Research Article

Quality Prediction of Strip in Finishing Rolling Process Based on GBDBN-ELM

Shuang Li , Jian Wang, and Sen Chen

Computer Integrated Manufacturing System Research Center, College of Electronics and Information Engineering, Tongji University, Shanghai 201800, China

Correspondence should be addressed to Shuang Li; 1930746@tongji.edu.cn

Received 18 March 2021; Revised 10 April 2021; Accepted 20 April 2021; Published 21 May 2021

Academic Editor: Aijun Yin

Copyright © 2021 Shuang Li et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

With the continuous development of the manufacturing industry, the requirement for strip steel quality is becoming higher and higher in automobile manufacturing, mechanical processing, and electronic and electrical industries. The precise control of strip quality depends on the accurate prediction of strip quality to a certain extent. However, the data collected by a large number of sensors on the complex strip production line and generated by the computer control system presents the characteristics of high dimensionality, high coupling, and nonlinearity, which brings difficulties to the prediction of strip quality. The continuous production of massive data in the production line also forces steel enterprises to seek new data mining methods, mining the relationship between sensor data to predict and control strip quality. To solve these problems, this paper proposes a GBDBN-ELM model, which is more efficient and more accurate than other algorithms. In this model, the RBM in DBN is replaced with GBRBM, so that RBM no longer depends on the binary distribution, can handle continuity values, and retain more data features. In order to solve the problem of too long DBN training time, this article replaces the BP network in DBN with an ELM regression model. The ELM model predicts the strip quality based on the extracted data abstract features, thereby improving the model's prediction accuracy and shortening the training time. In this paper, the GBDBN-ELM model is compared with the BP neural network, ELM, and DBN, and root mean square error, R square coefficient of determination, and training time are selected as evaluation indexes of the models. The experimental results show that the improved GBDBN-ELM model can not only improve the accuracy of strip steel quality prediction but also shorten the time of model training. The model proposed in this paper has achieved good results in prediction accuracy and performance.

1. Introduction

In the industrial field, the steel industry is one of the national basic industries. Most of the raw materials, resources, and equipment of other industries are provided by the steel industry. The development of the steel industry has also led to the progress of construction, machinery, transportation, and other industries. Although the current international steel production is increasing year by year, the technology for rolling high-quality steel still needs to be improved. With the rapid development of industry and technology, many industries have higher and higher requirements for the quality of strip steel, such as infrastructure engineering, automobile manufacturing, mechanical processing, and electronic and electrical industries. Therefore, the improvement of strip

quality has become one of the main tasks of the hot rolling production process. The strip quality can be estimated in advance through prediction, and then, the process parameters can be adjusted in time through computer calculations to achieve closed-loop control of the system, which can maximize the strip quality. Therefore, the strip quality prediction method has gradually become a hot spot in the steel industry.

The traditional rolling mill control relies on manual operation; the strip quality at the exit is controlled by simple electric pressing or manual pressing, without the participation of many sensors. The steel industry has bid farewell to traditional production modes with the extensive application and development of modern automatic control theory in the industrial field. The combination of modern equipment and advanced technology has made the strip production

process increasingly complex [1]. In the strip steel production process, multiple devices are organically connected. The process parameters and product quality parameters involved in the subproduction stage are various, and the relationship between the various parameters is complex. The parameters of each stage often present a hierarchical structure and are coupled with each other. It is difficult to describe these parameters with linear or simple nonlinear relationships [2].

Quality of the strips at the exit of hot continuous rolling mainly depends on the finishing mill. The change of strip width and thickness is caused by the rolling force from the vertical stand and horizontal stand in finishing rolling. The main factors affecting strip steel quality include rolling force, reduction position, inlet temperature, roll bending force, roll gap width, and rack speed. Moreover, factors such as water flow, motor current, oil film compensation, and lubrication also have a certain impact on the surface quality of the strip [3]. Most of these variables are coupled with each other and have serious nonlinearity, and some of them are difficult to measure, which also brings certain difficulties to the prediction of strip quality.

Moreover, with the development of technology, the production model has spread from physical space to virtual space, and the degree of digital production has gradually deepened. More sensors, data acquisition equipment, and computer network control system are involved in the production process. A large amount of raw data are produced in the strip production line every day. How to use these data reasonably and mine more knowledge for strip quality prediction and control is also a problem that needs to be studied.

In order to achieve closed-loop control of strip quality and advance adjustment of process parameters, and to solve the poor prediction accuracy of strip quality resulting caused by high-dimensional, highly coupled, nonlinear data, the main contributions of this paper are as follows.

This paper proposes a strip quality prediction model combining DBN and ELM. In the combined model, DBN is used to extract features from high-dimensional and high coupling input data, and ELM predicts strip quality according to the extracted data features.

Based on the DBN-ELM combined model, the RBM in the DBN model is replaced by GBRBM to solve the dependence on the binary distribution of the visible layer and hidden layer of RBM. The model is improved to the GBRBM-ELM model to suit the continuous value regression problem.

The feasibility of the model is analyzed from the aspects of prediction accuracy and model performance, and the prediction effect of the model is compared with that of BP, ELM, and DBN. The results show that the GBDBN-ELM model can improve the prediction accuracy while shortening the model training time.

The rest of this article is organized as follows: The second chapter introduces the research progress of strip quality prediction technology; the third chapter introduces the principle, network structure, and training method of the strip quality prediction model; the fourth chapter validates the

model through the data on the production line of a steel company; the last chapter is a summary of this article.

2. Related Works

Quality prediction and quality control problems often use two types of methods, mathematical model methods and data mining techniques.

2.1. Mathematical Model Methods. In traditional quality control, the mathematical model is used to predict the quality parameters, and the variables such as temperature, pressure, element, and their relationship are described by mathematical formulas [4, 5]. But the process of establishing the mathematical model of strip quality is very complex, because in the rolling process, not only many physical quantities but also much thermodynamic knowledge is involved. The strip quality prediction based on the mathematical model ignores and simplifies the influence of many on-site factors, and the dynamic effects sometimes produce false results, leading to large errors. Later, people paid attention to deformation laws in the forming process, and the finite element method and finite element simulation software were applied to the simulation of the strip production process and the quality control of the strip [6, 7]. However, the finite element analysis software has high requirements for the user's ability, and it often requires level-by-level training to use it proficiently.

On the other hand, the growing mass of data has made data mining methods centered on machine learning and deep learning more attention [8]. And it is used in manufacturing production scheduling [9], equipment monitoring [10], quality control [11], and other aspects. The data mining methods provide an effective way to predict and control the quality of hot strip rolling. It can break the data island, deeply mine, and utilize the data value. Specifically, initially, through data mining, the correlation between process parameters and quality parameters is discovered from the massive production history data. Further, the strip quality is predicted through these correlations. Finally, combined with the computer control system, a closed loop is formed to the greatest extent to control the quality of the strip.

2.2. Data Mining Techniques. Data mining techniques can be divided into two parts: machine learning and deep learning. Kotkunde et al. used artificial neural networks (ANN) and support vector machines (SVM) to evaluate the thickness distribution of alloy sheets at various temperatures and blank diameters [12]. Li and Dai used the k -means algorithm to divide the production data into k clusters and uses the BP neural network to predict the final strip rolling temperature to improve the prediction accuracy [13]. Wu et al. improved the ELM algorithm and created a two-hidden layer optimized ELM model, and they applied it to the prediction of bending force in the hot strip rolling process [14, 15].

The above studies are based on pure machine learning predictions, but machine learning cannot handle high-dimensional problems. Before using the machine learning method, the above research often needs to select data features to reduce the dimension of the input parameters [16].

However, for the high coupling strip quality prediction problem, too few features often cannot contain all the features of the data, which makes the prediction accuracy worse. The deep learning network has deeper network layers and more complex network structure compared with machine learning. For example, the deep belief network (DBN) superimposes RBM in front of the BP network. DBN has good feature extraction ability and shows good performance when processing high-dimensional input variables. It is widely used in manufacturing.

Liu et al. used DBN to process a large amount of real-time quality data collected by sensors and constructed a real-time quality monitoring and diagnosis plan for the manufacturing process [17]. Yang and Frangopol predict the remaining life cycle of ships based on DBN and then propose a ship life cycle management framework [18]. Zheng et al. combine DBN and SVM, use DBN to extract high-level features of signals, and use the SVM classifier for defect recognition, providing a new method for the nondestructive testing of bolt anchorage [19].

It can be concluded from a large number of studies that since deep belief networks were primarily used to classify problems at first, now it is mainly used for classification problems such as defect identification and quality classification in manufacturing quality problems and less often applied to regression problems such as quality parameter prediction. However, because the deep belief network has strong high-dimensional feature extraction capabilities and good model generalization, it can be improved to suit continuous value prediction problems on the basis of maintaining the feature extraction capabilities. For example, in existing research, DBN is combined with Particle Swarm Optimization (PSO) [20], Firefly Algorithm (FA) [21], Support Vector Machine (SVM) [22], Extreme Learning Machine (ELM) [23], and other algorithms to improve the prediction accuracy and model feasibility.

Considering the complexity of the DBN network structure, this paper chooses the combination of DBN and ELM to simplify the DBN training method and shorten the training time while improving the prediction accuracy.

3. Quality Prediction Model and Network Structure

A deep belief network is one of the core algorithms in deep learning. The deep belief network is composed of several restricted boltzmann machines (RBM) and a BP neural network, which can solve the high-dimensional and high-coupling problem well. However, DBN has problems such as unsuitable for continuous value and too long training time. In this paper, DBN is improved to make it more suitable for the quality prediction of the strip finishing process.

3.1. Deep Belief Network Model

3.1.1. Basic Structure of DBN. The deep belief network is composed of multiple series-connected RBMs and a BP neural network. It has a powerful feature learning ability. The structure of the deep belief network for strip steel quality prediction is shown in Figure 1.

FIGURE 1: Deep belief network structure.

The first visible layer V and the second hidden layer h_1 constitute RBM1. The hidden layer of RBM1 is also the visible layer of RBM2, forming RBM2 together with the third hidden layer, and so on for each layer, stacking to form multilayer RBMs. RBM in DBN uses unsupervised learning, mainly used for feature extraction; the BP network uses supervised learning, mainly used for regression and outputting the predicted value of quality parameters.

3.1.2. DBN Training Process. It can be seen from the figure that the training process of DBN is divided into two stages, namely, the forward pretraining stage and the reverse fine-tuning stage. DBN uses a greedy unsupervised learning mechanism to complete layer-by-layer forward training from bottom to top and extracts the abstract features of the bottom-level data as the high-level input, until the features are sent to the top-level regression unit. Then, it calculates the error between the regression result and the real result and uses the back propagation algorithm of the BP network to complete the reverse fine-tuning of the parameters, further reducing the model error and improving the training accuracy of the system.

DBN gives full play to and combines the advantages of RBM and BP neural network, uses multilayer RBM to extract and abstract high-dimensional data, retains important feature information as much as possible, uses the BP network to complete regression, and uses the BP algorithm to fine-tune the parameters of each layer, so as to achieve the optimal state.

3.1.3. Shortcomings of DBN. Although the traditional DBN has particularly good feature extraction capabilities, after analyzing the model, it can be known that the traditional DBN model also has the following shortcomings:

- (a) The visible layer and hidden layer of traditional RBM obey the binary distribution and have a good function of extracting feature signals for discrete data. In the problem of strip quality prediction, the continuous input signals need to be digitized, which leads to the loss of information and reduces the accuracy of the model
- (b) In the process of DBN training, an important parameter that needs to be adjusted is the number of neurons in each hidden layer, which directly affects the prediction accuracy and training time of the model. For the problem of strip quality prediction, the dimension of input data involved is relatively high, so it is more difficult to select the number of neurons
- (c) Since the fine-tuning process of DBN is based on the gradient descent algorithm, the convergence speed of the BP network is relatively slow. In addition, the BP algorithm is a local search algorithm, which may cause the network to fall into a local optimum due to improper selection of the initial network weights, which may lead to network training failures

In order to solve the above problems, this paper introduces Gauss-Bernoulli RBM instead of RBM in traditional DBN to save the signal of continuous input data, introduces particle swarm optimization to calculate the optimal number of neurons in the hidden layer in the process of parameter adjustment, and introduces extreme learning machine to shorten the training time of the model, improve the generalization ability, and avoid falling into local optimization.

3.2. Gaussian-Bernoulli RBM (GBRBM)

3.2.1. Basic Structure of GBRBM. Restricted Boltzmann machine (RBM) is a shallow random generation network proposed by Hinton et al. It is an energy model for unsupervised learning. It divides all neurons into the visible layer and hidden layer. Data is input from the visible layer to express data features. The hidden layer can extract features to express the relationship between input variables, so the hidden layer is also called a feature extractor. The two layers of neurons are fully connected, and there is no connection between the neurons in the same layer.

Suppose $v_1 \sim v_m$ is the visible layer cell node, $h_1 \sim h_n$ is the hidden layer cell node, $a_1 \sim a_m$ is the visible layer node offset, $b_1 \sim b_n$ is the hidden layer node offset, and $w_{n \times m}$ is the weight matrix between the visible layer and the hidden layer.

When the state of (v, h) is determined, the energy function of RBM can be defined as

$$E(v, h | \theta) = -a^T v - b^T h - v^T w_{n \times m} h. \quad (1)$$

The visible layer and hidden layer of the traditional RBM are limited by the binary distribution [24], which has a good performance when dealing with classification problems. But Boolean variables are no longer suitable for the calculation of continuous data when dealing with regression problems. Therefore, this paper introduces GBRBM when carrying out strip quality prediction.

Gaussian-Bernoulli RBM (GBRBM) is a restricted Boltzmann machine for nonbinomial data proposed by Krizhevsky and Hinton. GBRBM introduces Gaussian function between visible and hidden elements to process continuous numbers between 0 and 1. The energy function expression of GBRBM is as follows:

$$E(v, h | \theta) = \sum_{i=1}^m \frac{(v_i - a_i)^2}{2\sigma_i^2} - \sum_{j=1}^n b_j h_j - \sum_{i=1}^m \sum_{j=1}^n \frac{v_i}{\sigma_i} w_{ij} h_j. \quad (2)$$

The lower the energy of the system is, the more stable the system is and the smaller the error of quality parameter prediction results is. In equation (2), θ is the parameter to be solved, $\theta = (w, a, b, \sigma)$, and σ_i is the tolerance corresponding to v_i . When θ is determined, the joint probability distribution of (v, h) can be obtained through the energy function:

$$P(v, h | \theta) = \frac{e^{-E(v, h | \theta)}}{Z(\theta)}. \quad (3)$$

In equation (3), $Z(\theta) = \sum_{v, h} e^{-E(v, h | \theta)}$ is the normalization factor, also called the distribution function.

Since there is no connection between neurons in the same layer of RBM, the activation states between the visible layer and the hidden layer unit are independent of each other, so when the V and H states are determined, the activation probability of the visible layer and the hidden layer unit can be obtained as

$$P(v_i = x | h) = \frac{1}{\sigma_i \sqrt{2\pi}} \exp \left(-\frac{\left(x - a_i - \sum_{j=1}^n w_{ij} h_j \right)^2}{2\sigma_i^2} \right), \quad (4)$$

$$P(h_j = 1 | v) = \text{sigmoid} \left(a_j + \sum_{i=1}^m \frac{v_i}{\sigma_i} w_{ij} \right). \quad (5)$$

3.2.2. GBRBM Training Process. The purpose of RBM model training is to calculate the optimal value of parameter $\theta = (w, a, b, \sigma)$, so as to obtain the optimal model. Usually, it can be achieved through the maximum likelihood estimation formula:

$$\ln L(\theta | v) = \ln \left(\frac{1}{Z(\theta)} \sum_h e^{-E(v, h | \theta)} \right). \quad (6)$$

In order to calculate the updated equation of each parameter, we use the contrast divergence (CD) algorithm proposed by Hinton to train the model and add the adjustment

of σ_i in GBRBM; the training process is as follows: (a) at the beginning of training, assign the input data to the nodes of the visible layer to obtain v_i and obtain the data features mapped from the visible layer V to the hidden layer H according to equation (5), (b) calculate reversely according to equation (4) and map the output obtained in (a) to the visible layer V , and (c) calculate the error between the samples according to the comparison between the reconstructed results and the original data and adjust the interlayer weight W to reduce the error. The updating process of parameter vector $\theta = \{w, a, b, \sigma\}$ is as follows:

$$\begin{cases} \Delta w_{ij} = \gamma (\langle v_i h_j \rangle_{\text{data}} - \langle v_i h_j \rangle_{\text{recon}}), \\ \Delta a_i = \gamma (\langle v_i \rangle_{\text{data}} - \langle v_i \rangle_{\text{recon}}), \\ \Delta b_j = \gamma (\langle h_j \rangle_{\text{data}} - \langle h_j \rangle_{\text{recon}}), \\ \Delta \sigma_i = \gamma (\langle \sigma_i \rangle_{\text{data}} - \langle \sigma_i \rangle_{\text{recon}}). \end{cases} \quad (7)$$

In equation (7), γ is the RBM learning rate, $\langle \cdot \rangle_{\text{data}}$ is the mathematical expectation of the input data, and $\langle \cdot \rangle_{\text{recon}}$ is the mathematical expectation of the reconstructed data. The output of the trained model forward passing can represent the original input of the visible layer; thus, the feature extraction of the input data is completed.

3.3. Extreme Learning Machine. The BP neural network is used in the upper layer of DBN. Although the BP neural network has better adaptive ability, it adopts the gradient descent algorithm in the training process. When the neuron is close to 0 or 1, the convergence speed is relatively slow, resulting in a longer training time for the model. Moreover, the BP algorithm may fall into a local optimum for complex nonlinear problems such as strip quality prediction. In order to solve these problems, this paper introduces the extreme learning machine model.

Extreme learning machine (ELM) is a single hidden layer feedforward neural network proposed by Huang Guangbin in 2004, including the input layer, hidden layer, and output layer. The structure is shown in Figure 2. The offset of the hidden layer node and the weight of the input layer in ELM are randomly assigned during initialization, which greatly shortens the training time of the model. The output weight of ELM is adjusted by the regularized minimum mean square error, which can ensure the global optimization ability of ELM. Therefore, ELM has relatively high learning efficiency and strong generalization ability and is more suitable for complex production scenarios such as the steel finishing rolling process.

Suppose there are M sample (X_i, y_i) , $X_i = [x_{i1}, x_{i2}, \dots, x_{in}]^T \in R^n$ and $y_i = [y_{i1}, y_{i2}, \dots, y_{im}]^T \in R^m$ are the input samples and their corresponding expected output, respectively. Assuming that the number of hidden layer nodes is L , the ELM model can be expressed as

$$\sum_{i=1}^L \beta_i g(w_i \cdot x_j + b_i) = o_j, \quad j = 1, \dots, M. \quad (8)$$

FIGURE 2: Structure of ELM.

In equation (8), $g(x)$ is the activation function of the hidden layer; w_i and β_i are the weight vectors between the input layer and hidden layer and between the hidden layer and output layer, respectively; b_i is the offset of the hidden layer node; and o_j is the output of ELM. The purpose of network training is to minimize the output error and find a special w_i, β_i, b_i so that the output value is the target value:

$$\sum_{i=1}^L \beta_i g(w_i \cdot x_j + b_i) = y_j, \quad j = 1, 2, \dots, M. \quad (9)$$

Expressed as a matrix:

$$H\beta = Y, \quad (10)$$

$$H = \begin{bmatrix} g(w_1 \cdot x_1 + b_1) & \cdots & g(w_L \cdot x_1 + b_L) \\ \vdots & \ddots & \vdots \\ g(w_1 \cdot x_M + b_1) & \cdots & g(w_L \cdot x_M + b_L) \end{bmatrix}. \quad (11)$$

In equation (10), h is the hidden layer output matrix, β is the weight matrix, and Y is the network output matrix. Since ELM randomly generates w_i and b_i in the initialization stage, the matrix H is uniquely determined. The training process of the network can be transformed into a linear system solving the problem. The approximate solution of β can be obtained according to the Moore-Penrose generalized inverse matrix:

$$\beta^* = H^+ Y. \quad (12)$$

In equation (12), H^+ is the Moore-Penrose generalized inverse of the hidden layer output matrix H .

3.4. GBDBN-ELM Model. In this article, the RBM in the traditional DBN is replaced with GBRBM to form GBDBN, and then, the GBDBN model and the ELM model are combined, as shown in Figure 3. For an N -layer GBDBN-ELM model, the strip quality sample data is assigned to the visible layer V of the first layer of GBRBM, the first hidden layer h_1 and the second hidden layer h_2 form GBRBM, the output of the former GBRBM is also the input of the latter GBRBM, and so on, until the $N-2$ layer of the model; the $N-2$ layer, $N-1$ layer, and the last output layer are the ELM. The

FIGURE 3: GBDBN-ELM network structure for strip quality prediction.

$N - 2$ layer of the model is both the output of the last layer of GBRBM and the input of ELM.

In this model, the strip quality input data is extracted by multilayer GBRBM to form a low-dimensional feature expression, which ensures the features of the original input data set as much as possible. Then, input the extracted features into ELM for regression prediction to obtain the predicted strip quality prediction data.

For an N -layer GBDBN-ELM model, suppose the number of neurons in the $N - 1$ layer network is n , and the number of neurons in the $N - 2$ layer network is m , the network can be expressed as

$$\sum_{i=1}^m \beta_i g(w_i \cdot h_{N-2} + b_i) = o_j, \quad j = 1, \dots, n. \quad (13)$$

According to the ELM algorithm, the output matrix of the $N - 1$ layer of the network and the solving equation of β can be obtained as

$$H_{N-1} = \begin{bmatrix} g(w_1 \cdot h_{N-2,1} + b_1) & \cdots & g(w_L \cdot h_{N-2,1} + b_L) \\ \vdots & \ddots & \vdots \\ g(w_1 \cdot h_{N-2,n} + b_1) & \cdots & g(w_L \cdot h_{N-2,n} + b_L) \end{bmatrix}, \quad (14)$$

$$\beta^* = H_{N-1}^+ Y. \quad (15)$$

The GBDBN-ELM model combines the unsupervised learning characteristics of DBN with high learning efficiency and strong generalization ability of ELM. It can improve the training speed and prediction accuracy.

4. Experimental Study

The indexes to measure the quality of strip steel mainly include the thickness, width, and surface temperature, among which the thickness is the most important index to evaluate whether the steel is up to the standard [25]. Therefore, this paper verifies the feasibility of the improved deep confidence network by predicting the thickness of the finished rolled strip and compares the improved model with other machine learning algorithms and deep learning algorithms to illustrate the superiority of the model.

4.1. Data Preparation

4.1.1. Data Source. The experimental data in this paper comes from a 1580 mm hot strip finishing line of a steel company. The production line consists of 7 units. After 5~7 passes of rough rolling, we can get intermediate billet of 25~60 mm thick, which can be sent to the finishing mill after the hot coil box, flying shear, and dephosphorization box. The control of strip thickness is mainly in the finishing mill. After the finishing mill, we can obtain the finished strip with thickness of 1.2~12.7 mm. The production line consists of seven finishing mills, namely, F1~F7. A work roll bending device is adopted on 7 rolling mills, among which F2~F4 are PC rolling mills with crossed rolls in pairs. Looper rolls are installed between each two rolling mills to balance the rolling tension and prevent plate stacking. The threading speed, acceleration, reduction of each stand, and bending force of each stand of the F1~F7 rolling mill are calculated and set by a computer control system according to the variety and specification of rolled strip and can be adjusted dynamically. The exit of the F7 finishing mill is equipped with rolling line detection instruments for thickness, width, temperature, and crown of strip steel quality, which can monitor the quality in real time and modify the process parameters to improve the quality of rolled products.

In this experiment, the process parameters set by the computer control system of the seven finishing mills in the finishing rolling stage and the strip quality parameters detected by the sensor at the F7 exit are collected within 8 days. The sampling time interval is 90 seconds, and a total of 3350 sets of production data are collected. Each set of data includes 7 sets of finishing mills' reduction position, rolling force, stand speed, oil film compensation, eccentric compensation, and other process parameters, as well as their confidence and number of points, totaling 234 columns of data.

4.1.2. Data Preprocessing. As there are 234 process parameters collected, if all these data are used to predict the strip thickness, the deep learning network will be very complex and the training time will be very long. However, some of the data are not highly correlated with the final strip exit thickness. In this paper, the importance of each element is sorted by the gradient boosting decision tree method, as shown in Figure 4. Finally, 69 factors are selected as the input parameters for strip quality prediction, including entrance thickness, exit temperature, roll gap of each stand, rolling force, stand speed, roll bending force, back tension, and looper angle.

FIGURE 4: The importance of each element to the thickness of the strip (the first 20).

Due to the complex production environment, water vapor, and other interference factors, and the instability of the computer system and sensor itself, the data collected on-site has certain errors, missing data, and abnormal values. For the problem of missing data, this paper uses the mean method to supplement the missing value. For outliers, first, calculate the Euclidean distance between samples by the k -means clustering method and extract outliers, and then, eliminate the outliers. The min-max normalization method is used to carry out linear transformation on the original data, and the data is mapped between $[0, 1]$, so as to eliminate the influence of parameter dimension on the prediction results.

According to the holdout verification method, 3000 groups of data are randomly selected as the training set, and the remaining 350 groups of data are used as the verification set of the model after training.

4.2. Parameter Setting

4.2.1. Key Parameter. Before training and prediction, some relevant parameters need to be set in advance. These parameters cannot be updated in the training process but given in advance through the parameter setting method. These parameters have a great impact on the learning ability of the model and need to be adjusted continuously to maximize the advantages of the model.

By analyzing the structure and principle of the network model, the superparameters of the GBRBM-ELM model need to be set in advance, including the number of GBRBM layers in DBN, the number of hidden layer nodes in DBN and ELM, the number of visible layer nodes in the first RBM layer, the number of ELM output layer nodes, the size of data blocks in the network training phase, the number of training rounds, the learning rate and momentum term.

Since 69 input parameters are selected to predict the strip thickness, the number of visible layer nodes is 69 and the number of output layer nodes is 1. This paper uses different methods to set and tune different parameters.

4.2.2. Grid Search Method. Grid search is to use prior knowledge to specify the value range of parameters. In this range, the parameters are listed hierarchically. Based on the experimental results, the optimal parameter value with a small prediction error can be selected.

Taking GBRBM layers as an example, it is one of the important parameters of the DBN network structure. The number of RBM layers directly affects the prediction effect of the model. When the number of RBM layers is too small, the model will not be able to take advantage of deep learning, and the prediction effect will be poor. But too many layers will lead to the training time process or cause overfitting. According to prior knowledge, the change range of the number of layers is set to be between 1 and 10. The prediction effect of the model is shown in Figure 5.

According to the comparison results, when the number of RBM layers is 4, the model error is the smallest, so this paper uses a 4-layer RBM network structure.

Using the same method, after multiple comparison experiments, the number of hidden layer nodes in ELM, data block size, training rounds, learning rate, and momentum can be obtained. The optimal parameters of the network are shown in Table 1.

4.2.3. Particle Swarm Optimization. Another main parameter of the DBN model structure is the number of nodes in each hidden layer. Because the hidden layers in the 4-layer RBM are related to each other, the number of nodes varies widely, and there are many node combinations; it is difficult to use grid search to enumerate one by one to find the optimal

FIGURE 5: Variation of residual sum of squares with RBM layers.

TABLE 1: Parameters of GBDBN-ELM model.

Number of RBM layers	4
Number of hidden layer nodes in ELM	60
Data block size	150
Training rounds	20
Learning rate	0.01
Momentum	0.5

combination of the number of nodes. In this paper, particle swarm optimization (PSO) is used to automatically calculate the number of hidden layer nodes in each layer.

The particle swarm algorithm compares the optimized solution of each objective function to the particles in the search space. Each particle has two parameters, position and velocity, and the fitness of the particle can be calculated from the objective function. By comparing the fitness of the particle at the current time with that at the previous time, the individual optimal position p_{id} can be obtained. Similarly, the group optimal position g_d can be obtained. According to equation (14), the velocity and position of the particle can be updated, and the global optimal solution satisfying the termination condition can be found:

$$v_{id}(t+1) = wv_{id} + c_1r_1(p_{id} - x_{id}(t)) + c_2r_2(g_d - x_{id}(t)), \quad (16)$$

$$x_{id}(t+1) = x_{id}(t) + v_{id}(t+1). \quad (17)$$

In equations (16) and (17), $v_{id}(t)$ and $x_{id}(t)$ are the velocity and position of particles at time t , c_1 and c_2 are the learning factors, and r_1 and r_2 are random numbers in (0,1).

Set the population size of PSO as 10 and the number of alternations as 10, and finally, find the number of hidden layer nodes of 4-layer DBN as [63, 54, 46, 35].

4.3. Model Training. The training of the GBDBN-ELM combined model is divided into two parts:

- GBDBN Module Training.** First, initialize the network parameters, weights, and the number of hidden layer nodes of the model. The first $N - 2$ layers of the combined model are the GBDBN model, and the pre-processed input data is allocated to the visible layer nodes to establish v_i . Next, the contrast divergence (CD) algorithm is used to train each RBM layer by layer from bottom to top. When one RBM layer is trained, the parameters of the layer are fixed and used as the input of the upper RBM to train the upper RBM, and so on, until all RBM training is completed. Finally, the bottom features are gradually gathered into the high-level features and finally sent to the regression unit.
- ELM Module Training.** The connection weights of the $n - 1$ layer and $n - 2$ layer are initialized. The $N - 2$ layer is the feature extraction layer of the last layer of GBDBN. The preprocessed high-dimensional labeled data is used as the input of the GBDBN module after training, and the feature extraction result is used as the input of the initial elm module. The elm algorithm is used for training to obtain better model parameters.

Based on GBDBN-ELM module training, effective DBN and ELM are obtained, respectively. The test data set is pre-processed to obtain high-dimensional sample data to be detected. The trained GBDBN model is used for feature extraction to obtain better feature data. The predicted strip thickness can be obtained by the ELM module. The overall process is shown in Figure 6.

4.4. Result Analysis and Comparison

4.4.1. Model Evaluation Index. In this paper, five indexes are used to evaluate the prediction effect of the model, including

FIGURE 6: Strip steel quality prediction process based on GBDBN-ELM.

FIGURE 7: The curve of the predicted value and the true value of the strip thickness.

the sum of squares of residuals (SSR), root mean square error (RMSE), R square coefficient of determination (R^2), and training time (T). The index is calculated as follows:

$$\begin{cases} \text{SSE} = \sum_{i=1}^N (y_i^* - y_i)^2, \\ \text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^N (y_i^* - y_i)^2}, \\ R^2 = 1 - \frac{\sum_{i=1}^N (y_i^* - \bar{y}_i)^2}{\sum_{i=1}^N (y_i - \bar{y}_i)^2}. \end{cases} \quad (18)$$

The smaller SSE, RMSE, and MAE, the better the prediction effect. R^2 is the fitting degree of the model, and the closer R^2 is to 1, the better the regression effect of the model is; T is the time from the beginning to the end of training, and the smaller t is, the faster the model training is and the better the model performance is.

4.4.2. Prediction Results. 350 sets of data were used in the test set to evaluate the performance of the model. In this paper, the simulation results of the prediction model are assessed by analyzing the curve of the predicted value and the true value of the strip thickness, the curve of the prediction error, and the curve of the prediction relative error.

FIGURE 8: The curve of the prediction error.

FIGURE 9: The curve of the prediction relative error.

TABLE 2: Comparison of prediction results of different models.

Index\model	BP	ELM	DBN	DBN-ELM	GBDBN-ELM
SSR	59.3560	78.2718	14.3091	22.398	8.5856
RMSE	0.4118	0.4728	0.2022	0.2530	0.1565
R^2	0.9872	0.9607	0.9926	0.98854	0.9956
T	3.1548	0.5127	96.9802	32.0120	33.8380

FIGURE 10: Comparison of relative errors of BP and GBDBN-ELM (part).

FIGURE 11: Comparison of relative errors of ELM and GBDBN-ELM (part).

It can be seen from Figure 7 that the predicted value of strip thickness obtained by the GBDBN-ELM model is very close to the real value of strip thickness, and the fluctuation trend and variation range of them are basically consistent. By analyzing Figure 8, we can get the following results: 93.7% of the absolute error between the predicted value and the real value is between $[-0.3, 0.3]$, and only a few points have relatively large error between $[-0.4, 0.8]$. The reason is

that the test samples are randomly selected, and these points have large mutation compared with the surrounding points, and the change range of the model prediction is less than the real change range. As shown in Figure 9, the relative error of strip thickness predicted by the model is within 10%, among which 80.9% is less than 5%. In conclusion, the prediction model of strip thickness based on GBDBN-ELM has high accuracy.

FIGURE 12: Comparison of relative errors of DBN and GBDBN-ELM (part).

FIGURE 13: Comparison of relative errors of DBN-ELM and GBDBN-ELM (part).

4.4.3. Comparison of Different Models. In order to comprehensively analyze the prediction performance of the model for strip steel quality, this paper compares it with the BP neural network, ELM, traditional DBN network, and DBN-ELM model and evaluates the above models according to SSR, RMSE, and R^2 . The results are shown in Table 2. In order to intuitively compare the prediction error of the model, Figures 10–13 show the relative error of the prediction results of the first 100 groups of test data of GBDBN-ELM and show the comparison with other models, respectively.

Comparing the prediction results of the BP neural network and ELM in Table 2, we can find that the training time of ELM is shorter, since it generates hidden layer offset and input layer weights randomly during initialization, and there is no need to update them during training. However, the BP neural network can adjust the model to a better state during training, so its prediction accuracy is higher than that of ELM. However, by analyzing Figures 10 and 11, it can be found that a large part of the error of the two neural network prediction results is greater than 10%,

so it is not enough to use simple machine learning for strip quality prediction.

It can be seen from Table 2 that although DBN-ELM significantly shortens the training time, the prediction accuracy is slightly lower than DBN. GBDBN-ELM improves the RBM network on the basis of DBN-ELM to make it suitable for the continuous value regression problem and can retain more data features when predicting strip thickness. Therefore, GBDBN-ELM combines the advantages of DBN and ELM.

By comparing the results in Table 2, it is found that compared with DBN, GBDBN-ELM reduces the training time by 65.1%, the sum of squares error by 39.9%, and the root mean square error by 22.6%; compared with DBN-ELM, GBDBN-ELM reduces the sum of square error by 61.7% and the root mean square error by 38.1%, and the training time only increases by one second, which indicates that the improved model can effectively improve the accuracy and shorten the training time.

From Figures 12 and 13, we can see the prediction effect of the three deep learning models for each test data more intuitively. It can be seen that the overall relative error of the prediction results of GBDBN-ELM is smaller than that of DBN and DBN-ELM. The relative error of GBDBN-ELM hardly exceeds 8%, and it is certain that the prediction accuracy of the improved model meets industry requirements. In 350 sets of test data, the error of GBDBN-ELM's prediction results is 63.7% less than DBN, and 65.7% less than DBN-ELM. The average relative error is reduced, respectively, by 40.4% and 32.6%.

Especially for data with large prediction errors of DBN and DBN-ELM, GBDBN-ELM can significantly reduce the error and achieve better prediction results. The advantages of the improved model are also reflected here to a large extent.

The analysis of Table 2 and Figures 10–13 shows that the improved GBDBN-ELM model can improve the prediction accuracy and shorten the training time to a large extent.

5. Conclusion

This paper proposes an improved DBN strip quality prediction method to solve the problem that the strip quality prediction accuracy is not high because there are many sensors involved in the strip production process, and most of the process parameters are coupled with each other and have serious nonlinearity. In this paper, the RBM in DBN is changed into GBRBM to eliminate the dependence on binary distribution, extract the features of high-dimensional and high coupling input data, combine GBDBN with ELM, replace the BP network in DBN with ELM, and input the extracted data features into ELM for strip quality prediction. The GBDBN-ELM model is verified by the data of the steel finishing line and used to predict the strip thickness. We can draw the following conclusions.

The simple BP neural network and ELM model cannot deal with the high dimension and high coupling nonlinear data produced by the complex production process. Due to the simple network structure, they cannot fully extract the data features and mine the knowledge contained in the data,

resulting in the accuracy of strip thickness prediction being not enough.

The GBDBN model proposed in this paper can solve the problem of low prediction accuracy caused by complex input data. The GBDBN network can retain as many abstract features of input data as possible, so that ELM can obtain higher prediction accuracy.

Through the comparison with the DBN network, it can also be known that using the ELM algorithm for GBDBN network training and prediction calculations can greatly shorten the time and solve the problem of excessive training time caused by the complexity of the DBN network.

Data Availability

The raw/processed data required to reproduce the experiments in this article cannot be shared due to corporate confidentiality.

Conflicts of Interest

The authors declare that there is no conflict of interest.

Acknowledgments

This work was supported by the National Science and Technology Innovation 2030 of China Next-Generation Artificial Intelligence Major Project, Data-Driven Tripartite Collaborative Decision-Making and Optimization, under Grant 2018AAA0101801.

References

- [1] C. Klinkenberg, B. Kintscher, K. Hoen, and M. Reifferscheid, "More than 25 years of experience in thin slab casting and rolling current state of the art and future developments," *Steel Research International*, vol. 88, no. 10, article 1700272, 2017.
- [2] Q. K. Pan, Q. D. Chen, T. Meng, B. Wang, and L. Gao, "A mathematical model and two-stage heuristic for hot rolling scheduling in compact strip production," *Applied Mathematical Modelling*, vol. 48, pp. 516–533, 2017.
- [3] D. H. Zhang, W. Peng, J. Sun, J. G. Ding, and X. Li, "Key intelligent technologies of steel strip rolling process," *Journal of Iron and Steel Research*, vol. 31, no. 2, pp. 174–179, 2019.
- [4] J. Y. Chen, P. H. Hao, J. P. Song, and H. J. Guo, "Study on temperature prediction model of cold rolling strip," in *International Conference on Artificial Intelligence and Big Data (ICAIBD)*, pp. 214–217, Chengdu, China, May 2018.
- [5] C. Yao, A. He, J. Shao, and J. Zhao, "A real-time quasi-3D metal flow model for hot strip rolling," *International Journal of Mechanical Sciences*, vol. 159, pp. 91–102, 2019.
- [6] S. Ganguly, X. Wang, K. Chandrashekhara, M. F. Buchely, and S. Lekakh, "Modeling and simulation of void closure during steckel mill rolling for steel plate," *Steel Research International*, vol. 92, no. 2, article 2000293, 2020.
- [7] C. Yao, A. He, J. Shao et al., "Finite difference modeling of the interstand evolutions of profile and residual stress during hot strip rolling," *Metals*, vol. 10, no. 11, article 1417, 2020.
- [8] A. Dogan and D. Birant, "Machine learning and data mining in manufacturing," *Expert Systems With Applications*, vol. 166, article 114060, 2021.

- [9] B. Waschneck, A. Reichstaller, L. Belzner, T. Altenmuller, and T. Bauernhansl, "Deep reinforcement learning for semiconductor production scheduling," in *Annual SEMI Advanced Semiconductor Manufacturing Conference (ASMC)*, pp. 301–306, Saratoga Springs, NY, 2018.
- [10] M. Azamfa, X. Li, and J. Lee, "Deep learning-based domain adaptation method for fault diagnosis in semiconductor manufacturing," *IEEE Transactions on Semiconductor Manufacturing*, vol. 33, no. 3, pp. 445–453, 2020.
- [11] C. B. Lemos, P. C. M. A. Farias, E. E. Simas, and A. G. S. Conceicao, "Convolutional Neural Network Based Object Detection for Additive Manufacturing," in *International Conference on Advanced Robotics (ICAR)*, pp. 420–425, Belo Horizonte, Brazil, Dec 2019.
- [12] N. Kotkunde, A. Balu, A. K. Gupta, and S. K. Singh, "Development of predictive models for formability study of Ti-6Al-4V alloy at elevated temperatures," *Materials Today-Proceedings*, vol. 4, no. 2, pp. 937–945, 2017.
- [13] X. Li and Y. R. Dai, "Prediction model of hot rolled strip quality based on K-means clustering and neural network," in *11th International Symposium on Computational Intelligence and Design (ISCID)*, vol. 2, pp. 150–153, Hangzhou, China, 2018.
- [14] Y. Wu, H. C. Ni, X. Li, F. Luan, and Y. D. He, "Prediction of bending force in the hot strip rolling process using multilayer extreme learning machine," *Mathematical Problems in Engineering*, vol. 2021, Article ID 6682660, 13 pages, 2021.
- [15] Z. Xu, D. D. Ye, J. J. Chen, and H. T. Zhou, "Novel terahertz nondestructive method for measuring the thickness of thin oxide scale using different hybrid machine learning models," *Coatings*, vol. 10, no. 9, p. 805, 2020.
- [16] A. Singh and A. Tiwari, "A study of feature selection and dimensionality reduction methods for classification-based phishing detection system," *International Journal of Information Retrieval Research*, vol. 11, no. 1, pp. 1–35, 2021.
- [17] Y. M. Liu, H. F. Zhou, F. Tsung, and S. Zhang, "Real-time quality monitoring and diagnosis for manufacturing process profiles based on deep belief networks," *Computers & Industrial Engineering*, vol. 136, pp. 494–503, 2019.
- [18] D. Y. Yang and D. M. Frangopol, "Evidence-based framework for real-time life-cycle management of fatigue-critical details of structures," *Structure and Infrastructure Engineering*, vol. 14, no. 5, pp. 509–522, 2018.
- [19] H. Q. Zheng, Y. R. Yang, X. Y. Sun, and C. Wen, "Nondestructive detection of anchorage quality of rock bolt based on DS-DBN-SVM," in *International Conference on Machine Learning and Cybernetics (ICMLC)*, pp. 288–293, Chengdu, China, 2018.
- [20] Z. M. Li, C. X. Liang, and M. S. Wang, "Short-term power generation output prediction based on a PSO-DBN neural network," *Power System Protection and Control*, vol. 48, no. 8, pp. 149–154, 2020.
- [21] S. Goudarzi, M. N. Kama, M. H. Anisi, S. A. Soleymani, and F. Doctor, "Self-organizing traffic flow prediction with an optimized deep belief network for internet of vehicles," *Sensors*, vol. 18, no. 10, article 3459, 2018.
- [22] X. X. Zhang, Y. Miao, X. Z. Liu, and K. Wu, "Hot strip thickness prediction based on least squares support vector machine with feature extraction by deep belief network," *Metallurgical Industry Automation*, vol. 44, no. 2, pp. 17–22, 2020.
- [23] Y. Xie, J. Zou, Z. Li, F. Gao, and C. Peng, "A novel deep belief network and extreme learning machine based performance degradation prediction method for proton exchange membrane fuel cell," *IEEE Access*, vol. 8, pp. 176661–176675, 2020.
- [24] J. Melchior, N. Wang, and L. Wiskott, "Gaussian-binary restricted Boltzmann machines for modeling natural image statistics," *Plos One*, vol. 12, no. 2, article e0171015, 2017.
- [25] R. R. Dema, R. N. Amirov, and O. R. Latypov, "Mathematical model for assessing the management of quality parameters of hot-rolled strips according to the criterion of local thickness variation," *Materials Today: Proceedings*, vol. 19, no. 5, pp. 2417–2421, 2019.