

Article

Prediction of Coiling Temperature for Hot-Rolled Strip Steel Based on WOA-CNN-GRU-SE Model

Tiejun Sun, Hongjiang Cao, Xiaodan Zhang *, Luyao Sun, Zhiheng Meng and Yanming Cheng

College of Electrical and Information Engineering, Beihua University, Jilin 132013, China; suntj@beihua.edu.cn (T.S.); caohongjiang@stu.beihua.edu.cn (H.C.); sunluyao@stu.beihua.edu.cn (L.S.); mengzhiheng@stu.beihua.edu.cn (Z.M.); mycheng@beihua.edu.cn (Y.C.)

* Correspondence: zxd@beihua.edu.cn

Abstract

Coiling temperature is a pivotal process parameter for hot-rolled strip steel, which directly determines the microstructure and mechanical properties of final products. Affected by the coupling of multiple process variables, coiling temperature presents strong nonlinearity and complex time-varying characteristics. Traditional heat transfer mechanism models, Random Forest (RF), Extreme Learning Machine (ELM) and single Long Short-Term Memory (LSTM) networks fail to fully explore the deep correlation among variables. In addition, their hyperparameters are generally selected by manual trial-and-error, leading to unsatisfactory prediction accuracy and poor robustness in practical production. To address the above limitations, this paper proposes a novel prediction model named WOA-CNN-GRU-SE, where the Whale Optimization Algorithm (WOA) is adopted for parameter optimization. Firstly, Convolutional Neural Network (CNN) is utilized to extract local coupling features from various working condition parameters. Secondly, the Squeeze-and-Excitation (SE) attention mechanism is applied to adaptively recalibrate channel weights, which enhances key features closely related to temperature variation and suppresses redundant interference information. Afterwards, Gated Recurrent Unit (GRU) is employed to conduct in-depth learning of temporal features. Furthermore, WOA is used to globally optimize critical hyperparameters, including learning rate, the number of GRU hidden units and L2 regularization coefficient, so as to eliminate the drawbacks of manual parameter tuning. Comparative experiments are conducted on actual production data from a hot rolling line. The results demonstrate that the proposed model outperforms CNN-GRU, CNN-GRU-SE, LSTM, RF and ELM in prediction performance. Its hit rate reaches 92.56% within the industrial error range of ± 6 °C. This model effectively realizes accurate prediction of coiling temperature under complex working conditions and possesses great application potential in industrial practice.

Keywords: coiling temperature prediction; whale optimization algorithm; convolutional neural network; Gated recurrent unit; squeeze-and-excitation attention mechanism

Academic Editors: Lukasz Sobolewski, Marcel Luzar, Wiesław Miczulski, Marcin Mrugalski and Andrzej Obuchowicz

Received: 17 June 2026

Revised: 8 July 2026

Accepted: 9 July 2026

Published: 13 July 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\) license](https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Coiling temperature is a critical process parameter in the production of hot-rolled continuous strip steel, and its control accuracy directly affects the microstructure evolution, mechanical properties of the strip steel, and the quality of the final product [1–

3]. In actual industrial production, excessively high coiling temperature easily leads to microstructure coarsening and decreased strength, and negatively affects the performance of subsequent cold working processes. Conversely, excessively low coiling temperature may cause problems such as difficulty in coiling, deteriorated strip shape, and increased residual stress [4]. Therefore, achieving accurate prediction and reasonable control of coiling temperature is of great significance for improving the quality of hot-rolled strip products and ensuring the stable operation of production.

The cooling and coiling process of hot-rolled strip after rolling is jointly affected by multiple factors, involving complex physical processes such as radiative heat transfer, convective heat transfer, heat conduction and phase transformation, and is also subject to the coupled influence of multiple factors including finishing rolling temperature, strip thickness, rolling speed, cooling water flow rate and fluctuations in on-site working conditions [5,6]. Since this process exhibits significant nonlinearity, complexity and uncertainty, traditional coiling temperature prediction models established based on heat transfer mechanisms, empirical formulas or mathematical statistical methods, despite having a certain theoretical basis and being widely applied in industrial sites, often fail to accurately describe the nonlinear mapping relationship between multiple process variables and coiling temperature under complex working conditions, and still have certain limitations in model adaptability and prediction accuracy [7].

With the continuous accumulation of on-site industrial data, data-driven intelligent prediction methods have gradually become an important research direction for coiling temperature modeling. PENG et al. [8] proposed a real-time adaptive setting algorithm based on historical data. This algorithm uses data mining technology to cluster similar coils, solves the quadratic programming problem via the gradient descent method, and dynamically optimizes the heat transfer learning coefficient. Hou et al. [9] established a coiling temperature prediction model based on improved cuckoo search algorithm optimized support vector regression (ICS-SVR), and improved the prediction accuracy under multiple working conditions by adjusting penalty factors and kernel parameters. Sun Tiejun et al. [10] used an improved genetic algorithm to optimize the weights and thresholds of the BP neural network, and introduced “atavism” and “eugenics” operations into the algorithm to improve population diversity and global optimization ability, which controlled the prediction error within ± 10 °C. Zhang Shuai et al. [11] proposed a coiling temperature prediction model for hot-rolled strip steel based on improved gray wolf optimization optimized extreme learning machine, and the proposed IGWO-ELM model achieves superior prediction accuracy for coiling temperature. Hao Qiuyu et al. [12] established a coiling temperature prediction model for hot-rolled strip steel based on random forest, used the Bayesian optimization algorithm to optimize the hyperparameters of the random forest model, and compared and verified the performance of this model with the Bayesian optimized decision tree model, support vector regression model and on-site mechanism model. The results show that the proposed model outperforms all other comparison models in terms of error control and prediction stability. Zhou Jianxin et al. [13] proposed the NMWOA-LSTM prediction model, which uses the improved whale optimization algorithm to optimize the key parameters of LSTM, further improving the prediction accuracy of coiling temperature.

Compared with traditional methods, data-driven models can directly learn the nonlinear relationship between input variables and target variables from production data, and have promising application prospects in complex process modeling. However, although data-driven intelligent methods have been widely adopted for coiling temperature prediction, existing approaches still have several inevitable deficiencies. Traditional machine learning models such as Random Forest (RF) and Extreme Learning Machine (ELM) can only fit simple linear relationships and struggle to describe the strong

nonlinear characteristics caused by coupled process variables. Single temporal networks like Long Short-Term Memory (LSTM) focus on temporal information learning but lack the ability to extract local combined features of working parameters. Moreover, the existing CNN-GRU hybrid models cannot distinguish the importance of different features, and redundant features will interfere with model training. In addition, key hyperparameters of deep learning models are usually determined by manual trial-and-error, which is time-consuming and unable to obtain global optimal solutions, further restricting the prediction accuracy and generalization performance. Against this background, improving the coiling temperature prediction model by combining optimization algorithms and deep learning methods has important research significance and engineering value.

To overcome these technical gaps, this paper develops a hybrid framework named WOA-CNN-GRU-SE by integrating feature extraction, attention mechanism, temporal learning and swarm intelligence optimization. CNN is adopted to extract local features that cannot be effectively captured by single temporal networks. The Squeeze-and-Excitation (SE) attention mechanism is introduced to adaptively screen features and highlight valid information. Gated Recurrent Unit (GRU) is used to explore the dynamic temporal rules of temperature changes. Meanwhile, Whale Optimization Algorithm (WOA) with excellent global search capability is employed to optimize hyperparameters instead of manual adjustment. This work aims to achieve high-accuracy prediction of coiling temperature under complex industrial conditions and provide a reliable reference for on-site temperature control.

Based on the above analysis, this paper proposes a WOA-CNN-GRU-SE coiling temperature prediction model. The model takes CNN-GRU-SE as the main prediction framework, and embeds the SE attention mechanism between the CNN feature extraction module and the GRU module to enhance the expression ability of key features. Meanwhile, the whale optimization algorithm (WOA) is introduced to optimize the learning rate, the number of GRU hidden units and the L2 regularization coefficient. Finally, experiments were conducted based on actual production data from a hot rolling mill line, and the effectiveness of the proposed model was verified through ablation studies and control experiments.

The main innovations of this work are summarized as follows:

- (1) A hybrid CNN-GRU-SE architecture is constructed. It combines the local feature extraction of CNN, feature enhancement of SE attention mechanism and temporal modeling of GRU, to fully explore the nonlinear mapping relationship between multi-process parameters and coiling temperature.
- (2) The Whale Optimization Algorithm (WOA) is applied to globally optimize the learning rate, number of GRU hidden units and L2 regularization coefficient, which solves the low efficiency and suboptimal parameters caused by manual tuning.
- (3) Multiple comparative experiments and ablation studies are carried out based on real production data from a hot rolling line. The results verify the effectiveness of each module and the practicality of the proposed model in industrial scenarios.

2. Mechanistic Model and Data Processing

2.1. Physical Mechanism of Laminar Flow Cooling

In the hot continuous rolling production process, the post-rolling cooling procedure is a critical link to achieve the target coiling temperature. After exiting the last finishing mill stand, the strip steel generally has a finish rolling temperature of 800~900 °C, and needs to be cooled to approximately 500~600 °C after passing through a laminar cooling zone over one hundred meters in length, so as to meet the requirements of downstream

mechanical properties. This drastic temperature reduction process involves complex energy exchange between the strip steel, cooling water, air and roller tables. To accurately describe this process, existing mechanistic models mainly rely on classical heat transfer theory to establish one-dimensional unsteady heat conduction differential equations for quantitative characterization of the aforementioned energy release process.

For temperature prediction in hot continuous rolling lines, existing mechanism-based models primarily construct heat conduction differential equations based on classical heat transfer theory. By setting corresponding initial and boundary conditions, the finite difference method is employed to calculate the real-time temperature distribution of the strip steel. Taking the finishing rolling temperature as the initial condition, and taking the heat flux density of radiative heat transfer of the strip steel, convective heat transfer of cooling water, and heat conduction of the transport rollers as boundary conditions, the real-time temperature field of the strip steel throughout the entire cooling zone is obtained by solving the difference equation at each time step. In the actual post-rolling cooling process, heat dissipation along the thickness direction of the strip steel is absolutely dominant. Therefore, the cooling process of the strip steel is simplified as a one-dimensional unsteady heat conduction problem, and the one-dimensional heat conduction differential equation is shown in Equation (1).

$$\frac{\partial T}{\partial t} = \frac{\lambda}{\rho c} \left(\frac{\partial^2 T}{\partial x^2} \right) \quad (1)$$

In this equation: T represents the temperature of the strip steel, in °C; t represents time, in s; λ represents the thermal conductivity of the strip steel, in W/(m·°C); ρ represents the material density, in kg/m³; c represents the specific heat capacity of the material, in J/(kg·°C); x represents the coordinate along the thickness direction of the strip steel.

Although the differential equation of heat conduction defines the intrinsic spatiotemporal relationship of the temperature field, solving the actual dynamic temperature process of strip steel still relies on the constraints of initial conditions and boundary conditions. Specifically, the initial condition, which refers to the temperature distribution across the entire spatial domain at the starting moment, is completely determined by the exit temperature of the strip steel when it leaves the finishing mill; the boundary condition reflects the heat transfer characteristics on the strip surface, which requires real-time adjustment based on the actual cooling working conditions, and its specific calculation is shown in Equation (2).

$$-\lambda \frac{\partial T}{\partial x} \bigg|_{x=0,m} = \alpha(T - T_f) \quad (2)$$

In the equation, T_f denotes the temperature of the cooling medium on the strip surface, in °C; α denotes the heat transfer coefficient on the strip surface, in W/(m²·°C); m denotes the thickness of the strip, in mm.

The calculation of boundary conditions in the mechanistic model is highly dependent on the surface heat transfer coefficient a . Depending on the cooling medium, this coefficient can be subdivided into a_{air} for air-cooled sections and a_{water} for water-cooled sections. Currently, conventional calculations often use regression equations to approximate the water-cooling heat transfer coefficient, with the specific mathematical model shown in Equation (3).

$$\alpha_{water} = Z \cdot a_1 \cdot c_w \cdot e^{-a_2 \cdot (T - T_w)} \cdot \left(\frac{V}{V_{base}}\right) \cdot \frac{f}{w_b \cdot l_s} \quad (3)$$

where a_1 , a_2 , c_w , V_0 are model parameters; Z is the heat transfer coefficient correction term; V is the real-time segmented strip speed, m/s; V_{base} is the finishing mill strip entry speed, m/s; T_w is the water temperature, °C; f is the water flow rate, m³/h; w_b is the cooling zone width, mm; l_s is the cooling section length, mm.

2.2. Data Collection and Preprocessing

The historical raw production data of thin hot-rolled wide strip steel from a domestic hot strip mill production line of a Chinese steel plant were adopted as the modeling data. The strip steel grade involved in this study is Q235B plain carbon steel (as defined in the Chinese National Standard GB/T 700-2006) [14], and the dataset contains data collected from 26 coils of strip steel. To meet the requirements of different product specifications, this batch of steel coils covers 4 different strip thickness specifications, namely 4.5 mm, 5.5 mm, 7.0 mm and 17.5 mm. During the laminar cooling process, the secondary process control system implements segmented control and precise tracking for the strip steel. Through high-frequency periodic sampling, the system dynamically acquires various process parameters of the strip steel during cooling, and conducts temporal discretization to convert the continuous temperature measurement process into “strip steel feature points” that strictly correspond to each process parameter.

To improve the prediction accuracy of the model, the Pauta criterion (Equation (4)) is employed to eliminate abnormal data points hidden in the modeling dataset. Finally, the strip thickness, finish rolling exit temperature, average strip speed and the number of open water valves are selected as the input variables of the prediction model, while the actual coiling temperature is taken as the output variable of the model. The prediction results of the model are used for comparison with the actual coiling temperature. Moreover, it should be noted that the modeling dataset used in this study relies on proprietary historical data from an actual industrial hot rolling production line. Due to inherent limitations of field data acquisition systems and confidentiality requirements for industrial data, the dataset lacks several specific physical parameters. Specifically, missing parameters include complete geometric dimensions (such as the exact strip width), direct cooling medium parameters (such as water flow rate and cooling water temperature), and seasonal environmental variables (such as ambient air temperature). Although the currently selected input variables (strip thickness, average strip speed, final rolling temperature, and valve opening) are sufficient to effectively capture the primary dynamic heat transfer mapping relationships, the absence of these variables inevitably limits the full physical interpretability of the proposed model. These limitations arising from industrial data constraints are further discussed in Section 5, and addressing them will be a key focus of our future research. A total of 803 process sampling points from four strip thickness specifications are extracted from the dataset, which are then divided into a training set, a validation set and a test set according to the ratio of 7:1.5:1.5. Prior to model training, this study shuffled the 803 sampling points and converted them into independent discrete state points. As shown in Table 1, to ensure the stability of model training and reduce the impact of differences across different feature dimensions on the network training process, a normalization method is adopted to map the input features and output target values into a unified numerical interval, thereby providing a reliable data foundation for subsequent model training and prediction. The normalization formula is presented as Equation (5).

$$|x_i^j - \bar{x}^j| > 3\sigma, i = 1, 2, \dots, n; j = 1, 2, \dots, k \quad (4)$$

Table 1. Partial sample data of coiling temperature.

Serial Number of Cooling Sampling Point	Strip Thickness (mm)	Finish Rolling Temperature (°C)	Average Strip Speed (m/s)	Actual Coiling Temperature (°C)	Open Valves
1	17.69	800	3.27	665	45
2	17.711	800	3.27	626	45
3	7.006	863	4.15	657	27
4	4.542	863	10.16	669	46
5	7.011	862	4.13	658	27
6	7.017	869	4.43	687	34
7	7.024	868	4.42	683	34
8	7.041	879	4.11	662	27
9	7.006	862	4.52	683	34
10	4.551	863	9.26	677	51
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.
794	5.061	905	9.08	737	35
795	5.499	919	8.1	742	31
796	4.537	853	8.3	663	33
797	6.98	869	4.43	687	34
798	5.594	905	9.05	734	35
799	5.545	901	8.18	731	29
800	4.551	862	9.08	673	51
801	5.573	905	9.23	734	35
802	17.816	800	3.27	544	59
803	5.506	919	8.1	741	31

In the equation, x_i^j denotes the i feature of the j sample, $\bar{x}^j$ and σ represent the mean and standard deviation of the j feature (label), respectively, n is the number of samples, and k is the number of features and labels per sample.

$$x_i^* = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}} \quad (5)$$

wherein x_i denotes the raw data, x_i^* represents the normalized data, $x_{\min}$ and $x_{\max}$ are the minimum and maximum values of the sample data, respectively.

To further analyze the correlation between the input process parameters and the actual coiling temperature, this paper adopts the Pearson correlation coefficient to conduct correlation analysis on the sample data, and the results are shown in Figure 1.

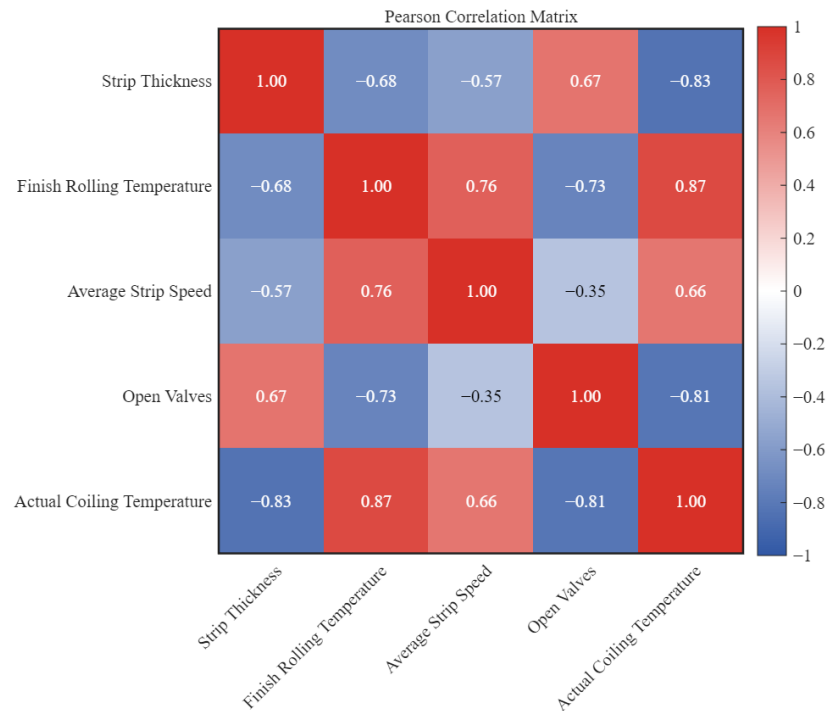


Figure 1. Pearson correlation matrix of different variables.

As shown in Figure 1, both the finishing rolling temperature and the average velocity in the water-cooling zone are positively correlated with the actual coiling temperature. Among them, the Pearson correlation coefficient between the finishing rolling temperature and the actual coiling temperature reaches as high as 0.87. From the perspective of the solution mechanism of the one-dimensional unsteady heat conduction differential equation in heat transfer (as shown in Equation (1)), the finishing rolling temperature accurately characterizes the initial enthalpy field of the strip steel at the instant it leaves the last finishing rolling stand and enters the cooling system, which constitutes the absolute initial condition for the time integration of the entire cooling partial differential equation. This indicates that the initial temperature state after rolling has a significant influence on the final coiling temperature; the correlation coefficient between the average velocity in the water-cooling zone and the actual coiling temperature is 0.66, demonstrating that the average velocity in the water-cooling zone also affects the final temperature.

Meanwhile, both the strip steel thickness and the number of open valves are negatively correlated with the actual coiling temperature, with correlation coefficients of -0.83 and -0.81 respectively. Strip thickness is an important geometric parameter affecting the cooling behavior of hot-rolled strip steel. A thicker strip has a larger heat storage capacity, requiring more heat removal during the laminar cooling process, thereby affecting the final coiling temperature. In addition, strip width is also an important geometric parameter of thin hot-rolled wide strip steel, which may influence the heat transfer behavior during cooling. However, the available dataset used in this study did not contain strip width as an independent input variable, and its influence was not quantitatively analyzed.

In this context, the equivalent mapping relationship between the “valve on-off status” and the core cooling dynamic condition (cooling water flow rate) needs to be specifically noted. In actual laminar cooling engineering, the water supply network generally operates under constant pressure, which means that the flow rate through a single valve is a known constant. Therefore, the on-off status recorded by the control system has an

approximately linear positive proportional relationship with the total cooling water flow rate actually acting on the strip steel surface. A larger number of open valves corresponds to an increased water flow rate, which first causes a sharp rise in the water-cooling convective heat transfer coefficient on the strip steel surface, leading to more heat being removed and a significant decrease in the coiling temperature. Directly adopting “the number of open valves” as a characteristic indicator is physically equivalent to introducing the water flow rate, and it is also directly consistent with the discrete valve control instructions of the on-site PLC in engineering practice. It should be pointed out that the Pearson correlation coefficient mainly reflects the degree of linear correlation between variables and cannot fully describe complex nonlinear relationships. Therefore, a CNN-attention mechanism combined with the GRU network model is still required to further capture the deep mapping relationships between these variables.

3. Construction of a Predictive Model

To improve the prediction accuracy of coiling temperature for hot-rolled strip steel, this paper constructs a CNN-GRU-SE prediction model optimized based on the Whale Optimization Algorithm (WOA), the overall structure of which is shown in Figure 2. The model is mainly composed of an input module, a WOA-optimized CNN-GRU-SE coiling temperature prediction module, and an output module. First, the original coiling temperature data is normalized and divided into a training set, a validation set and a test set according to a certain proportion, so as to reduce the impact of dimensional differences among different variables on model training. The processed data is input into the CNN-GRU-SE prediction network to realize nonlinear modeling and prediction of coiling temperature.

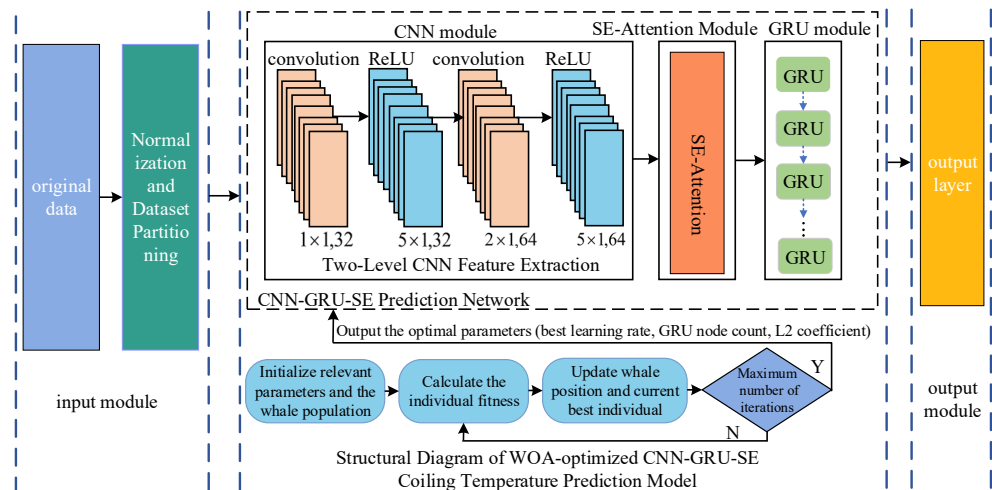


Figure 2. WOA-CNN-GRU-SE Attention Mechanism Network Structure Diagram.

In the CNN-GRU-SE prediction network, the CNN module is used to extract local correlation features among input working condition parameters. This paper adopts a two-layer convolution structure: the first convolution layer has a kernel size of 1×1 and 32 convolution kernels, and the second convolution layer has a kernel size of 2×1 and 64 convolution kernels. A ReLU activation function is introduced after convolution to enhance the nonlinear expression capability of the model. Subsequently, the SE-Attention module recalibrates the channel weights of the features extracted by CNN, enabling the model to pay more attention to key features that make greater contributions to coiling temperature prediction. The weighted features are further input into the GRU module,

which conducts in-depth learning and fusion of feature information through a gating mechanism. Finally, the coiling temperature prediction result is output through the fully connected layer.

To avoid the dependence of key hyperparameters of the model on manual experience selection, this paper introduces WOA to perform hyperparameter optimization for the CNN-GRU-SE model. WOA takes the learning rate, the number of GRU nodes, and the L2 regularization coefficient as optimization objects, and adopts normalized mean square error as the fitness evaluation index. In the optimization process, relevant parameters and the whale population are initialized first, then the model fitness under the parameter combination corresponding to each individual whale is calculated, and the whale position and the current optimal solution are continuously updated. When the maximum number of iterations is reached, the optimal parameter combination is output and used for the training of the final CNN-GRU-SE prediction model. Among them, the learning rate and the L2 regularization coefficient act as network training parameters in the overall model training process, and the number of GRU nodes is used to determine the structure of the GRU module.

In conclusion, the proposed WOA-CNN-GRU-SE model extracts local features via CNN, enhances the representation of key features through the squeeze-and-excitation (SE) attention mechanism, conducts in-depth feature learning with GRU, and performs adaptive optimization for key hyperparameters by combining with the whale optimization algorithm (WOA), thereby improving the model's modeling capability and prediction accuracy for the complex nonlinear relationship of coiling temperature.

3.1. CNN Module

Convolutional Neural Network (CNN) is a deep learning model characterized by local perception and parameter sharing. It can automatically extract local correlation features from input data through the sliding operation of convolution kernels on the input data [15,16]. Compared with traditional fully connected networks, CNN can reduce the number of network parameters while enhancing the model's ability to represent the local combination relationship of input features, and thus is commonly used for feature extraction of complex nonlinear data [17]. In predicting the coiling temperature of hot-rolled strip steel, significant coupling relationships exist among process parameters such as Strip Thickness, Finish Rolling Temperature, Average Strip Speed, and Open Valves. Therefore, it is difficult to fully capture the coiling temperature variations by relying on a single variable. The CNN module can extract local combination features among multiple input variables through convolution operation, thereby providing more effective feature representations for the subsequent attention mechanism and GRU module. The relevant formula of the convolution layer is shown in Equation (6).

$$Y_{i,j,n} = \sum_{m=1}^k \sum_{l=1}^k X_{i+m-1,j+l-1} \cdot W_{m,l,n} + b_n \quad (6)$$

wherein X denotes the input feature, Y represents the convolutional output feature, W is the convolution kernel weight, b is the bias term, k is the size of the convolution kernel, m is the number of input channels, and n is the number of filters.

To enhance the nonlinear representation capability of the model, the Rectified Linear Unit (ReLU) activation function is introduced after the convolutional layer, and its mathematical expression is given as follows:

$$\text{ReLU}(x) = \max(0, x) \quad (7)$$

The Rectified Linear Unit (ReLU) activation function preserves positive response features and suppresses invalid negative responses, endowing the network with stronger nonlinear fitting capability while contributing to improved training efficiency [18]. In this paper, a two-layer convolutional structure is adopted in the Convolutional Neural Network (CNN) module to extract features from input working condition parameters. The first convolutional layer primarily completes the preliminary mapping of input features, while the second convolutional layer further extracts deeper local combination features. After processing by the CNN module, the original working condition parameters are converted into a multi-channel feature representation, which lays a foundation for the channel weight allocation of the subsequent Squeeze-and-Excitation (SE) attention mechanism and the deep feature learning of the Gated Recurrent Unit (GRU) module.

3.2. SE Attention Mechanism Module

To further enhance the expressive capability of the model for key features, this paper introduces the Squeeze-and-Excitation (SE) attention mechanism after the CNN feature extraction module [19]. As a typical channel attention architecture, the core idea of the SE module is to adaptively recalibrate the multi-channel features extracted by CNN by learning the importance weights of features across different channels, thereby highlighting the key features that contribute more significantly to coiling temperature prediction and suppressing redundant or interfering information [20]. As opposed to directly inputting CNN output features into the GRU, the SE module completes weight allocation before features enter the GRU, enabling the subsequent network to pay more attention to the feature information closely related to coiling temperature changes. The structural diagram of the SE attention mechanism module is shown in Figure 3.

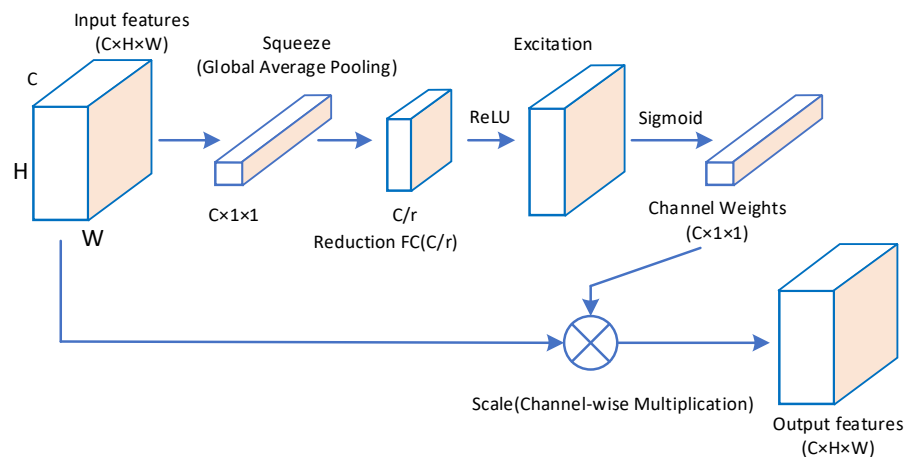


Figure 3. SE attention mechanism module structure.

The Squeeze-and-Excitation (SE) module mainly consists of three processes: Squeeze, Excitation, and Recalibration [21]. First, in the squeeze stage, global average pooling is performed on the feature maps output by the CNN, and the input is the multi-channel feature maps obtained after the second layer of convolution and ReLU activation. According to the CNN structure proposed in this paper, the number of channels of these feature maps is 64. The SE module adaptively recalibrates the features output by the CNN by learning the weight of each channel. The spatial features on each channel are compressed into a global statistic to obtain the overall response information of each channel. For the c -th channel, the compression process can be expressed as:

$$z_c = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W u_c(i, j) \quad (8)$$

In the equation, H and W denote the height and width of the feature map respectively, $u_c(i, j)$ represents the feature value of the c -th channel at position (i, j) , and z_c denotes the global descriptive information of the c -th channel.

In the excitation stage, the channel description vector is fed into a two-layer fully connected network, and the weight coefficients of each channel are generated via the Sigmoid function. The calculation process is formulated as follows:

$$s = \sigma(W_2 \delta(W_1 z)) \quad (9)$$

In the equation, z denotes the channel descriptor vector obtained via global average pooling, W_1 and W_2 are the weights of the fully connected layers, $\delta(\cdot)$ represents the Rectified Linear Unit (ReLU) activation function, $\sigma(\cdot)$ denotes the Sigmoid activation function, and s is the resulting channel weight vector.

During the rescaling stage, the generated channel weights are multiplied element-wise channel-by-channel with the original convolutional features, to obtain the SE-weighted feature output:

$$\tilde{u}_c = s_c \cdot u_c \quad (10)$$

Among them, s_c denotes the weight coefficient corresponding to the c -th channel, and $\tilde{u}_c$ denotes the weighted feature of the c -th channel. After processing by the Squeeze-and-Excitation (SE) module, the channel features with strong correlation to coiling temperature prediction are enhanced, while redundant or interfering features are suppressed, thereby providing more discriminative weighted feature inputs for the subsequent Gated Recurrent Unit (GRU) module.

3.3. GRU Module

Gated Recurrent Unit (GRU) is a neural network structure with gating regulation capability [22,23]. Compared with traditional recurrent neural networks, GRU controls the information transmission process through update gates and reset gates, which can effectively realize the screening, retention and updating of feature information while reducing the number of model parameters [24]. In this paper, GRU is introduced after the convolutional neural network (CNN) and the squeeze-and-excitation (SE) attention mechanism to conduct deep learning and fusion on the weighted features, so as to further explore the complex nonlinear mapping relationship between multiple process parameters and coiling temperature.

The main computational procedure of the GRU is shown in Equations (11)–(14).

$$r_t = \sigma(W_r \cdot [h_{t-1}, x_t] + b_r) \quad (11)$$

$$z_t = \sigma(W_z \cdot [h_{t-1}, x_t] + b_z) \quad (12)$$

$$\tilde{h}_t = \tanh(W_{\tilde{h}} \cdot [r_t * h_{t-1}, x_t] + b_{\tilde{h}}) \quad (13)$$

$$h_t = (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t \quad (14)$$

In the aforementioned equation, σ is the sigmoid activation function, $\tanh$ denotes the hyperbolic tangent activation function, z_t is the update gate, r_t is the reset gate, W represents the weight matrix, and b denotes the bias term. The reset gate controls how much information from the previous feature state is incorporated into the calculation of the current candidate state, while the update gate regulates the retention ratio between existing information and current candidate information. Through the gating mechanism, the Gated Recurrent Unit (GRU) can further screen and fuse the Squeeze-and-Excitation (SE) weighted features, enabling the model to more sufficiently excavate the nonlinear mapping relationship between process parameters and coiling temperature. Finally, the deep features output by the GRU are fed into the fully connected layer to generate the prediction result of coiling temperature.

3.4. WOA Hyperparameter Optimization Module

The predictive performance of deep learning models is substantially influenced by hyperparameters including learning rate, number of hidden units, and regularization coefficient [25,26]. The conventional empirical trial-and-error method not only incurs high computational cost, but also struggles to obtain optimal parameter combinations in complex parameter spaces [27]. To improve the prediction accuracy and training stability of the proposed CNN-GRU-SE model, this paper introduces the Whale Optimization Algorithm (WOA) to perform optimization searching for the key hyperparameters of the model.

WOA (Whale Optimization Algorithm) is a swarm intelligence optimization algorithm that simulates the foraging behavior of humpback whales. It achieves the balance between global exploration and local exploitation primarily through mechanisms including prey encirclement, spiral updating and random search [28]. In the model proposed in this paper, each individual whale represents a set of candidate hyperparameter combinations:

$$X_i = [\alpha_i, H_i, \lambda_i] \quad (15)$$

In this equation, α_i denotes the learning rate corresponding to the i -th whale individual, H_i represents the number of GRU hidden units, and λ_i is the L2 regularization coefficient. The whale optimization algorithm (WOA) initializes the whale population within the specified search range, and substitutes each set of candidate parameters into the CNN-GRU-SE model for training.

Based on the model training process of this paper, the normalized mean square error of the training set is adopted as the fitness function to evaluate the performance of different parameter combinations. A smaller fitness value indicates a smaller prediction error of the model under this set of hyperparameters, which corresponds to a better parameter combination. During the iteration, the Whale Optimization Algorithm continuously updates the positions of whales based on individual fitness and records the current optimal individual. After the maximum number of iterations is reached, the optimal hyperparameter combination is output as follows:

$$X^* = [\alpha^*, H^*, \lambda^*] \quad (16)$$

In the formula, α^* , H^* and λ^* denote the optimal learning rate, the optimal number of GRU hidden units, and the optimal L2 regularization coefficient obtained via WOA search, respectively. Finally, the optimal parameter combination is input into the

CNN-GRU-SE prediction network to complete the final model training and prediction on the test set.

4. Experimental Setup and Result Analysis

4.1. Performance Evaluation Metrics

To comprehensively evaluate the prediction performance of each model, this paper adopts root mean square error (RMSE), mean absolute error (MAE), mean absolute percentage error (MAPE) and coefficient of determination (R^2) as evaluation indicators. Meanwhile, in accordance with the requirements of industrial applications, this paper further calculates the hit rate of prediction errors falling within the ranges of ± 2 °C, ± 4 °C, ± 6 °C, ± 8 °C and ± 10 °C, which is used to measure the engineering applicability of the models under different accuracy requirements. The specific formulas are as follows:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (17)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (18)$$

$$MAPE = \frac{100\%}{n} \sum_{i=1}^n \frac{|y_i - \hat{y}_i|}{y_i} \quad (19)$$

$$R^2 = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y})^2} \quad (20)$$

Let the true value be y_i , the predicted value be $\hat{y}_i$, and the number of samples be n . A lower value of MAE (Mean Absolute Error) or RMSE (Root Mean Square Error) indicates a smaller prediction error, while a coefficient of determination R^2 closer to 1 demonstrates a better goodness-of-fit of the model.

4.2. Parameter Settings

The experimental environment in this study is MATLAB 2025a, the operating platform is a 64-bit Windows operating system, and the processor is Intel Core i5-1035G1. For the proposed fusion model, the population size and the maximum number of iterations of the Whale Optimization Algorithm (WOA) are both set to 30. The parameters to be optimized include the learning rate, the number of GRU hidden units, and the L2 regularization coefficient. The convolutional neural network (CNN) adopts a two-layer convolution structure, where the sizes of convolution kernels are 1×1 and 2×1 respectively, and the numbers of convolution kernels are 32 and 64 respectively. A batch normalization layer and a ReLU activation layer are added after each convolution layer. During the training process, the maximum number of training epochs is set to 500, the batch size is set to 32, and the Adam optimizer is adopted for model training.

4.3. Comparison of Optimization Algorithm Performance

To validate the effectiveness of WOA in optimizing hyperparameters for the CNN-GRU-SE model, this study selects the particle swarm optimization (PSO) algorithm as a comparative method. Under identical datasets, data partitioning schemes, network architectures, and fitness functions, both WOA and PSO are employed to optimize the key hyperparameters of the model. The population size for both algorithms is set to 20, and

the maximum number of iterations is set to 20. The parameters being optimized include the learning rate, the number of GRU hidden units, and the L2 regularization coefficient, with search ranges defined as $[1 \times 10^{-4}, 5 \times 10^{-3}]$, $[32, 256]$, and $[1 \times 10^{-6}, 1 \times 10^{-4}]$, respectively. The normalized root mean square error (RMSE) on the validation set is used as the fitness function; a lower fitness value indicates a smaller prediction error of the corresponding model with that set of hyperparameters on the validation set.

The fitness convergence curves of the two optimization algorithms are shown in Figure 4. As can be seen from Figure 4, both Particle Swarm Optimization (PSO) and Whale Optimization Algorithm (WOA) can reduce the fitness value during the iteration process, indicating that both optimization algorithms possess certain hyperparameter optimization capability. Specifically, WOA achieves a faster decline in fitness during the early iteration stage: its fitness value decreases significantly after the second iteration and is further optimized in subsequent iterations. In comparison, the fitness value of PSO decreases at a relatively slower rate and enters a stable state earlier. In terms of the final convergence result, the fitness value obtained by WOA is lower than that of PSO, which demonstrates that WOA has better convergence performance and optimization effect in the hyperparameter optimization process of the proposed CNN-GRU-SE model in this paper.

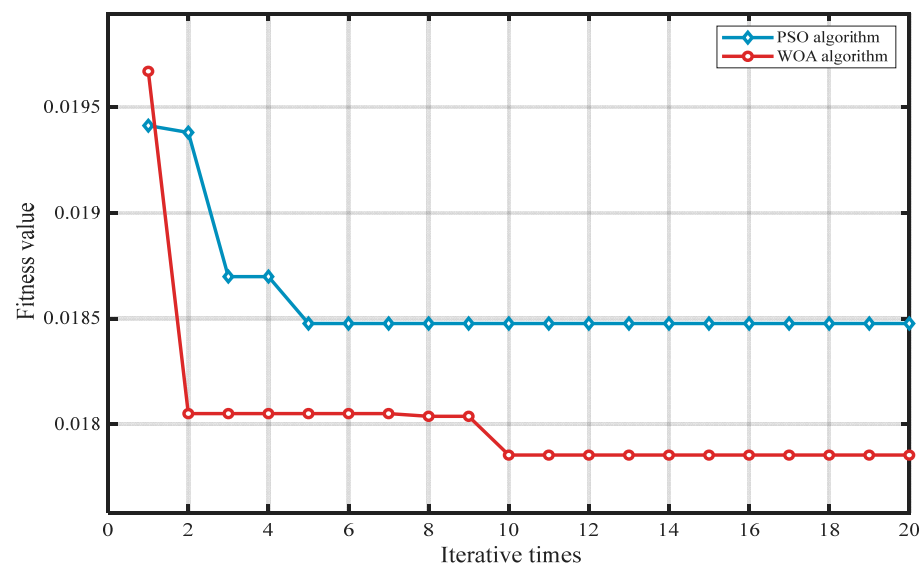


Figure 4. Fitness convergence curves of WOA and PSO.

4.4. Ablation Study Analysis

To verify the effectiveness of the proposed WOA-CNN-GRU-SE model in predicting coiling temperature, three models—CNN-GRU, CNN-GRU-SE, and WOA-CNN-GRU-SE were selected for comparative analysis under identical dataset conditions. To ensure fairness in comparison, all models employed the same data preprocessing methods, consistent data partitioning strategies, and uniform training and testing procedures, with performance evaluated using identical metrics. The prediction results on the test set are shown in Figure 5. From the prediction curves, it is evident that all three models effectively capture the overall trend of coiling temperature changes. The inclusion of the SE attention mechanism leads to a slight improvement in prediction accuracy. However, in regions with significant local fluctuations or dense temperature samples, the WOA-CNN-GRU-SE model demonstrates superior prediction accuracy compared to the other two models.

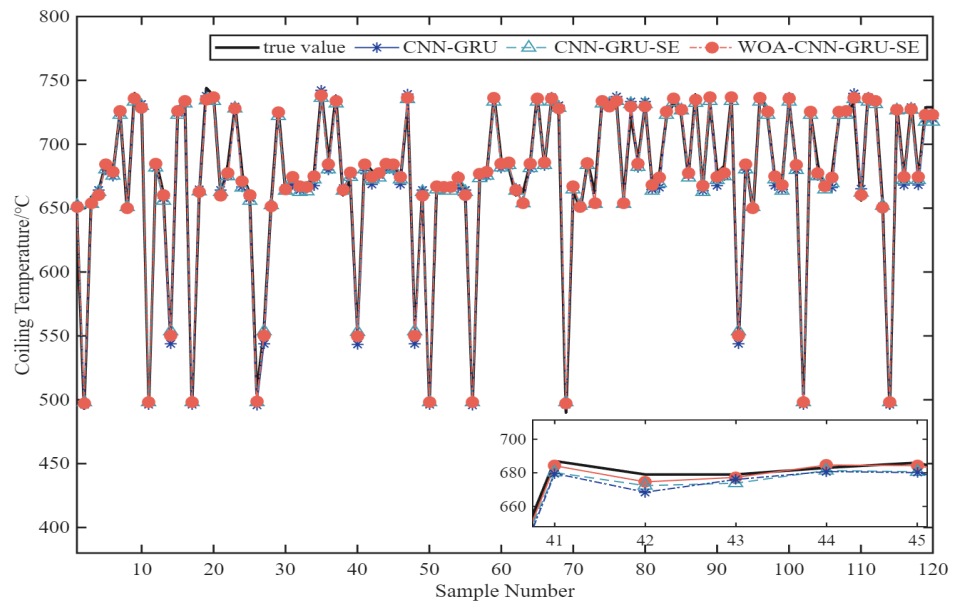


Figure 5. Ablation experiments results.

To objectively compare the prediction accuracy of multiple models, the evaluation results of the multi-models are shown in Table 2. Compared with the CNN-GRU-SE and CNN-GRU models, the proposed WOA-CNN-GRU-SE attention mechanism model reduced the root mean square error by 26.17% and 32.52%, respectively. The mean absolute error decreased by 23.26% and 30.98%, while the mean absolute percentage error was reduced by 22.44% and 31.41%. The coefficient of determination improved from 0.9937 to 0.9971, indicating that the proposed model achieves better fitting performance than the other two models.

Table 2. Ablation comparison.

Model	RMSE	MAE	MAPE/%	R ²
WOA-CNN-GRU-SE	3.4477	2.8864	0.4363	0.9971
CNN-GRU-SE	4.6699	3.7612	0.5625	0.9947
CNN-GRU	5.1093	4.1821	0.6361	0.9937

Using only metrics such as RMSE, MAE, and R² can reflect the overall prediction performance of a model from an error perspective. However, for engineering applications like coiling temperature prediction, average error indicators alone are insufficient to fully capture the model's accuracy across different error tolerances. To further evaluate the practical applicability of each model, this study calculates the hit rates of all models under various temperature error thresholds, as shown in Figure 6. As the error threshold increases, the hit rates of all models generally rise. The WOA-CNN-GRU-SE model maintains a high hit rate across all error thresholds, demonstrating a particularly significant advantage in medium-to-low error ranges. Within a ± 6 °C error range, the WOA-CNN-GRU-SE model achieves a hit rate of 92.56%, outperforming CNN-GRU-SE (77.69%) and CNN-GRU (75.21%). In the ± 8 °C error range, the hit rate of our proposed model approaches 100%. This indicates that the model not only performs well in overall error metrics but also exhibits superior stability and practical value under high-precision prediction requirements.

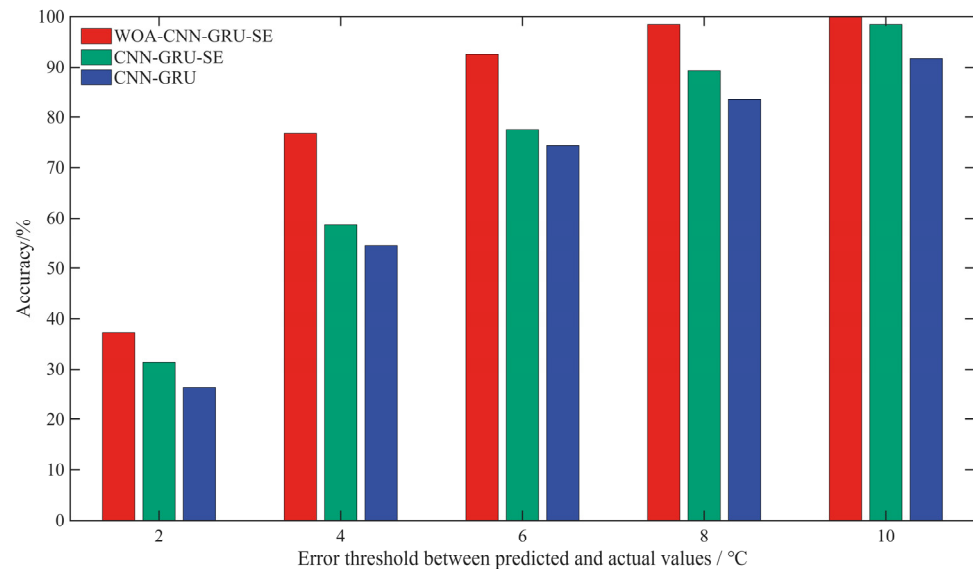


Figure 6. Comparison of prediction accuracy of ablation experiments.

To further evaluate the reliability of the improvements in the proposed model, we performed a statistical significance analysis based on the prediction errors of test samples. The standard deviations of the absolute prediction errors of WOA-CNN-GRU-SE and CNN-GRU-SE are 1.8935 and 2.7794, respectively, demonstrating that the proposed model achieves more stable prediction performance. Furthermore, we employed the Wilcoxon signed-rank test to analyze the difference between the two models. The obtained *p*-value is less than 0.05, indicating that the improvements in the proposed model are statistically significant.

4.5. Multi-Model Comparative Analysis

To verify the superiority of the proposed model over other typical prediction methods, LSTM (Long Short-Term Memory), ELM (Extreme Learning Machine), and RF (Random Forest) models were selected as benchmark models. The prediction results of each model on the test set are shown in Figure 7, and the evaluation results across multiple models are presented in Table 3. From the figure, it can be observed that while LSTM, ELM, and RF models are able to capture the general trend of coiling temperature to some extent, their predictions still exhibit noticeable deviations from the actual values. In contrast, the WOA-CNN-GRU-SE model demonstrates a prediction curve that more closely follows the true values. Further analysis based on the evaluation metrics in Table 3 reveals that compared with LSTM, RF, and ELM models, the WOA-CNN-GRU-SE attention mechanism model developed in this study reduces the root mean square error by 42.35%, 46.15%, and 52.12%, respectively. The mean absolute error is reduced by 41.20%, 45.36%, and 48.14%, respectively. The mean absolute percentage error decreases by 42.2%, 45.33%, and 47.57%, respectively. Moreover, the coefficient of determination increases from 0.9874 to 0.9971, indicating that the proposed model achieves better fitting performance than the other three models.

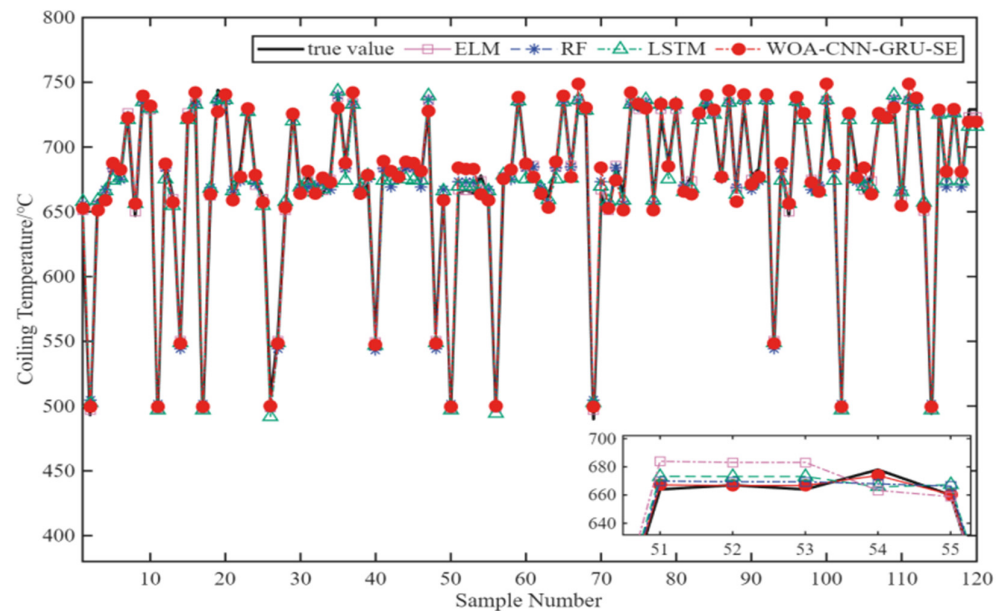


Figure 7. Prediction of different models.

Table 3. Prediction comparison of different models.

Model	RMSE	MAE	MAPE/%	R ²
WOA-CNN-GRU-SE	3.4477	2.8864	0.4363	0.9971
LSTM	5.9799	4.9091	0.7548	0.9913
RF	6.4027	5.2822	0.7980	0.9901
ELM	7.2006	5.5655	0.8322	0.9874

Based on comprehensive evaluation metrics, the WOA-CNN-GRU-SE model demonstrates the best overall prediction performance, followed by the LSTM model, while the ELM model exhibits relatively larger errors. To further compare the differences in prediction accuracy among models in engineering applications, the hit rates of each model under different temperature error thresholds are shown in Figure 8. As the error tolerance increases, the hit rates of all models improve. Within a ± 6 °C error range, the WOA-CNN-GRU-SE model achieves a hit rate of 92.56%, outperforming LSTM at 67.77%, RF at 61.98%, and ELM at 64.46%. In contrast, the proposed model consistently achieves the highest hit rates across all error thresholds, with a particularly significant advantage within the stricter ± 6 °C error range. As the error threshold continues to widen, the gap in hit rates between models gradually narrows, yet the proposed model remains superior, indicating not only high prediction accuracy but also strong stability and practical value for engineering applications.

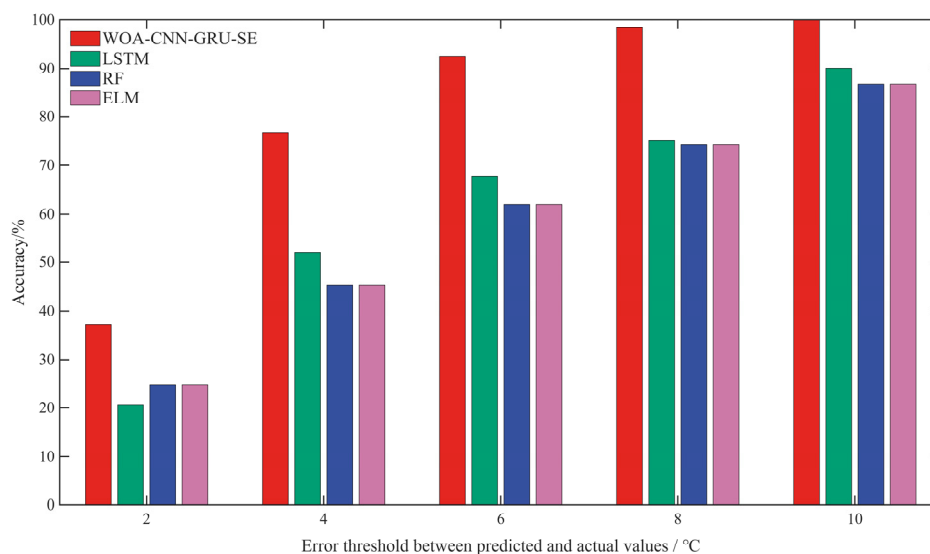


Figure 8. Comparison of prediction accuracy.

5. Conclusions

Aiming to address the challenges of multi-variable coupling, strong nonlinearity, low prediction accuracy of traditional models and difficult hyperparameter tuning in hot-rolled strip coiling temperature prediction, this paper proposes the WOA-CNN-GRU-SE model. Experimental results show that the SE attention mechanism can effectively screen key features and improve the feature representation ability of the network. Compared with PSO, WOA achieves better convergence performance and can quickly obtain the optimal hyperparameter combination. Ablation studies verify that both the SE module and WOA optimization strategy can greatly improve prediction accuracy. Compared with CNN-GRU, LSTM, RF and ELM, the proposed model achieves lower prediction errors and a hit rate of 92.56% within the industrial error range of ± 6 °C, showing outstanding overall performance. The proposed model effectively addresses the limitations of traditional methods in complex industrial environments. Its prediction accuracy and stability meet the requirements of practical production, which can provide a useful reference for intelligent control of coiling temperature in hot rolling lines.

Although the proposed model demonstrates high prediction accuracy, the dataset lacks geometric dimensions such as strip width, cooling medium parameters like coolant flow rate and temperature, and environmental variables, which limits the model's current physical interpretability. In future work, we will collect comprehensive production data including full-dimensional process variables and conduct multiple repeated experiments for validation, thereby further enhancing the model's physical interpretability and cross-seasonal generalization capability to adapt to more complex rolling conditions.

Author Contributions: Conceptualization, T.S. and H.C.; methodology, H.C. and X.Z.; software, L.S.; validation, Z.M., X.Z. and Y.C.; writing—original draft, X.Z. and Y.C.; writing—review and editing, T.S. and H.C. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by Science and Technology Research Project of the Education Department of Jilin Province under Grant JJKH20230064KJ and JJKH20261143KJ, and Jilin Provincial Department of Science and Technology under Grant 20250102141JC.

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Fu, Z.X.; Yang, G.W.; Han, R.Y.; Xu, Y.W.; Mao, X.P.; Bao, S.Q.; Zhao, G. Influence of coiling temperature on microstructure and mechanical properties of a hot-rolled high-strength steel microalloyed with Ti, Mo and V. *J. Iron Steel Res. Int.* **2022**, *29*, 484–493.
2. Singh, P.P.; Ghosh, S.; Mula, S. Strengthening behaviour and failure analysis of hot-rolled Nb+ V microalloyed steel processed at various coiling temperatures. *Mater. Sci. Eng. A* **2022**, *859*, 144210.
3. Cui, K.Y.; Li, Z.R.; Li, H.B.; Chen, S.; Wang, C.W.; Xiong, X.G.; Liu, Y.B. Effects of finishing rolling temperature and coiling temperature on microstructure and properties of V-Ti microalloyed weathering steel. *Heat Treat. Met.* **2026**, *51*, 179–188.
4. Wang, Y.D.; Tang, Z.H.; Xiao, S.F.; Siyasiya, C.W.; Wei, T. Effects of final rolling temperature and coiling temperature on precipitates and microstructure of high-strength low-alloy pipeline steel. *J. Iron Steel Res. Int.* **2022**, *29*, 1236–1244.
5. Sun, N.; Yu, H.; Yang, L.M.; Li, T.; Xu, Z.Y.; Li, X.H.; Wang, J.T. Influence of hot rolling curling temperatures on microstructure and texture of Al-Mg-Si alloy. *Forg. Stamp. Technol.* **2025**, *50*, 139–150.
6. Sun, T.J.; Wang, M.R.; Liu, B.; Cui, W.C.; Li, P.W.; Qu, L.P. Multi-objective optimization of cooling after hot-rolling of strip steel based on big data mining technology. *J. Beihua Univ. (Nat. Sci.)* **2023**, *24*, 820–827.
7. Wang, Q.B.; Ning, X.Y.; Zhang, J.Y.; Zhang, H.J. Prediction model for post-roll cooling coil temperature based on data and mechanism fusion method. *Metall. Ind. Autom.* **2025**, *49*, 12–18.
8. Peng, L.-G.; Wang, D.-G.; Li, J.; Xing, J.-F.; Gong, D.-Y. Data-driven adaptive setting algorithm for coiling temperature model parameter. *Chin. J. Eng.* **2020**, *42*, 778–786.
9. Hou, J.; Li, J.; Tang, J. A coiling temperature prediction model for the laminar cooling process of hot rolled strip based on ICS-SVR. In *Proceedings of the 2025 IEEE 5th International Conference on Data Science and Computer Application (ICDSCA)*; IEEE: New York, NY, USA, 2025; pp. 349–354.
10. Sun, T.J.; Yang, W.D.; Cheng, Y.M.; Duan, F.-Y. Improved genetic algorithm for optimizing prediction model of strip coiling temperature. *Control Theory Appl.* **2015**, *32*, 1106–1113.
11. Zhang, S.; Wang, J.J.; Li, A.L.; Cui, G.M. Improved GWO-ELM based hot rolled strip coiling temperature prediction. *Electron. Meas. Technol.* **2021**, *44*, 50–55.
12. Hao, Q.Y.; Gong, D.Y.; Tian, B.Q.; Ding, L.; Xu, J. Coiling temperature prediction of hot rolling strip based on random forest. *Metall. Ind. Autom.* **2023**, *47*, 93–102.
13. Zhou, J.X.; Huo, T.M. Prediction model of coiling temperature based on NMWOA-LSTM. *Electron. Meas. Technol.* **2023**, *46*, 60–66.
14. Standardization Administration of the People's Republic of China. *GB/T 700-2006*; Carbon Structural Steels. Standards Press of China: Beijing, China, 2006.
15. Wang, M.; Guo, X.F.; Cao, X.M.; Tian, Z.K.; Pan, F.L. Sea surface temperature prediction based on MPPO-CNN-LSTM-Attention model. *Mar. Sci.* **2026**, 1–16. Available online: <https://link.cnki.net/urlid/37.1151.p.20260203.1000.002> (accessed on 16 June 2026).
16. Zha, W.; Liu, Y.; Wan, Y.; Luo, R.; Li, D.; Yang, S.; Xu, Y. Forecasting monthly gas field production based on the CNN-LSTM model. *Energy* **2022**, *260*, 124889.
17. Zheng, Y.; Hu, J.; Chen, J.; Deng, H.; Hu, W. State of health estimation for lithium battery random charging process based on CNN-GRU method. *Energy Rep.* **2023**, *9*, 1–10.
18. Purwono, P.; Ma'arif, A.; Rahmانيar, W.; Fathurrahman, H.I.K.; Frisky, A.Z.K.; Haq, Q.M.U. Understanding of convolutional neural network (CNN): A review. *Int. J. Robot. Control Syst.* **2022**, *2*, 739–748.
19. Han, M.B.; Wang, H.N.; Zheng, H.W.; Liu, D.L.; Li, Q. Big data-driven CNN-GRU-Attention model for blast furnace hot metal temperature prediction. *China Metall.* **2025**, *35*, 168–177.
20. Li, Y.H.; Hao, S.Z.; Lu, J.J.; Zhang, Y. Mid-long term runoff intelligent prediction research based on CNN-SE Attention-GRU coupling mechanism. *J. Huazhong Univ. Sci. Technol. (Nat. Sci. Ed.)* **2026**, 1–7.
21. Wang, H.; Song, R.; He, W.; Cai, J.J.; Long, Z.Y.; Cong, M. Railway container transportation time prediction based on CNN-GRU-Attention model. *J. Beijing Jiao Tong Univ.* **2025**, *49*, 115–121.
22. Xie, Y.; Wang, J.; Li, H.; Dong, A.; Kang, Y.; Zhu, J.; Wang, Y.; Yang, Y. Deep learning CNN-GRU method for GNSS deformation monitoring prediction. *Appl. Sci.* **2024**, *14*, 4004.
23. Mahjoub, S.; Chrifi-Alaoui, L.; Marhic, B.; Delahoche, L. Predicting energy consumption using LSTM, multi-layer GRU and drop-GRU neural networks. *Sensors* **2022**, *22*, 4062.
24. Li, Z.G.; Zhang, X. CNN-GRU model predicts blast furnace gas production. *Mach. Des. Manuf.* **2022**, *4*, 69–72.
25. Xiao, Y.; Yin, Y.; Xu, J.; Zhang, Y. Transformer core loosening diagnosis based on fusion feature extraction and CPO-optimized CatBoost. *Processes* **2025**, *13*, 3247.

26. Li, K.L.; Wang, J.B.; Shen, Y.J.; Chen, B. Prediction of photovoltaic power generation based on CEEMD-WOA-LSTM. *J. Eng. Therm. Energy Power* **2025**, *40*, 136–147.
27. Yang, X.; Li, Z.; Cao, L.; Chen, L.; Huang, Q.; Bi, G. Process optimization and quality prediction of laser aided additive manufacturing SS 420 based on RSM and WOA-Bi-LSTM. *Mater. Today Commun.* **2024**, *38*, 107882.
28. Yang, Y.P.; Chen, Z.G.; Yu, Z.H.; Wang, Y.; Chen, L. Fault diagnosis of rolling bearings based on WOA-VMD and PSO-SVM. *Manuf. Technol. Mach. Tool* **2025**, *2*, 23–29.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.