. An attention branch is added to the standard 1DCNN network to extract fine-grained features of the local rapid change heat load. Coarse-grained features of latent temporal sequence are extracted by BiLSTM and realize

multi-model fusion prediction under the full connection layer.

Attention-1DCNN-based feature extraction

In the BF ironmaking process, the heat load changes sharply. The features of long temporal series often contain more information due to the multi-granularity features in short sub-series, which influence the changing trend of actual heat load. Therefore, the soft attention mechanism is employed in the 1DCNN model, weighing all input features one by one and focusing on specific spaces and channels to extract significant fine-grained features of temporal series. The structure of the 1DCNN unit based on the attention mechanism is shown in Fig. 5, which mainly includes three parts: temporal feature extraction, parallel module, and feature fusion.

Temporal feature extraction Different segmentation scales are designed for the 1DCNN model in the original temporal series data, and an attention mechanism is designed to extract the multi-granularity features of temporal series data. Each 1DCNN module sets the input subsequence to s . Accordingly, in the attention mechanism module, the input subsequence is set to sa . In addition, by setting $sa < s$, the sensory field of the attention mechanism is made to focus on the local area, which can better grasp the process of rapid changes in heat load compared to the sensory field of 1DCNN. Therefore, it

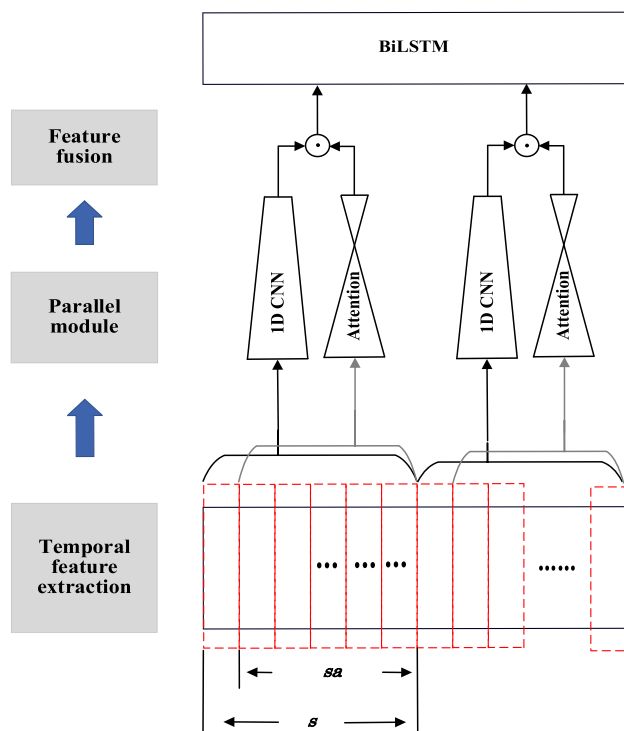


Fig. 5 The structure of attention mechanism-based 1DCNN model

can accurately obtain the salient features with sharp transitions. The features extracted by the 1DCNN model and the attentional mechanism are calculated to (9) and (10).

$$e = \sigma(W_{attention} \cdot sa + B_{attention}) \quad (9)$$

$$x = \sigma(W_{1dcnn} \cdot s + B_{1dcnn}) \quad (10)$$

where $W_{attention}$ and $B_{attention}$ denote the weight and bias of the attention mechanism, W_{1dcnn} and B_{1dcnn} represent the weight and bias of the 1DCNN model.

Parallel module The parallel module refers to adding a parallel soft attention branch to the standard 1DCNN to extract trend features of significant changes in heat load. The attention branch is designed with a smaller input scale than the 1DCNN to focus on the perceptual field of the input, resulting in complete access to temporal contextual information and learning the importance of local sequence features. The attention module reduces the influence of non-important features in the final model by enhancing the impact of critical temporal features in the final model. This effectively responds to the model's inability to distinguish the variability in the importance of temporal features. Therefore, more abundant short sequence features are extracted through multi-scale information to prevent input scale deviation of the single-scale model, which increases the corresponding matching features of the prediction model.

Feature fusion The output features of the 1DCNN module and the corresponding attention mechanism module output significance features are fused by multiplying the feature elements one by one in the existing literature (Hyndman & Koehler 2006; Wang et al. 2017). Precisely, the importance of the features extracted by the 1DCNN model is measured by parameter λ . If the output features of 1DCNN are more critical, the corresponding attention mechanism module is closer to 0. Conversely, the less essential the output features of 1DCNN are, the corresponding attention mechanism module is closer to 1. In general, the features extracted by the default parallel module are equally important, i.e., $\lambda = 0.5$. Therefore, parameter λ reflects the importance of features to identify essential features accurately. The feature fusion of the parallel module is shown in (11).

$$\tilde{x} = \lambda e \cdot (1 - \lambda)x \quad (11)$$

where e and x represent the attention feature and the convolution feature in the parallel module respectively.

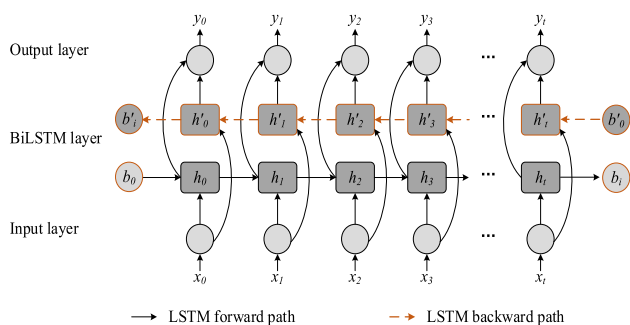


Fig. 6 Structure of BiLSTM unit

BiLSTM-based prediction output

After feature extraction based on the attention mechanism, features associated with changes in heat load trends are obtained, and this feature information is used as input parameters for subsequent prediction models. BiLSTM predicts the output based on the entire temporal series, including the LSTM network in positive and negative directions (Ribeiro et al. 2019). The training sequences are two different LSTM networks. Both are connected to an output layer, which provides complete past and future context information for each point in the input sequence of the output layer. The network weights are updated during the training process through forward and backpropagation of the output neurons (Wang et al. 2018a, 2018b; Zhang et al. 2018). The structure of the BiLSTM neural unit is shown in Fig. 6.

In the structure of the bidirectional recurrent neural network, the forward hiding state and the reverse hiding state of the temporal step are set as h_t and h'_t respectively, and the forward hiding state and the reverse hiding state can be calculated respectively, refers to (12):

$$\begin{cases} h_t = \sigma(W \cdot x_t + W \cdot h_{t-1} + b_h) \\ h'_t = \sigma(W' \cdot x_t + W' \cdot h_{t+1} + b'_h) \end{cases} \quad (12)$$

where x_t and h_{t-1} represent state input and hidden layer output at the hidden forward state, x_t and h_{t+1} represent current state input and hidden layer output at the reverse hiding state, b_h and b'_h represent bias at different levels. Furthermore, connect the hidden states h_t and h'_t in both directions to obtain the whole hidden state H_t , and input it to the output layer. The output layer calculates the output y_t refers to (13):

$$y_t = \tanh(H_t \cdot W + b) \quad (13)$$

where $\tanh(\cdot)$ is the activation function, W and b are the model parameters of the output layer.

Algorithm 1: Attention-based deep learning for heat load prediction

Input: Dataset $D=(x_{(t)}, y_{(t)}), t=1, 2, \dots, k$ with $x_{(t)} = \{x_{(t)}^1, x_{(t)}^2, \dots, x_{(t)}^n\}$; Learning rate η ; Hidden neurons N ; Batch size B ; Training epochs E ; Window size T
Output: Predicted heat load variable $\hat{y}_{(t)}, t=1, 2, \dots, k$

```

1  Steps:
2  Standardized Dataset D ;
3  for t ← 1 to k do
4      Calculate the MI I(t) between x(t) and y(t) respectively ;
5      If I(t) > 0.3 ;
6          Then retain relevant features;
7  Random Initialization of Network Parameters;
8  for t ← 1 to k do
9      Attention mechanism-based IDCNN:
10     for j ← 1 to T do
11         calculate soft attention
12         e(t-j+1) ← σ(W_attention · sα + B_attention);
13         IDCNN-based feature extraction
14         x(t-j+1) ← σ(W_idcnn · s + B_idcnn);
15         feature fusion
16         x̃(t-j+1) ← λe(t-j+1) · (1 - λ)x(t-j+1);
17         feature output
18         h'(t-j+1) ← tanh(x̃(t-j+1), h(t-j));
19     BiLSTM:
20     for k ← 1 to T do
21         BiLSTM -based joint training
22         h_t ← σ(W · x_t + W · h'(t-k+1) + b_h);
23     context vector c(t) ← β(t)h(t) + β'(t)h'(t-1) + ... + β''(t)h'(t-T+1);
24     decoder hidden states s(t) ← f_bilstm(c(t), s(t-1), y(t-1));
25     predicted quality variable ŷ(t) ← f_linear(s(t));

```

Joint training mechanism

The fusion output of the IDCNN model and attention mechanism is taken as the input of nodes in the BiLSTM unit. BiLSTM performs serial modeling to realize the fused prediction of features on the long-period and short-period sequences, and the final heat load prediction result is obtained. The fusion strategy can achieve the coarse-grained features by considering the long-period coarse-grained features in temporal series and the sharp changes in short-period fine-grained features. Algorithm 1 gives the details of the proposed attention-based deep learning method. Firstly, the raw data set needs to be normalized to obtain a standard data format for model training and testing. A mutual information-based feature selection follows this. Each variable is computed concerning the heat load to measure their correlation, and parameters with correlation values greater than 0.3 are retained. These two steps allow the key parameters that affect the heat load and the attention mechanism-based heat load prediction model to be further initialized. The feature extraction and prediction modeling are achieved through attention mechanism-based IDCNN and BiLSTM models. Then, BPNN is used to compute the gradient of parameters, and the Adam algorithm is used to update the network parameters.

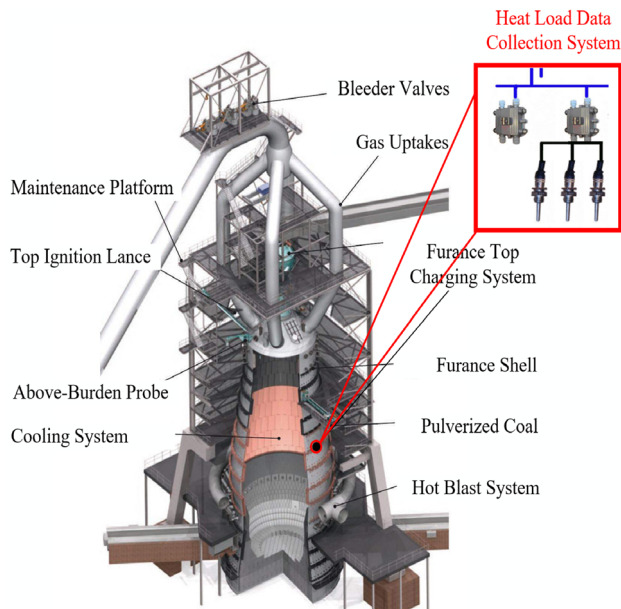


Fig. 7 Schematic diagram of BF and heat load data collection system

Industrial data-based experiments

In this section, the actual industry data and the comparative ablation experiments are introduced first, and then the feature selection experiment is analyzed. Finally, the validity of the proposed model is examined by comparing the prediction results of different models through experiments. The comparative prediction methods include Multi-Layer Perceptron (MLP), LSTM, and 1DCNN-BiLSTM models in these experiments, and these are the more practical and novel models in the existing literature.

Data description and experiment setting

The proposed model has been validated by experiments on the No.4 BF of a large iron and steel plant in Shanghai. The BF has a production capacity of 4 million tons per year, equivalent to a pig iron output of 10,000 tons per day, with the most significant internal volume in operation in China. Figure 7 is a schematic diagram of this real BF ironmaking system and its heat load data collection system (Li et al. 2021). We collected 26 features from the No. 4 BF in the company. After sifting and eliminating some vacant or invalid data, we finally collated 14 features as shown in Table 1. The sampling collection frequency is one hour, including six months of production data. Table 1 lists the multi-granularity features measured directly, including fine-grained features of the same temperature type but distributed in different parts and coarse-grained features of varying measurement objects. Based on the actual ironmaking process data, the mutual information-based feature selection model is designed, and the statistical results of

the prediction model under each evaluation metric are further analyzed. Finally, the predicted and the actual heat load in the next three days are compared and analyzed.

The accuracy of the prediction model is evaluated by comparing the predicted heat load with the actual data. We use the macro averaging criterion to assess the proposed prediction model. They are mean absolute error (MAE), root means square error (RMSE), mean absolute percentage error (MAPE), and maximum absolute error (MAX Error). The four evaluation metrics are as follows:

$$MAE(y, \hat{y}) = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (14)$$

$$RMSE(y, \hat{y}) = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (15)$$

$$MAPE(y, \hat{y}) = \frac{100\%}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \quad (16)$$

$$MAXError(y, \hat{y}) = MAX(|y_i - \hat{y}_i|) \quad (17)$$

where, y_i represents the actual data, \hat{y}_i represents the predicted heat load.

Ablation comparative experiments

The attention-based 1DCNN-BiLSTM proposed in this paper was compared with the current mainstream deep learning methods and temporal-series models: MLP, LSTM, and 1DCNN-BiLSTM models. Then tenfold cross-validation of the proposed attention-based 1DCNN-BiLSTM model is performed. The optimal case of loss function evaluation parameters is finally selected as the final model hyperparameters, as shown in Table 2. we use parameter a to equalize the length of 1DCNN and attention sequences lengths. We have designed three alternative values for parameter a , which are 0.25, 0.5, 0.75. Finally, the parameter tests have revealed that the best results are achieved when $a = 0.25$ identifies sharp heat load changes. Thus, the sequence length CNN_s = 24 for 1DCNN feature extraction and the sequence length Attention_sa = 6 for feature extraction of the attention branch. In the BF ironmaking process, timely and effective acquisition of the heat load change trend is critical for taking preventive measures in advance and ensuring the safety of the BF body. In this case, the fast and stable convergence of the prediction model is vital to the application deployment of the algorithm. Therefore, the convergence speed of the proposed model is evaluated and compared with the MLP, LSTM, and 1DCNN-BiLSTM models. The comparison of loss functions in the validation set is shown in Fig. 8. It can be seen from the figure that the MLP model without

Table 1 Input and output variables using for the prediction model

	Variables	Minimum	Mean	Maximum	Units
1	Temperature				
	Throat	467.60	499.91	532.21	[°C]
	Shaft	486.21	649.36	812.51	[°C]
	Belly	1487.40	1518.46	1549.51	[°C]
	Bosh	1623.25	1647.79	1672.32	[°C]
	Hearth	1715.21	1856.78	1998.34	[°C]
2	Feeding speed	141.03	181.97	234.39	[m ³ /min]
3	Pulverized coal injection	58.71	61.33	63.51	[m ³ /min]
4	Blast capacity	6885.13	6982.23	7061.73	[m ³ /h]
5	Oxygen consumption	5947.81	7636.40	7983.07	[m ³ /h]
6	Humidity	12.08	16.87	20.10	[g/m ³]
7	Blast temperature	1211.19	1226.83	1229.77	[°C]
8	CO use efficiency	48.67	51.28	52.97	[%]
9	Focal ratio	305.93	309.91	466.88	[m ³ /h]
10	Coal ratio	131.68	148.96	184.35	[m ³ /h]
11	Fuel ratio	473.30	491.60	684.99	[m ³ /h]
12	Fe content	7217.97	7458.14	7517.86	[m ³ /h]
13	FeO content	351.14	578.91	679.69	[m ³ /h]
14	Total heat load	9.28259	10.1392	11.46821	[GJ/h]

Table 2 Hyperparameters setting of the attention-based 1dcnn- bilstm model

NO	Hyperparameters description	Value or method
1	CNN_filters_1	112
2	CNN_filters_2	64
3	CNN_kernel_size	2
4	CNN_padding	1
5	CNN_s	24
6	Attention_sa	6
7	BiLSTM_neurons	20
8	BiLSTM_time_steps	3
9	BiLSTM_features	14
10	Dropout	0.2
11	Dense_neurons	1
12	Dense_activation	linear
13	Loss_function	mse
14	Optimizer	adam

temporal-series data processing ability has a slow convergence speed during the training process. The LSTM and 1DCNN-BiLSTM temporal-series prediction models take more time to achieve the final stable convergence under the validation set. However, the proposed attention-based 1DCNN-BiLSTM algorithm has a faster convergence speed than other prediction algorithms, which meets the requirements of real-time online prediction.

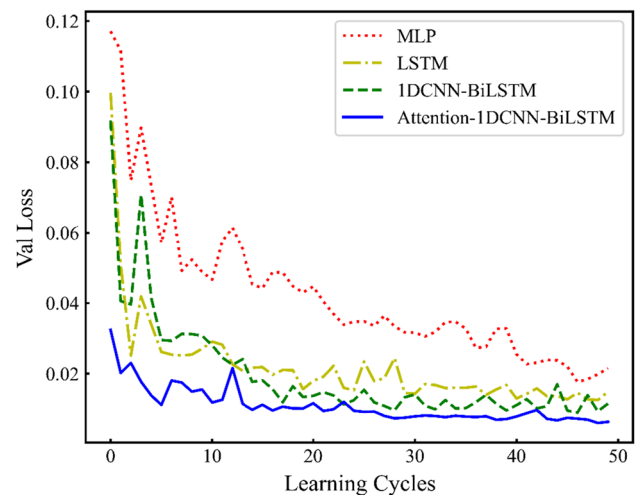


Fig. 8 The comparison of loss functions in validation set

Feature selection results

In this section, the feature selection method is applied by referring to the technique in existing literature (Xu et al. 2020), and mutual information selects the critical influencing factors. Finally, these key factors are then used as inputs to the prediction model, providing necessary reliable data for timely and accurate grasping of the changing trend of heat load and elevating the prediction model’s response time.

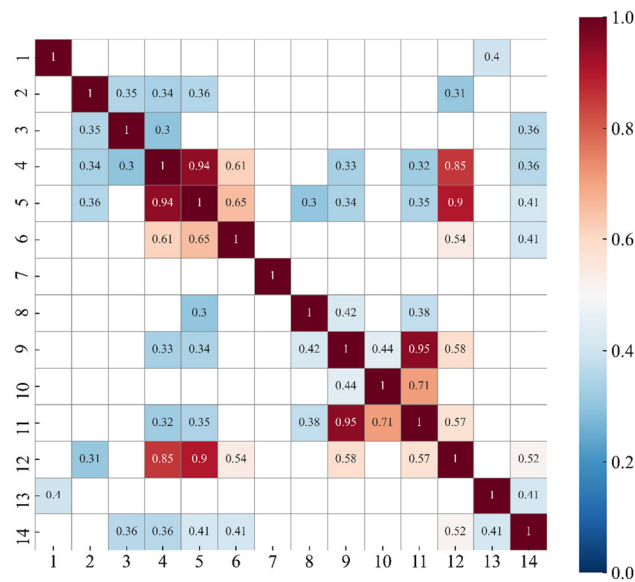


Fig. 9 The heat map of mutual information based relationships

By calculating the nonlinear correlation between variables through mutual information, the heat map of the relationships can be obtained, as shown in Fig. 9. However, based on the description by Xu et al., (2020), if the correlation coefficient between the two variables is less than 0.3, between the variables is weak and is usually considered irrelevant. Therefore, we finally select the six key variables the correlation from the 14 input variables: 3-pulverized coal injection, 4-blast capacity, 5-oxygen consumption, 6-humidity, 12-Fe content, and 13-FeO content.

Metrics evaluation of different algorithms

During the process of the statistical results, each model was run 30 times separately, then recorded and drew a box plot to reduce the contingency of a single prediction experiment. The results are shown in Fig. 10. It can be seen from the plot that the MLP model performs the worst in each evaluation indicator. The fluctuation of the predictive effect in each metric is the most prominent. Compared with the MLP model, the prediction results of the LSTM model decreased in all indicators, indicating that the model with time-series feature processing capability is more suitable for characterizing the trend of heat load. However, further improvement of the LSTM model by adding the 1DCNN model led to a decrease in the mean value of each metric of the prediction model. Still, the magnitude of the abnormal fluctuations of each metric was greater than that of the LSTM model.

Further analysis of each index's box plot result values is shown in Table 3. Although the mean value of the improved index decreased, the maximum and minimum values of the

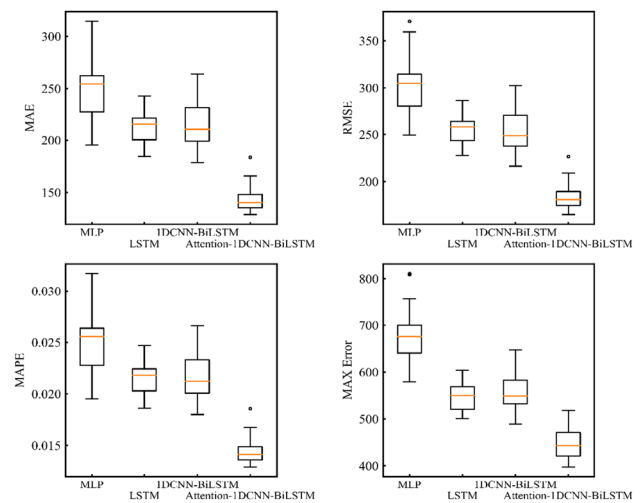


Fig. 10 The performance of the four prediction algorithms in different evaluation metrics

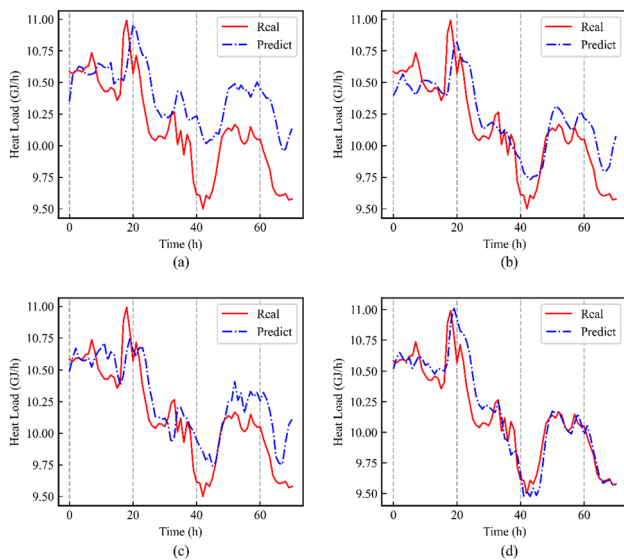
prediction results did not decrease. Still, they increased, indicating that the improved 1DCNN-BiLSTM model could not effectively extract the characteristics of rapid changes in the process of heat load change. Finally, the attention mechanism is added to the 1DCNN-BiLSTM model to extract the short-term fluctuation characteristics of heat load in a targeted manner. The results of each index show that this improvement is effective, significantly reducing the index error, and the fluctuation of the index is decreased considerably.

Comparison of the heat load prediction results of different algorithms

With the selected key variables as inputs and the heat load as outputs over the next three days, the prediction results of the different models are shown in Fig. 11. As shown in Fig. 11a, The prediction results of the MLP model were relatively accurate in the first 20 h. Still, the following prediction errors became larger, especially from the highest point at the 20th hour to the lowest point at the 40th hour. It indicates that the MLP model has some advantages in short-term time series prediction, and the ability to extend the prediction to a more extended period is insufficient. In Fig. 11b, The LSTM model outperforms the MLP model in predicting the same situation due to its time-series feature extraction and memory function. Although it is similar to the actual heat load variation trend, the results for the highest and lowest points of heat load are not accurately predicted. The analysis of the real data collected shows that the data related to the heat load has multi-granularity features, i.e., both coarse and fine granularity features. Therefore, a 1DCNN feature extraction module is added to the LSTM model to enhance the feature extraction ability. The LSTM is changed to a BiLSTM model to enhance

Table 3 Statistical results of the MLP, LSTM, 1DCNN- BiLSTM, Attention- 1DCNN- BiLSTM Prediction model

Evaluation indicators	Prediction models	Min	1st Q	Median	3rd Q	Max	Interquartile	Range
MAE	MLP	195.54	227.20	254.31	262.27	314.37	35.07	118.83
	LSTM	184.52	200.42	215.18	221.27	242.75	20.85	58.23
	1DCNN-BiLSTM	178.60	199.12	210.36	231.10	263.74	31.99	85.15
	Attention-1DCNN-BiLSTM	128.73	135.14	140.26	148.11	183.61	12.97	54.88
RMSE	MLP	249.37	279.92	304.85	314.43	370.57	34.51	121.20
	LSTM	227.52	243.48	258.07	263.66	286.09	20.18	58.58
	1DCNN-BiLSTM	216.30	237.76	248.53	270.35	302.40	32.59	86.09
	Attention-1DCNN-BiLSTM	164.95	174.74	180.70	189.44	226.47	14.70	61.52
MAPE	MLP	0.020	0.023	0.026	0.026	0.032	0.004	0.012
	LSTM	0.019	0.020	0.022	0.022	0.025	0.002	0.006
	1DCNN-BiLSTM	0.018	0.020	0.021	0.023	0.027	0.003	0.009
	Attention-1DCNN-BiLSTM	0.013	0.014	0.014	0.015	0.019	0.001	0.006
MAX Error	MLP	579.21	640.59	675.68	699.82	810.81	59.23	231.60
	LSTM	500.61	520.67	550.10	568.96	603.84	48.29	103.23
	1DCNN-BiLSTM	489.16	532.17	549.12	582.91	647.26	50.74	158.11
	Attention-1DCNN-BiLSTM	397.65	421.02	442.74	470.74	517.77	49.72	120.12

**Fig. 11** The comparison between the predicted values and the real values. **a** MLP; **b** LSTM; **c** 1DCNN-BiLSTM; and **d** Attention-1DCNN-BiLSTM

the memory of temporal features. The improved 1DCNN-BiLSTM model does not significantly improve heat load trend prediction compared with the original LSTM model, as shown in Fig. 11c. The maximum and minimum values of heat load are not accurately predicted. The reason for this may be that the addition of new functional modules simultaneously increases the complexity of the model, and the model's ability to sense extreme cases is insufficient, resulting in poor model prediction. Further attempts are made

to add an attention mechanism to the improved 1DCNN-BiLSTM model. And the model is simplified by reducing the feature extraction and temporal prediction layers so that the model can focus more on learning heat load trends in extreme cases. Sure enough, the above improvements led to a significant improvement in the prediction accuracy of the latest model, with great precision in both the prediction of heat load extremes and the prediction of change trends, as shown in Fig. 11d.

Statistical analysis of the heat load prediction results of different algorithms

Further, we perform a statistical analysis of the prediction results of different models. The scatter plots corresponding to the four groups of prediction models are shown in Fig. 12. The horizontal and vertical coordinates represent the actual and predicted heat load. The red and blue dividing lines represent the prediction error value greater than or less than $\pm 0.03\%$, respectively. When the point falls between the two lines, the prediction result is closer to the actual heat load, and the error is considered acceptable in the engineering application. It can be seen from Fig. 12a that only 47.89% of the points in the MLP model fall within the acceptable error range, while 49.3% of the points have error values greater than 0.03%, and only 2.82% of the points have error values less than 0.03%. From Fig. 12b, it can be seen that more than 87.32% of the heat load prediction errors in the LSTM model are within the interval $[-0.03\%, +0.03\%]$, which is a great improvement compared to the MLP model. As shown

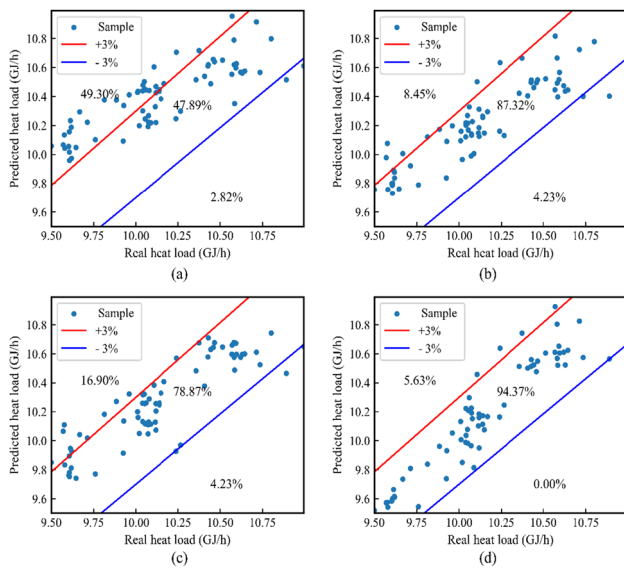


Fig. 12 Scatter plots of heat load with different models **a** MLP; **b** LSTM; **c** 1DCNN-BiLSTM; and **d** Attention-1DCNN-BiLSTM

in Fig. 12c, the prediction results of the 1DCNN-BiLSTM model were further counted. It was found that 78.87% of the heat load predictions were within the acceptable error range, significantly lower than the LSTM model. It shows that the high model complexity causes a decrease in prediction accuracy. Finally, from Fig. 12d, the attention-based 1DCNN-BiLSTM model was obtained by simplifying the 1DCNN-BiLSTM model and adding an attention mechanism. Its prediction error within the acceptable range reaches 94.37%, with higher accuracy than MLP, LSTM, and other advanced models. The prediction models we have proposed above as well as the traditional LSTM models have achieved good prediction accuracy. Although the complexity of the LSTM model is lower than that of our proposed model, the convergence speed and accuracy of our proposed model are better. However, our proposed model is more effective in convergence speed and accuracy. This is because the complexity of the original data is significantly reduced through a series of dimensionality reductions and 1DCNN feature extraction based on the attention mechanism. Then the BiLSTM model can learn the intrinsic mechanism of the parameters better. This results in better performance than the LSTM model alone.

The SOTA comparison tests

In order to verify the superiority of the proposed method in the paper, we further carried out state-of-the-art (SOTA) comparison tests. Such as Temporal Convolutional Network (Kok et al. 2020) (TCN), and the Transformer model (Acheampong et al. 2021). The Attention-1DCNN-BiLSTM model is compared with these two typical SOTA models. The predicted output of each model is compared with the true value

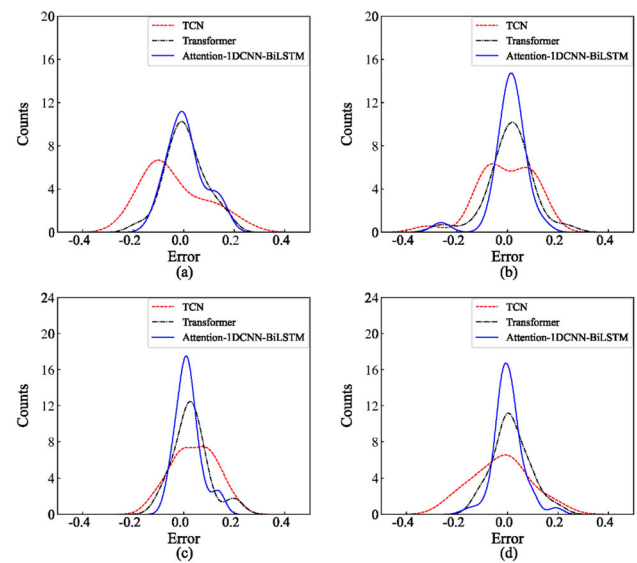


Fig. 13 Error frequency histogram of heat load with different models **a** the first cycle; **b** the second cycle; **c** the third cycle; and **d** the fourth cycle

to obtain the error comparison result at each moment, and the error value of each moment is counted in 24 h, the heat load was counted for four consecutive days, and the error at each moment is counted, as shown in Fig. 13. From the comparison chart, it can be seen that the TCN model and Transformer model do not perform well. This may be because TCN models are generally unidirectional in structure and fail to accurately identify sudden heat loads with backward and forward correlations. The Transformer model is not as good as the RNN and CNN models in acquiring local information, and the top gradient tends to disappear. Therefore, the attention mechanism network model designed in this paper for the characteristics of rapid changes in heat load can achieve better prediction results.

Conclusion

This research proposes an attention mechanism-based deep learning for heat load prediction in the BF ironmaking process. The main focus of this research is to offer a systematic framework of data-driven methods for the BF heat load prediction problem. Two parts are included, a two-stage data pre-processing and a prediction model based on an attention mechanism. In particular, in the second part, we focus on the lagging nature of traditional methods and the unpredictability of trends when the heat load changes with sharp fluctuations. An attention mechanism is added to ensure real-time heat load prediction and the identification of abnormal fluctuations. Finally, the experiment results of different prediction models show that the proposed prediction model can accurately grasp the rapid change trend of heat load, and the

accuracy of the prediction model achieves 94.37%, which is higher than that of the comparison prediction models. However, there are still some deficiencies in this research. The feature selection section provides a nonlinear correlational relationship feature selection method based on mutual information, whose thresholds are determined mainly based on empirical values. Next, we also consider adaptive threshold determination methods and correlational coupling effects between variables. Besides, the impact of model complexity on prediction performance is evident from the experimental part. In future research works, we will investigate other models and work on the simplification methods of complex models to achieve high prediction accuracy of simple models.

Acknowledgements This work was supported by the National Key Research and Development Program of China (No. 2020YFB1711100, No. 2019YFB1704401). Furthermore, the authors wish to thank the Editor and anonymous referees for their constructive comments and recommendations, which have significantly improved the presentation of this paper.

Funding This work was supported by the National Key Research and Development Program of China (No. 2020YFB1711100, No. 2019YFB1704401).

Declarations

Conflict of interest No potential conflict of interest was reported by the author(s).

References

- Acheampong, F. A., Nunoo-Mensah, H., & Chen, W. (2021). Transformer models for text-based emotion detection: a review of BERT-based approaches. *Artif Intell Rev*, 54(8), 5789–5829. <https://doi.org/10.1007/s10462-021-09958-2>
- Chen, X. H., Zhang, B. K., & Gao, D. (2020). Bearing fault diagnosis base on multi-scale CNN and LSTM model. *J Intell Manuf.* <https://doi.org/10.1007/s10845-020-01600-2>
- Deng, W. H., & Wang, G. Y. (2017). A novel water quality data analysis framework based on time-series data mining. *J Environ Manage*, 196, 365–375. <https://doi.org/10.1016/j.jenvman.2017.03.024>
- Dong Z, Xiang W, Xue X, Chen S, Wang X (2011) On-line identification of thermal process using a modified TS-type neuro-fuzzy system. In: 2011 2nd International Conference on Artificial Intelligence, Management Science and Electronic Commerce (AIMSEC)
- Eren, L., Ince, T., & Kiranyaz, S. (2019). A generic intelligent bearing fault diagnosis system using compact adaptive 1D CNN classifier. *J Signal Process Syst Signal Image Video Technol*, 91(2), 179–189. <https://doi.org/10.1007/s11265-018-1378-3>
- Hou, Z., Ma, K., Wang, Y., Yu, J., Ji, K., Chen, Z., & Abraham, A. (2021). Attention-based learning of self-media data for marketing intention detection. *Eng Appl Artif Intell.* <https://doi.org/10.1016/j.engappai.2020.104118>
- Hu, R., Wen, S. P., Zeng, Z. G., & Huang, T. W. (2017). A short-term power load forecasting model based on the generalized regression neural network with decreasing step fruit fly optimization algorithm. *Neurocomputing*, 221, 24–31. <https://doi.org/10.1016/j.neucom.2016.09.027>
- Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. *Int J Forecast*, 22(4), 679–688. <https://doi.org/10.1016/j.ijforecast.2006.03.001>
- Kok, C., Jahmunah, V., Oh, S. L., Zhou, X., Gururajan, R., Tao, X., Cheong, K. H., Gururajan, R., Molinari, F., & Acharya, U. R. (2020). Automated prediction of sepsis using temporal convolutional network. *Comput Biol Med*, 127, 103957. <https://doi.org/10.1016/j.compbiomed.2020.103957>
- Kurunov, I. F., Loginov, V. N., & Tikhonov, D. N. (2006). Methods of extending a blast-furnace campaign. *Metallurgist*, 50(11–12), 605–613. <https://doi.org/10.1007/s11015-006-0131-5>
- Lecun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proc IEEE*, 86(11), 2278–2324. <https://doi.org/10.1109/5.726791>
- Li, H. Z., Guo, S., Li, C. J., & Sun, J. Q. (2013). A hybrid annual power load forecasting model based on generalized regression neural network with fruit fly optimization algorithm. *Knowl-Based Syst*, 37, 378–387. <https://doi.org/10.1016/j.knsys.2012.08.015>
- Li, C. K., Hou, Y. H., Wang, P. C., & Li, W. Q. (2017). Joint distance maps based action recognition with convolutional neural networks. *IEEE Signal Process Lett*, 24(5), 624–628. <https://doi.org/10.1109/Lsp.2017.2678539>
- Li, J., Hua, C., & Yang, Y. (2021). Output space transfer-based MIMO RVFLNs modeling for estimation of blast furnace molten iron quality with missing indexes. *IEEE Trans Instrum Measure*, 70, 1–10. <https://doi.org/10.1109/tim.2021.3063200>
- Li, J., Hua, C., Yang, Y., & Guan, X. (2022). Data-driven Bayesian-based Takagi-Sugeno fuzzy modeling for dynamic prediction of hot metal silicon content in blast furnace. *IEEE Trans Syst Man Cybern Syst*, 52(2), 1087–1099. <https://doi.org/10.1109/tsmc.2020.3013972>
- Liao, S., Zhu, Q., Qian, Y., & Lin, G. (2018). Multi-granularity feature selection on cost-sensitive data with measurement errors and variable costs. *Knowl Based Syst*, 158, 25–42. <https://doi.org/10.1016/j.knsys.2018.05.020>
- Lv, Y. L., Ji, Q. H., Liu, Y., & Zhang, J. (2020). Data-driven sensitivity analysis of contact resistance to assembly errors for proton-exchange membrane fuel cells. *Measure Control*, 53(7–8), 1354–1363. <https://doi.org/10.1177/0020294020926604>
- Lv, Y., Zhou, Q., Li, Y., & Li, W. (2021). A predictive maintenance system for multi-granularity faults based on AdaBelief-BP neural network and fuzzy decision making. *Adv Eng Inform*, 49. <https://doi.org/10.1016/j.aei.2021.101318>
- Pan, Y. J., Yang, C. J., An, R. Q., & Sun, Y. X. (2018). Robust principal component pursuit for fault detection in a blast furnace process. *Indus Eng Chem Res*, 57(1), 283–291. <https://doi.org/10.1021/acs.iecr.7b03338>
- Qin, W., Zha, D. Y., & Zhang, J. (2020). An effective approach for causal variables analysis in diesel engine production by using mutual information and network deconvolution. *J Intell Manuf*, 31(7), 1661–1671. <https://doi.org/10.1007/s10845-018-1397-8>
- Ribeiro, G. T., Mariani, V. C., Coelho, L., & DS. (2019). Enhanced ensemble structures using wavelet neural networks applied to short-term load forecasting. *Eng Appl Artif Intell*, 82, 272–281. <https://doi.org/10.1016/j.engappai.2019.03.012>
- Semenov, Y. S., Shumel'chik, E. I., Gorupakha, V. V., Nasledov, A. V., Kuznetsov, A. M., & Zubenko, A. V. (2017). Monitoring blast furnace lining condition during five years of operation. *Metallurgist*, 61(3–4), 291–297. <https://doi.org/10.1007/s11015-017-0491-z>
- Taler, J., Zima, W., Oclon, P., Gradziel, S., Taler, D., Cebula, A., Jaremkiewicz, M., Korzen, A., Cisek, P., Kaczmarek, K., & Majewski, K. (2019). Mathematical model of a supercritical

- power boiler for simulating rapid changes in boiler thermal loading. *Energy*, 175, 580–592. <https://doi.org/10.1016/j.energy.2019.03.085>
- Tian, H. X., Ren, D. X., Li, K., & Zhao, Z. (2020). An adaptive update model based on improved long short term memory for online prediction of vibration signal. *J Intell Manuf.* <https://doi.org/10.1007/s10845-020-01556-3>
- Wang, J. L., Zhang, J., & Wang, X. X. (2018a). Bilateral LSTM: a two-dimensional long short-term memory model with multiply memory units for short-term cycle time forecasting in re-entrant manufacturing systems. *IEEE Trans Indus Inform*, 14(2), 748–758. <https://doi.org/10.1109/Tii.2017.2754641>
- Wang, L., Yang, C. J., Sun, Y. X., Zhang, H. F., & Li, M. L. (2018b). Effective variable selection and moving window HMM-based approach for iron-making process monitoring. *J Process Control*, 68, 86–95. <https://doi.org/10.1016/j.jprocont.2018.04.008>
- Wang, J. L., Yang, Z. L., Zhang, J., Zhang, Q. H., & Chien, W. T. K. (2019). AdaBalGAN: an improved generative adversarial network with imbalanced learning for wafer defective pattern recognition. *IEEE Trans Semicond Manuf*, 32(3), 310–319. <https://doi.org/10.1109/Tsm.2019.2925361>
- Wang, J. L., Xu, C. Q., Yang, Z. L., Zhang, J., & Li, X. O. (2020). Deformable convolutional networks for efficient mixed-type wafer defect pattern recognition. *IEEE Trans Semicond Manuf*, 33(4), 587–596. <https://doi.org/10.1109/Tsm.2020.3020985>
- Wang, X., Hu, T., & Tang, L. (2021). A multiobjective evolutionary nonlinear ensemble learning with evolutionary feature selection for silicon prediction in blast furnace. *IEEE Trans Neural Netw Learn Syst.* <https://doi.org/10.1109/TNNLS.2021.3059784>
- Wang HW, Yang GK, Pan CC, Gong QS (2015) Prediction of hot metal silicon content in blast furnace based on EMD and DNN. In: 2015 34th Chinese Control Conference (Ccc), pp 8214–8218
- Wang PC, Li WQ, Gao ZM, Zhang YY, Tang C, Ogunbona P (2017) Scene flow to action map: a new representation for RGB-D based action recognition with convolutional neural networks. In: 30th IEEE Conference on Computer Vision and Pattern Recognition (Cvpr 2017), pp 416–425. <https://doi.org/10.1109/Cvpr.2017.52>
- Xie, L. (2017). The heat load prediction model based on BP neural network-markov model. *Adv Inform Commun Technol*, 107, 296–300. <https://doi.org/10.1016/j.procs.2017.03.108>
- Xing, X., Rogers, H., Zhang, G. Q., Hockings, K., Zulli, P., Deev, A., Mathieson, J., & Ostrovski, O. (2017). Effect of charcoal addition on the properties of a coke subjected to simulated blast furnace conditions. *Fuel Process Technol*, 157, 42–51. <https://doi.org/10.1016/j.fuproc.2016.11.009>
- Xu, L. J., Khan, J. A., & Chen, Z. H. (2000). Thermal load deviation model for superheater and reheater of a utility boiler. *Appl Thermal Eng*, 20(6), 545–558. [https://doi.org/10.1016/S1359-4311\(99\)00049-6](https://doi.org/10.1016/S1359-4311(99)00049-6)
- Xu, H. W., Zhang, J., Lv, Y. L., & Zheng, P. (2020). Hybrid feature selection for wafer acceptance test parameters in semiconductor manufacturing. *IEEE Access*, 8, 17320–17330. <https://doi.org/10.1109/Access.2020.2966520>
- Xu, H. W., Qin, W., Lv, Y. L., & Zhang, J. (2022). Data-driven adaptive virtual metrology for yield prediction in multi-batch wafers. *IEEE Trans Indus Inform.* <https://doi.org/10.1109/Tii.2022.3162268>
- Yukun L, Peng X, Zhao K (2011) Hybrid modeling optimization of thermal efficiency and NO_x emission of utility boiler. *Proc CSEE.* <https://doi.org/10.13334/j.0258-8013.pcsee.2011.26.00>
- Zhang, J. J., Wang, P., Yan, R. Q., & Gao, R. X. (2018). Long short-term memory for machine remaining life prediction. *J Manuf Syst*, 48, 78–86. <https://doi.org/10.1016/j.jmsy.2018.05.011>
- Zhang, F. L., Yan, J. X., Fu, P. L., Wang, J. J., & Gao, R. X. (2020). Ensemble sparse supervised model for bearing fault diagnosis in smart manufacturing. *Robot Comput-Integr Manuf.* <https://doi.org/10.1016/j.rcim.2019.101920>
- Zhao, J. J., Peng, Y. X., & He, X. T. (2020). Attribute hierarchy based multi-task learning for fine-grained image classification. *Neurocomputing*, 395, 150–159. <https://doi.org/10.1016/j.neucom.2018.02.109>
- Zhao H, Wang PH (2009) Modeling and optimization of efficiency and NO_x emission at a coal-fired utility boiler. In: 2009 Asia-Pacific Power and Energy Engineering Conference (Apeec), vols 1–7, pp 2832–2835
- Zhong WL, Jiang LF, Zhang T, Ji JS, Xiong HL (2018) A multi-part convolutional attention network for fine-grained image recognition. In: 2018 24th International Conference on Pattern Recognition (ICPR), pp 1857–1862
- Zhou, P., Lv, Y. B., Wang, H., & Chai, T. Y. (2017). Data-driven robust RVFLNs modeling of a blast furnace iron-making process using cauchy distribution weighted M-estimation. *IEEE Trans Indus Elect*, 64(9), 7141–7151. <https://doi.org/10.1109/Tie.2017.2686369>
- Zhou, P., Guo, D. W., & Chai, T. Y. (2018). Data-driven predictive control of molten iron quality in blast furnace ironmaking using multi-output LS-SVR based inverse system identification. *Neurocomputing*, 308, 101–110. <https://doi.org/10.1016/j.neucom.2018.04.060>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.