



Prediction model of hot strip crown based on industrial data and hybrid the PCA-SDWPSO-ELM approach

Zhenhua Wang^{1,2,4} · Yuanming Liu^{1,2,4} · Tao Wang^{1,2,4} · Dianyao Gong³ · Dianhua Zhang³

Accepted: 31 January 2023 / Published online: 11 February 2023

© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2023

Abstract

The accurate prediction of strip crown is the precondition of the shape preset model in hot strip rolling. In this study, a new hybrid strip crown forecasting model is proposed in combination with extreme learning machine (ELM) and the industrial data. The production data of 1780 mm hot strip rolling are collected by the on-site data acquisition system to form dataset. Principal component analysis (PCA) is used to reduce the dimension of the input data for modeling samples. To improve the prediction accuracy of ELM, an improved PSO based on S-curve decreasing inertia weight (SDWPSO) is proposed to optimize the initial weights and biases of ELM. Finally, the optimal ELM model and simple production dataset are used to establish a strip crown prediction model of hot strip rolling named PCA-SDWPSO-ELM. The comprehensive performance of the proposed hybrid PCA-SDWPSO-ELM prediction model is evaluated by MAE, MAPE and RMSE. The superiority of the proposed model is also proved by comparing the prediction results with those of the other three comparison models. The research shows that the hybrid PCA-SDWPSO-ELM method can solve the problem of nonlinear and strong coupling in traditional engineering. It is suitable for parameter prediction and optimization in the iron and steel manufacturing industry, especially in the process of shape control in hot strip rolling.

Keywords Strip crown prediction · Extreme learning machine (ELM) · Industrial data · Principal component analysis (PCA) · Regression model · Hot strip rolling

1 Introduction

Hot strip rolling products are widely used in the national economy (Pittner and Simaan 2011; Peng et al. 2015). Strip shape, including the crown and flatness of strip, is one of the key indices of product quality. The crown represents the difference in thickness distribution between the two

sides and the middle on the cross section of the strip, and flatness represents the different elongations of the strip in the length direction (Peng et al. 2014). Poor strip shape quality not only affects the hot rolling process, but also adversely affects the subsequent processes, such as cold rolling and shearing. Therefore, shape control is a core, frontier and highly difficult technology in the hot strip rolling process. Naturally, research on shape prediction and control has great theoretical significance and practical value.

To date, many scholars have carried out intensive studies of shape control in rolling processes. The common research methods of strip shape control are traditional mathematical analysis and finite element analysis. For strip crown control, it is generally effective to combine all kinds of factors affecting the strip crown, such as the detection device, the mathematical model, the load distribution, the bending and shifting system (Pin et al. 2013). It is the most basic method of establishing the strip shape control model through traditional mathematical analysis. The core idea of this method is to fully consider the characteristics of the

✉ Zhenhua Wang
wangzhenhuaneu@yeah.net

¹ College of Mechanical and Vehicle Engineering, Taiyuan University of Technology, Taiyuan 030024, Shanxi, People's Republic of China

² Engineering Research Center of Advanced Metal Composites Forming Technology and Equipment, Ministry of Education, Taiyuan 030024, Shanxi, People's Republic of China

³ State Key Laboratory of Rolling and Automation, Northeastern University, Shenyang 110819, Liaoning, People's Republic of China

⁴ TYUT-UOW Joint Research Centre, Taiyuan 030024, Shanxi, People's Republic of China

rolling mill model and metal flow law and to calculate the elastic deformation and flexural deformation of the rolling mill by using the influence function method (Li et al. 2010). Combined with the bending roll and shifting roll strategies, the shape of the roll gap profile is calculated to realize crown control (Peng et al. 2014). The finite element method can flexibly simulate the metal flow law and stress and strain of strips under various rolling conditions (Moazeni and Salimi 2015). The reasonable grid dividing and setting boundary conditions provide the most accurate representation for the roll system force and deformation of the whole rolling mill; thus, the calculation results of the shape parameters can be obtained with high precision, and the influence of the shape actuator on the flatness, crown and edge drop in the actual rolling process can be verified (Linghu et al. 2014; Tran et al. 2015). However, it is increasingly difficult to improve the precision of traditional shape mathematical models because of the many simplified conditions in the process of solving mathematical analytic method and finite element method.

To continuously improve the strip shape control accuracy, artificial intelligence technology with data and algorithms as the core has attracted increasing attention from scholars. Intelligent technology is not only more and more used in mobile electronic information network (Nikjoo et al. 2018; Mohajer et al. 2022a, b), but also has emerged in the field of industrial control. Early intelligent research was on the optimization of mechanism model parameters using heuristic intelligent optimization algorithms, such as using a genetic algorithm (GA) to establish multiobjective optimal control strategy and apply it to the identification of strip shape parameters and the setting of rolling schedules. The optimal strip crown and flatness setting values can be obtained successfully, and the precision of strip shape control can be improved (Nandan et al. 2005). The GA and the ant colony algorithm (ACA) was used to optimize the crown model of hot strip rolling, which proved that an evolutionary algorithm was practical in the optimization of rolling process parameters (Chakraborti et al. 2006). In recent years, data-driven modeling theory based on machine learning has become the main direction of intelligent modeling. Common machine learning algorithms mainly include artificial neural networks (ANNs), support vector machines (SVMs), decision trees and random forests (RFs), deep learning. The earliest application of artificial intelligence in the rolling field was to build a rolling force prediction model of the leveling rolling process using a neural network and applied it to a practical production line, and good results were obtained (Larkiola et al. 1998; Pican et al. 1996; Moussaoui and Abbassi 2006). The pattern recognition method is generally used in shape control systems, and the accurate shape standard pattern characteristic coefficient is the premise of shape control. The T-S

cloud inference neural network (Zhang et al. 2015a, b), PID neural network (Zhang et al. 2015a, b) and radial basis function network (Zhang et al. 2016) can effectively identify common defects in cold rolling and improve the precision of shape control. The relationship model between input parameters and strip shape can also be established by combining ANN and GA to predict the minimization flatness of hot strip rolling (John et al. 2008). In addition, the transfer matrix between the characteristic flatness error and flatness adjustment parameters can be established by using a GA to optimize the BP neural network. The transfer matrix was successfully applied to the strip shape adjustment mechanism of a rolling mill for the accurate control of strip shape (Liu et al. 2005; Peng et al. 2008). Combining a shape control matrix with the differential evolution algorithm (DE) optimization ELM, an intelligent model of cold strip rolling can be established and applied to the shape control process (Yang et al. 2017).

According to rolling theory, hot strip crown control is a multivariable, strong coupling and nonlinear process. At present, the strip crown prediction model based on traditional mathematical analytic theory has poor accuracy and can no longer meet the needs of higher precision rolling. Therefore, it is urgent to establish a new type of strip crown model with high precision prediction ability by using the above artificial intelligence methods. Artificial intelligence method simulates the real process of human brain processing. Based on a large amount of industrial data, it can predict the target crown, which can effectively prevent the error caused by the assumption being divorced from reality and the simplification being too rough. Up to now, there have been a few studies of strip crown prediction based on AI techniques. ANNs are basic methods of intelligent modeling. A hybrid model was constructed by Deng et al. (2019) based on hot strip rolling production data and a deep learning network to predict hot strip crown, and the results show that the 97.04% absolute error of modeling samples is less than 5 μm . Wang et al. (2019) established a prediction model of hot strip crown by using the mind evolutionary algorithm (MEA) to optimize the multilayer perceptron. Based on research on accurate prediction of strip crown, the influence of various input factors on the strip crown was analyzed. SVM is another common machine learning algorithm that can realize model training with small sample sizes. Many studies have combined heuristic intelligent optimization algorithms and SVMs to establish strip crown prediction models. The basic strategy is to use an intelligent optimization algorithm to optimize the key parameters in the SVM model and use the selected optimal parameters to establish the prediction model and predict the strip crown (Wang et al. 2018; Ji et al. 2021; Song et al. 2022). Although the proposed SVM model has high prediction accuracy, it is not suitable for large-scale industrial data

modeling. In addition, the model based on a single machine learning method easily underfits or overfits, and the ensemble model can solve this problem. Tree-based ensemble algorithms have recently been recognized as one of the best and most commonly used supervised learning methods. The ensemble learning method with the decision tree as the base learner inherits the advantages of the decision tree, such as being simple, highly interpretable and robust to anomalies, overcoming the disadvantages of instability and high variance (Zhou 2012). RF is a typical ensemble learning algorithm, that can also be used to establish a strip crown prediction model of hot strip rolling with good generalization performance (Sun et al. 2021). The novel strip crown prediction model using well-performing and efficient tree-based ensemble learning algorithms, including extreme gradient boosting (XGBoost) and light gradient boosting machine (LightGBM) algorithms, has been established and deeply studied (Li et al. 2021). Based on the above analysis, the combination of artificial intelligence methods and big data technology is a new trend in studying how to further improve the precision of strip crown control in rolling processes. Although some strip crown prediction models based on artificial intelligence have been reported, the sophisticated modeling process and expensive calculation cost are still incompatible with the characteristics of industrial control, such as fast and simple. Therefore, it is still very urgent and important to establish a strip crown prediction model with high-precision, efficient and easy realization by using more advanced intelligent algorithms and combining with the production data of manufacturers.

In general, the production data of hot strip rolling have the characteristics of redundancy, high dimension and coupling, and there are many factors, including chemical composition parameters, strip parameters, variable parameters of each stand, affecting the strip crown. Because the dimension is too high, the distribution of data on each feature dimension becomes sparser, which is basically disastrous for machine learning algorithms. Therefore, directly modeling with these high-dimensional data will lead to slow training speed and low accuracy. PCA is a multielement statistical method for converting multiple indices into several comprehensive indices with a low loss of information by using dimension reduction and is based on research on the most significant variations of the variables (Samarasinghe 2007). Huang and Harris (1993) applied PCA to select the initial codebook to decrease the dimension of training vectors. Malvoni et al. (2016) used PCA to reduce the dimension of a training set data used for photovoltaic forecast modeling, which improved the accuracy of the forecasting model. Franceschi et al. (2018) used PCA to screen the input variables that affect the prediction effect of PM2.5 and PM10 and established an

environmental early warning model by multilayer perceptron. PCA is also used in image compression research to group all training vectors of the test image and obtain the principal components from all training vectors in the training image (Tsai et al. 2015). The reasonable use of PCA algorithm can greatly simplify the model structure and save the training time.

In this study, a new hybrid PCA-SDWPSO-ELM forecasting model is proposed in combination with artificial intelligence method and the industrial data to predict strip crown in hot rolling. First of all, the production data of 1780 mm hot strip rolling are collected, summarized, mined and sorted by the on-site data acquisition system to form an effective modeling dataset. Then PCA algorithm is used to reduce the dimension of high-dimensional dataset to further improve the simplicity of modeling data. In order to improve the prediction accuracy of ELM, an improved particle PSO based on S-curve is proposed to optimize the initial weights and biases of ELM, so as to determine the optimal ELM model parameters. Finally, the optimal ELM model and simple production dataset are used to establish a strip crown prediction model of hot strip rolling, and the comprehensive performance of the proposed model and other models are analyzed by means of comparative research, which verifies the superiority and effectiveness of the proposed model. This paper is organized as follows: Sect. 2 introduces the basic theory of shape control. Section 3 shows the collection and processing of modeling data and the related modeling process. The discussion of the strip crown forecasting results is described explicitly in Sects. 4 and 5 concludes this paper.

2 Theory of strip crown control

2.1 Strip crown and proportional crown

Strip crown is the thickness difference between the center of the strip cross section and the reference point of the edge. To eliminate the effect of strip edge thinning, the edge reference point is usually located 40 mm from the strip edge. The definition of a strip crown is shown in Fig. 1. The proportional crown is the ratio of the strip crown to the thickness of the strip center.

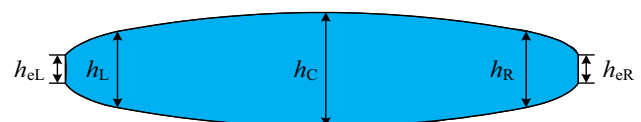


Fig. 1 Thickness variation in the strip cross section

$$C = h_C - \frac{(h_L + h_R)}{2} \tag{1}$$

$$C_{p_h} = \frac{C}{h_c} \tag{2}$$

where C is the strip crown, mm; h_C is the thickness of the center of the strip cross section, mm; h_L and h_R are the thickness of the reference point on the left and right side of the strip cross section, respectively, mm; and C_{p_h} is the proportional crown of the strip.

2.2 Unload roll gap crown model

The traditional crown control model consists of two parts: the unloaded roll gap crown model and the uniform load roll gap crown model. The unloaded roll gap crown is the roll gap crown of the rolling mill without a workpiece and without adding force, which reflects the effect of the roll crown on the shape of the strip, and it is an important factor that affects the shape of the load roll gap. As shown in Fig. 2, the unloaded roll gap crown consists of two parts: the roll gap between work rolls and roll gap between the backup roll and work roll.

2.2.1 Roll crown model

The calculation of the roll crown is the premise of the unload roll gap crown calculation. The roll crown is the diameter difference between the middle and the end of the roll, which is the sum of the original grinding crown, equivalent crown, thermal crown and wear crown. The roll thermal crown and roll wear crown are the crown formed by thermal expansion and wear during the rolling process. The equivalent crown of the roll is 0 for a conventional mill, and for a CVC mill, it can be calculated by interpolation of the transverse position.

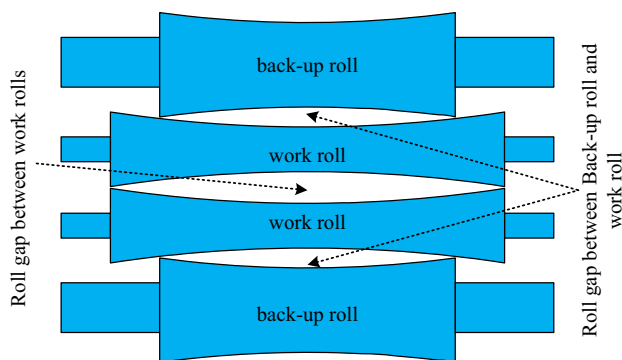


Fig. 2 Schematic plan of the unloaded roll gap

$$C_R = C_{grm} + C_{eqv} + C_t + C_w \tag{3}$$

where C_R is the roll crown, mm; C_{grm} is the original grinding crown of the roll, mm; C_{eqv} is the equivalent crown of the roll, mm; C_t is the thermal crown of the roll, mm; and C_w is wear crown of the roll, mm.

2.2.2 Roll gap crown between the backup roll and work roll

The gap crown between the backup roll and the work roll is determined by the work roll crown and the backup roll crown. Because it corresponds to the contact area between rollers, it is necessary to transform the crown of the work roll by assuming that the roll crown curve exhibits a conic distribution.

$$C_{br-wr} = C_{br} + C_{wr} \cdot \left(\frac{L_{br}}{L_{wr}}\right)^2 \tag{4}$$

where C_{br-wr} is the gap crown between the backup roll and the work roll, mm; C_{br} is the backup roll crown, mm; C_{wr} is the work roll crown, mm; and L_{br} is the backup roll length, mm. L_{wr} is the work roll length.

2.2.3 Roll gap crown between two work rolls

The work roll gap crown is the roll gap crown between two work rolls when no load is carried, and its size is equal to that of the work roll crown.

$$C_{wr-wr} = C_{wr} \tag{5}$$

where C_{wr-wr} is the roll gap crown between two work rolls, mm, and C_{wr} is the work roll crown, mm.

2.3 Uniform load roll gap crown model

The uniform load roll gap crown is the shape of the roll gap when the unit width rolling force is distributed in the contact area between the strip and the work rolls. The shape of the uniform load roll gap depends on the unloaded roll gap crown, roll system deflection and elastic flattening deformation of the rolls caused by the rolling force and bending force. The mathematical model of the uniform load roll gap crown can be described as a function of unit width rolling force, bending force, strip width, roll elastic modulus, roll diameter, gap crown between the backup roll and work roll, and roll gap crown between the two work rolls. The mathematical model is constructed as follows (Peng et al. 2014):

$$\begin{aligned}
 C_{\text{ufd}} = & b_0 \cdot p + b_1 \cdot p^{1.5} + b_2 \cdot C_{\text{wr-wr}} + b_3 \cdot C_{\text{br-wr}} \cdot p \\
 & + b_4 \cdot C_{\text{br-wr}} \cdot p^{1.5} + b_5 \cdot F_w + b_6 \cdot p \cdot F_w \\
 & + b_7 \cdot p^2 \cdot F_w + b_8 \cdot C_{\text{br-wr}} + b_9 \cdot p \cdot D_{\text{wr}} \\
 & + b_{10} \cdot F_w \cdot D_{\text{wr}} + b_{11} \cdot p \cdot D_{\text{br}} + b_{12} \cdot p^{1.5} \cdot D_{\text{wr}} \\
 & + b_{13} \cdot F_w \cdot D_{\text{wr}} + b_{14} \cdot C_{\text{br-wr}} \cdot D_{\text{wr}} \\
 & + b_{15} \cdot C_{\text{br-wr}} \cdot E_{\text{wr}} + b_{16} \cdot D_{\text{wr}} \cdot D_{\text{br}} + b_{17} \cdot D_{\text{br}} \cdot E_{\text{wr}}
 \end{aligned}
 \tag{6}$$

where C_{ufd} is the uniform load roll gap crown, mm; p is the unit width rolling force, kN/mm; F_w is the bending force, kN; D_{wr} and D_{br} are the diameters of the work roll and backup roller, respectively, mm; E_{wr} is the elastic modulus of the work roll, MPa; and b_i is the model coefficients. The model coefficients b_i are cubic polynomials of the strip width:

$$b_i = c_{i,0} + c_{i,1} \cdot W + c_{i,2} \cdot W^2 + c_{i,3} \cdot W^3, \quad i = 0 \sim 17
 \tag{7}$$

where W is the strip width, mm, and c_{ij} are polynomial coefficients.

3 Methodology

3.1 ELM regression modeling

The ELM is a kind of single-hidden-layer feedforward neural network (SLFN), which was proposed in 2004 (Huang et al. 2004, 2006, 2012). The purpose is to simplify the learning parameter setting while overcoming the defect of the BP algorithm, which easily falls into local minima and improving the learning efficiency (Lan et al. 2013).

w is defined as the connection weight of the input layer and the hidden layer, and β is defined as the connection weight of the hidden layer and the output layer.

$$w = \begin{bmatrix} w_{11} & w_{12} & \cdots & w_{1n} \\ w_{21} & w_{22} & \cdots & w_{2n} \\ \vdots & \vdots & & \vdots \\ w_{l1} & w_{l2} & \cdots & w_{ln} \end{bmatrix}_{l \times n}
 \tag{8}$$

$$\beta = \begin{bmatrix} \beta_{11} & \beta_{12} & \cdots & \beta_{1m} \\ \beta_{21} & \beta_{22} & \cdots & \beta_{2m} \\ \vdots & \vdots & & \vdots \\ \beta_{l1} & \beta_{l2} & \cdots & \beta_{lm} \end{bmatrix}_{l \times m}
 \tag{9}$$

b is defined as the biases of the hidden layer neurons; n , l and m are numbers of neurons in the input layer, hidden layer and output layer, respectively.

$$b = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_l \end{bmatrix}_{l \times 1}
 \tag{10}$$

The input and output matrices of the training set are X and Y respectively, and Q is the sample size.

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1Q} \\ x_{21} & x_{22} & \cdots & x_{2Q} \\ \vdots & \vdots & & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nQ} \end{bmatrix}_{n \times Q}
 \tag{11}$$

$$Y = \begin{bmatrix} y_{11} & y_{12} & \cdots & y_{1Q} \\ y_{21} & y_{22} & \cdots & y_{2Q} \\ \vdots & \vdots & & \vdots \\ y_{m1} & y_{m2} & \cdots & y_{mQ} \end{bmatrix}_{m \times Q}
 \tag{12}$$

$g(x)$ is the activation function of the hidden layer neuron. T is the output of the network.

$$\begin{aligned}
 T &= [t_1, t_2, t_3, \dots, t_Q]_{m \times Q}, \\
 t_j &= \begin{bmatrix} t_{1j} \\ t_{2j} \\ \vdots \\ t_{mj} \end{bmatrix}_{m \times 1} \\
 &= \begin{bmatrix} \sum_{i=1}^l \beta_{i1} g(w_i x_j + b_i) \\ \sum_{i=1}^l \beta_{i2} g(w_i x_j + b_i) \\ \vdots \\ \sum_{i=1}^l \beta_{im} g(w_i x_j + b_i) \end{bmatrix}_{m \times 1} \quad (j = 1, 2, \dots, Q)
 \end{aligned}
 \tag{13}$$

where $w_i = [w_{i1}, w_{i2}, \dots, w_{in}]$; $x_j = [x_{1j}, x_{2j}, \dots, x_{nj}]^T$.

H is the hidden layer output matrix, and the specific forms are as follows:

$$\begin{aligned}
 H(w_1, w_2, \dots, w_l, b_1, b_2, \dots, b_l, x_1, x_2, \dots, x_Q) \\
 = \begin{bmatrix} g(w_1 \cdot x_1 + b_1) & g(w_2 \cdot x_1 + b_2) & \cdots & g(w_l \cdot x_1 + b_l) \\ g(w_1 \cdot x_2 + b_1) & g(w_2 \cdot x_2 + b_2) & \cdots & g(w_l \cdot x_2 + b_l) \\ \vdots & \vdots & & \vdots \\ g(w_1 \cdot x_Q + b_1) & g(w_2 \cdot x_Q + b_2) & \cdots & g(w_l \cdot x_Q + b_l) \end{bmatrix}_{Q \times l}
 \end{aligned}
 \tag{14}$$

and represented by a matrix:

$$H\beta = T'
 \tag{15}$$

If the number of neurons in the hidden layer is equal to the number of training set samples, the training samples can be approximated by zero error for any w and b for the SLFN, which means that

$$\sum_{j=1}^Q \|t_j - y_j\| = 0 \tag{16}$$

$$y_j = [y_{1j}, y_{2j}, \dots, y_{mj}]^T (j = 1, 2, \dots, Q) \tag{17}$$

However, when the number of training samples Q is too large, to reduce the computational complexity, the number of hidden layer neurons K usually takes a number smaller than Q , and the training error of the SLFN can approach an arbitrary ε , $\varepsilon > 0$; that is,

$$\sum_{j=1}^Q \|t_j - y_j\| < \varepsilon \tag{18}$$

When the activation function $g(x)$ is infinitely differentiable, the parameters of the SLFN do not need to be completely adjusted and w and b can be randomly selected before training and remain unchanged during training. The connection weight β between the hidden layer and the output layer can be obtained by solving the least square solution of the following equations:

$$\min_{\beta} \|H\beta - T'\| \tag{19}$$

Its solution is

$$\hat{\beta} = H^+ T' \tag{20}$$

where H^+ is the Moore–Penrose generalized inverse of the implicit layer output matrix H .

3.2 S-curve decreasing inertia weight PSO algorithm

3.2.1 Standard PSO algorithm

The standard PSO algorithm originated from the study of the foraging behavior of birds and was originally proposed by Eberhart and Kennedy (1995). All the particles in a group adjust their velocity and position in accordance with the current global optimal solution found by the current individual extremum and the entire particle group that they have found. The velocity and position adjustment strategy are as follows:

$$V_{id}^{k+1} = wV_{id}^k + C_1 \text{rand}(0, 1)(P_{id}^k - X_{id}^k) + C_2 \text{rand}(0, 1)(P_{gd}^k - X_{id}^k) \tag{21}$$

$$X_{id}^{k+1} = X_{id}^k + V_{id}^{k+1} \tag{22}$$

where w is an inertial factor, C_1 and C_2 are acceleration constants, $\text{rand}(0, 1)$ is a random number belonging to 0–1, P_{id} represents the d th dimension of the i th variable individual extremum, P_{gd} represents the d th dimension of the

global optimal solution and k represents the number of iterations.

3.2.2 Improvement of the PSO algorithm

In the standard PSO algorithm, the default w is 1, which indicates that the particles always fly along a certain direction at a constant speed during the search process until the search boundary is reached. Only when the optimal solution is exactly on the particle trajectory can the optimal solution be found. Shi and Eberhart introduced a linear decreasing inertia weight coefficient into the particle swarm velocity update formula (Shi and Eberhart 1998). The linear decreasing inertial weight has the same decline rate, so the region with the larger inertia weight accounts for only a small part of the total area. Inspired by research of Monteil and Beghdadi (1999), combined with the value range of linear decreasing weight is [0.4,0.9], S-curve function can be introduced into the inertial weight function, and the shape of the function can be changed by changing the form of the independent variable. Therefore, in this paper, an S-curve decreasing inertia weight is constructed after repeated adjustment, which expands the large inertial weight area. Its mathematical form is as follows:

$$w = 0.25 \cdot (1 - \tanh(a \cdot (n_{\text{now}} \cdot 60/n_{\text{iter}} - b))) + 0.4 \tag{23}$$

where n_{now} is the current number of iterations, n_{iter} is the maximum number of iterations, and a and b are parameters of S-curve. a is used to adjust the steepness of the maximum and minimum transition areas, and b is used to adjust the range of the transition area.

By changing the inertia weight w , the effect of the velocity of the particle at the last moment on the current velocity will change. Research shows that when the inertia weight is small, particles will search near their positions, and in a short time, particles will quickly gather together. At this time, the algorithm is similar to the local search algorithm. When the inertia weight is large, particles will continue to explore new areas, and the algorithm tends to be a global search algorithm. When an appropriate inertia weight is selected, the algorithm will have a greater probability of finding the optimal solution. To search according to different inertia weights in the same iterative optimization process, an S-curve decreasing inertia weight coefficient is designed. In the initial stage of the search, the larger inertia weight of the S-curve decline function decreases slowly, while in the search, the larger inertia weight tends toward a global search, so the inertia weight of the S-curve function makes the algorithm have good ergodicity in theory. At the end of the iteration, the inertia weight of the S-curve function reaches a smaller value faster. Because the smaller inertia weight tends toward a local search, theoretically, the inertia weight of the S-curve

function can obtain better local search performance than the linearly decreasing inertia weight. Therefore, PSO based on S-curve decreasing inertia weight (SDWPSO) is selected to initialize the weights and biases of the ELM.

Different S-curves can be obtained by adjusting the values of a and b , and the process of decreasing the inertia weight will be different. Figure 3 shows curves of the inertia weight decline with different combinations of a and b .

3.3 Factory case study and modeling industrial data collection

3.3.1 Description of a hot strip rolling mill in Chengsteel

A hot strip rolling mill in the HBIS group of the Chengsteel company, as shown in Fig. 4, is used to demonstrate the design and implementation of the hybrid PCA-SDWPSO-ELM model. The production line consisting of furnace, vertical mill, roughing mill, flying shear, finishing mill, laminar cooling device, coiler and the arrangement of each equipment is shown in Fig. 5.

In the process of hot strip rolling, basic automation, process automation, man-machine interfaces, material tracking systems and measurement instruments will produce a large amount of real-time production data. Data exchange is realized through industrial ethernet, so that all functions of the computer control system can be completed. A flowchart of data communication is shown in Fig. 6.

3.3.2 Data collection

The main sources of data collected are as follows: The first source is communication with other process computers. This part of the rolling data mainly includes incoming data, product size data and performance requirement data. The



Fig. 4 Photograph of a real hot strip rolling mill at Chengsteel

incoming data include slab number, steel coil number, material, blank size and chemical composition. The finished product size data generally include the target thickness, target temperature and target width. Performance requirement data include target yield strength, cooling rate and cooling temperature. The second part is the data from the instruments, which are the actual rolling data measured in the rolling process, and these data are the key data in the modeling process. They mainly include the data related to the stands and the independent data of the stands. The relevant data of the stands mainly include the rolling force data, the roll gap data and the speed of the motor. Independent data of the stand mainly include measured thickness, measured width and measured temperature. The third part is the process intervention data of HMI operators. The flow of hot strip rolling is shown in Fig. 7.

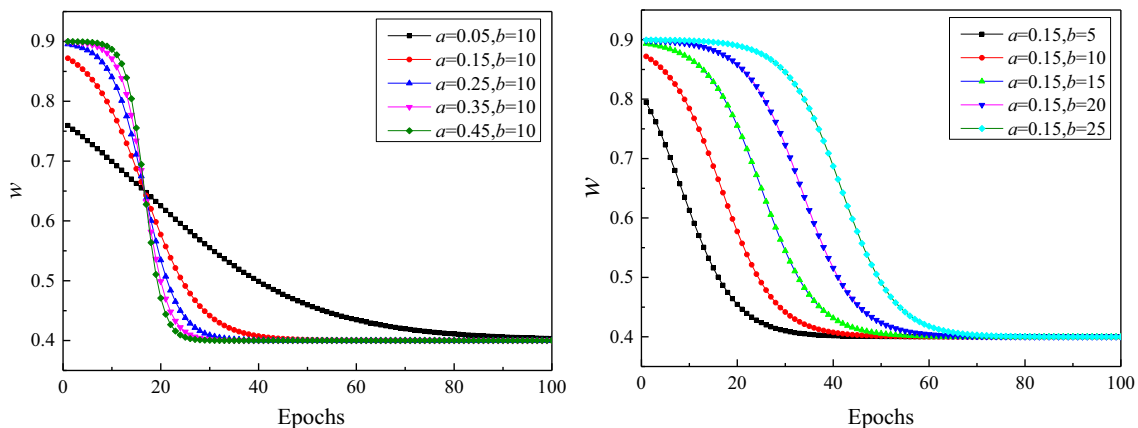


Fig. 3 Graphs of the S-curve function with different a and b values

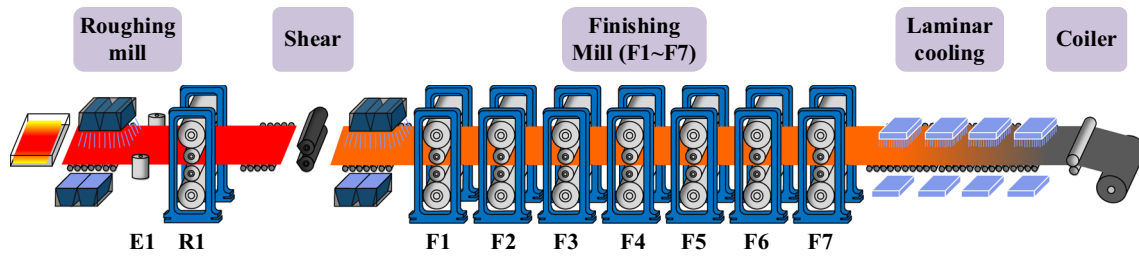


Fig. 5 Schematic layout of the hot strip rolling mill at Chengsteel

Fig. 6 Flowchart of data communication

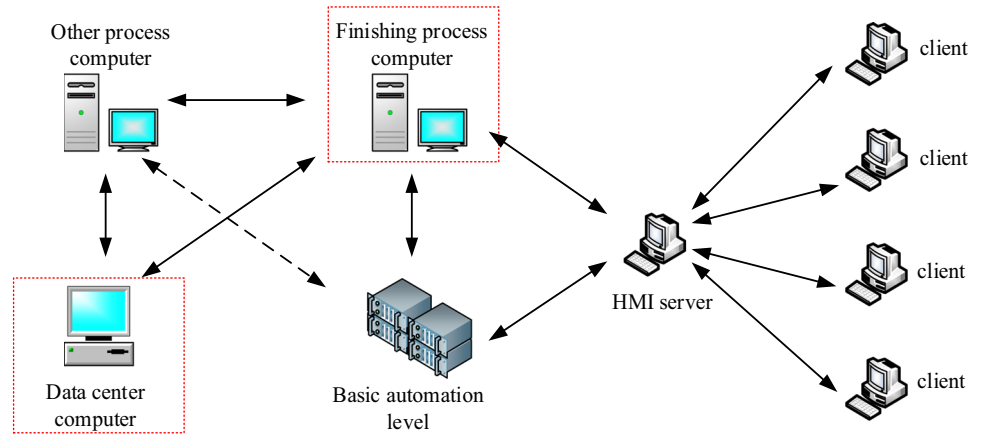
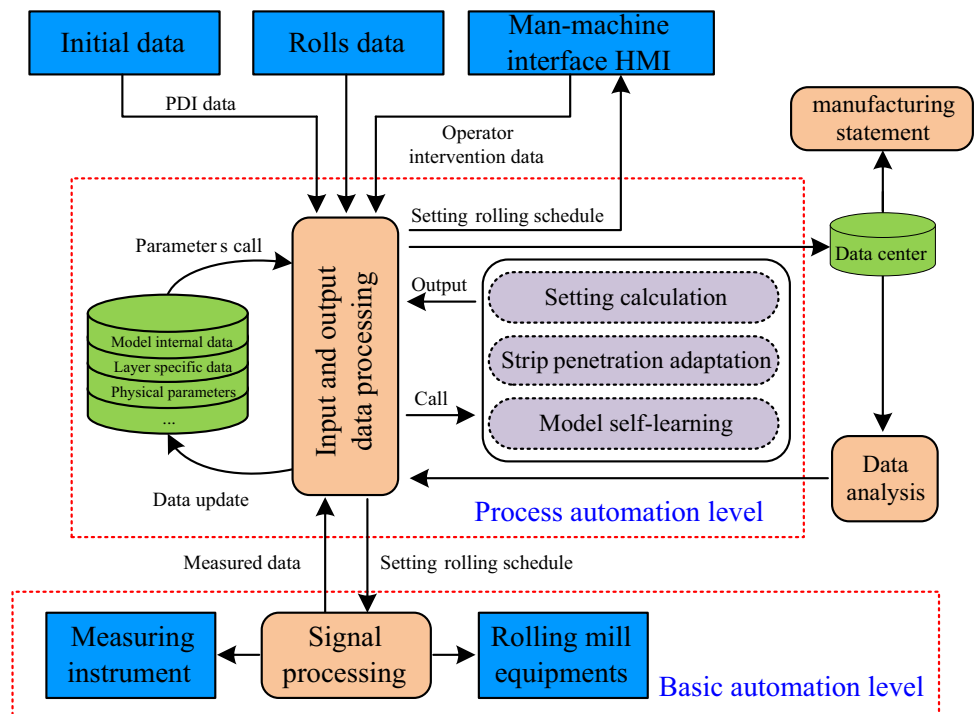


Fig. 7 Data flow in hot strip rolling



3.3.3 Determination of the input and output parameters of the model

According to the description of the theory of strip crown in Part 2, the influencing factors of strip crown mainly include the following aspects:

- Mill roll: diameter, roll length, roll thermal expansion, roll wear, etc.
- Strip steel: material (yield strength), strip width, strip thickness, temperature, etc.
- Rolling conditions: rolling force, bending force, roll shifting, roll speed, etc.

Based on the above principles, the variables listed in Fig. 8 are selected as input parameters of the model. The output variable is the exit strip crown after finishing mill.

3.3.4 Data preprocessing

Rolling data were collected from the data center for one week. Due to the erroneous data and outliers in these raw data, they cannot be directly used in modeling. Therefore, data preprocessing must be carried out. Preprocessing includes the following operations:

1. Removal of missing values.
2. Elimination of extraneous values.
3. Removal of outliers that are extremely deviated from the mean.

A total of 1809 strip samples are used as modeling datasets through the above operations. These data can be

divided into 8 layers according to the final rolling thickness. The sample number of each layer is shown in Fig. 9. From the perspective of modeling, the whole dataset can be divided into a training set and a test set. The sample data are normalized to $[-1, 1]$ (Han et al. 2013; Niu et al. 2016). The normalization formula as follows:

$$x'_i = 2 \times \frac{x_i - \min(x_i)}{\max(x_i) - \min(x_i)} + (-1), \quad i = 1, 2, 3, \dots, n \tag{24}$$

where $\max(x_i)$ and $\min(x_i)$ are the maximum and minimum numbers of data sequences, respectively.

3.3.5 Dimension reduction of sample data by PCA

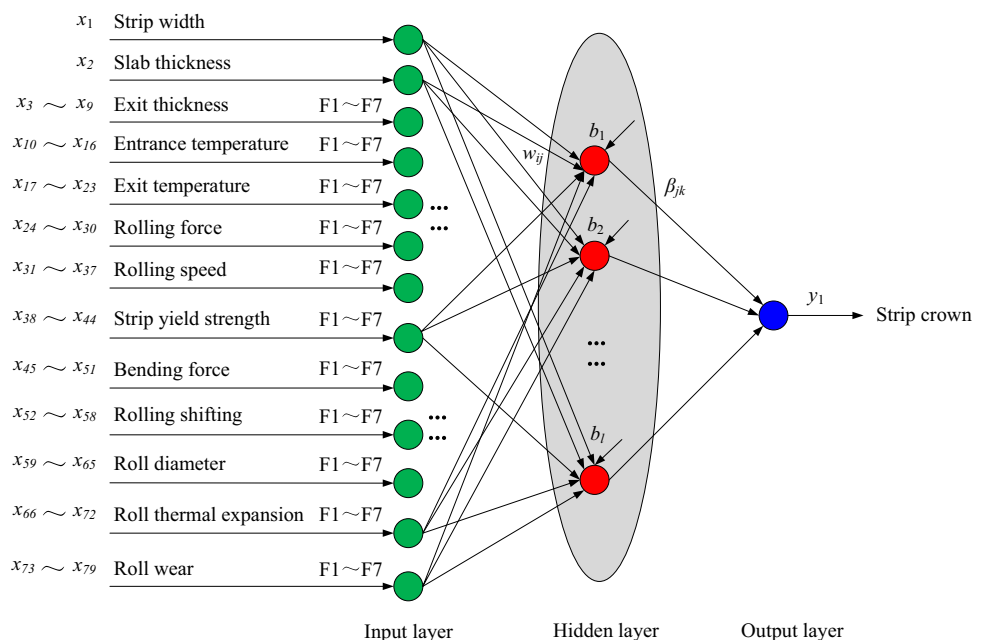
PCA is a multivariate statistical technique implemented to reduce dimensionality and to extract characteristic features from the original dataset. Suppose that the observed sample of the strip steel experimental data in this paper is \mathbf{X} :

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & & \vdots \\ x_{Q1} & x_{Q1} & \cdots & x_{Qn} \end{bmatrix} \tag{25}$$

\mathbf{X} is standardized by the following calculation:

$$x_{ij}^* = \frac{x_{ij} - \bar{x}_j}{\sqrt{\text{var}(x_j)}}, \quad i = 1, 2, \dots, Q; \quad j = 1, 2, \dots, n \tag{26}$$

Fig. 8 Input and output variables of the ELM models



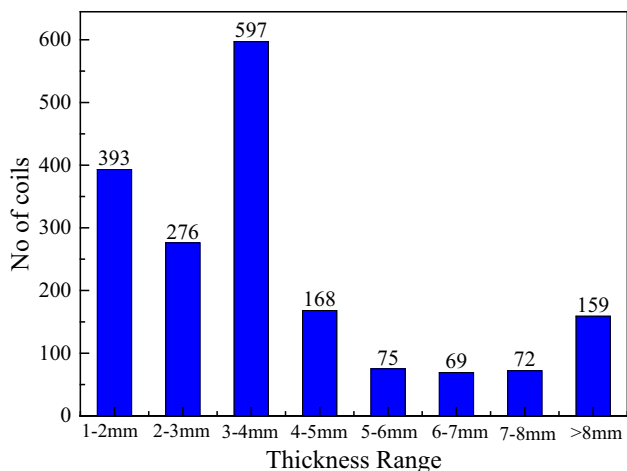


Fig. 9 Distribution of strip thickness for sample datasets

$$\bar{x}_j = \frac{1}{Q} \sum_{i=1}^Q x_{ij} \tag{27}$$

$$\text{var}(x_j) = \frac{1}{Q-1} \sum_{i=1}^Q (x_{ij} - \bar{x}_j)^2, \quad j = 1, 2, \dots, n$$

where \bar{x}_j is the mean value and $\sqrt{\text{var}(x_j)}$ is the standard deviation of $x_1, x_2, x_3, \dots, x_n$. The aim of data standardization is to eliminate the effects of different dimensions with respect to several data series.

The sample correlation coefficient matrix \mathbf{R} for the strip steel sample is calculated as

$$\mathbf{R} = \begin{bmatrix} r_{11} & r_{12} & \dots & r_{1n} \\ r_{21} & r_{22} & \dots & r_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ r_{n1} & r_{n2} & \dots & r_{nn} \end{bmatrix} \tag{28}$$

$$r_{ij} = \frac{\text{cov}(x_i, x_j)}{\sqrt{\text{var}(x_1)}\sqrt{\text{var}(x_2)}} = \frac{\sum_{k=1}^Q (x_{ki} - \bar{x}_i)(x_{kj} - \bar{x}_j)}{\sqrt{\sum_{k=1}^Q (x_{ki} - \bar{x}_i)^2} \sqrt{\sum_{k=1}^Q (x_{kj} - \bar{x}_j)^2}}, \quad Q > 1 \tag{29}$$

The eigenvalues and eigenvectors of the correlation coefficient matrix \mathbf{R} are calculated.

$$\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n \geq 0 \tag{30}$$

$$a_1 = \begin{bmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{n1} \end{bmatrix}, a_2 = \begin{bmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{n2} \end{bmatrix}, \dots, a_n = \begin{bmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{nn} \end{bmatrix} \tag{31}$$

n principal components can be obtained by PCA, and the first k principal components are selected according to the

cumulative contribution rate of each principal component. The contribution rate is defined as follows:

$$\eta_i = \frac{\lambda_i}{\sum_{k=1}^n \lambda_k} (i = 1, 2, \dots, n) \tag{32}$$

where η_i represents the explained variations of the i th principal component, and $\sum_{i=1}^n \eta_i = \sum_{k=1}^i \lambda_k / \sum_{k=1}^n \lambda_k (i = 1, 2, \dots, n)$ depicts the accumulative explained variations of principal components.

According to the principal component expression, the principal component score of the standardized sample is calculated, and the principal component matrix is defined as \mathbf{F} :

$$\mathbf{F} = \begin{bmatrix} F_{11} & F_{12} & \dots & F_{1k} \\ F_{21} & F_{22} & \dots & F_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ F_{Q1} & F_{Q2} & \dots & F_{Qk} \end{bmatrix} \tag{33}$$

$$F_{ij} = a_{j1}x_{i1} + a_{j2}x_{i2} + \dots + a_{jn}x_{in}, \tag{34}$$

$(i = 1, 2, \dots, Q; j = 1, 2, \dots, k)$

In this paper, the input variable that influences the strip crown in hot strip rolling is 79-dimensional data, which is catastrophic for the machine learning algorithm. Therefore, with a cumulative contribution rate of 0.95, the dimension of the modeling input variables is reduced.

3.4 Model development

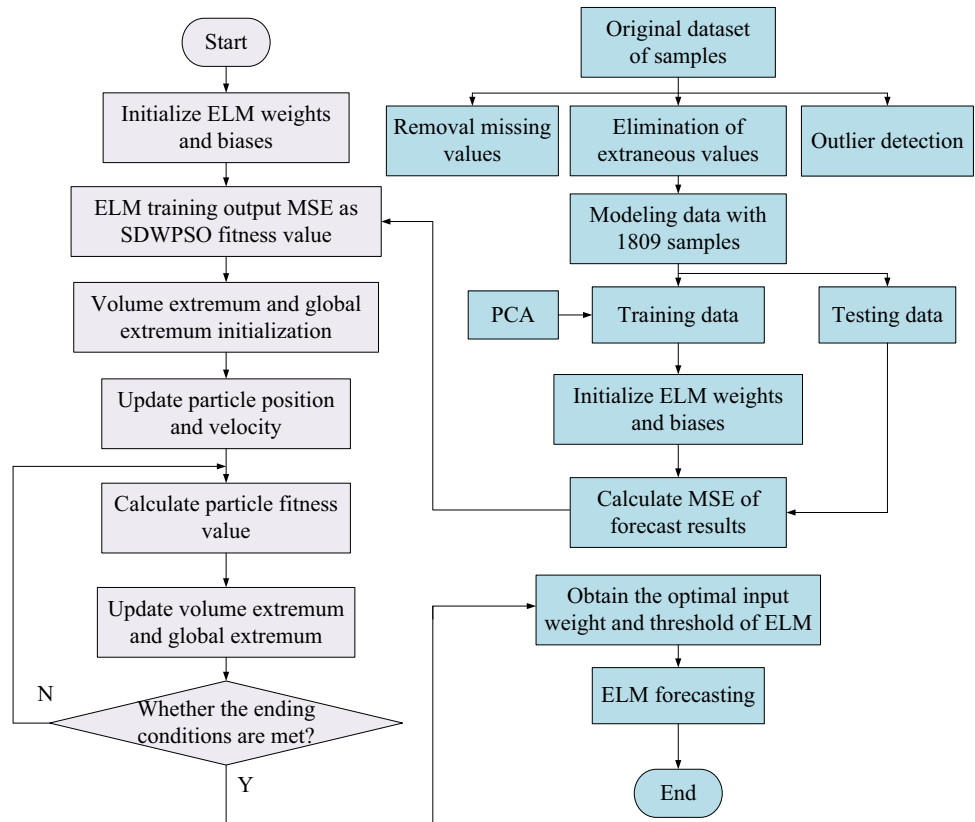
Among 1809 strip steel sample data, 83% (1500) are used as training set data, and 17% (309) are used as test set data. Four models are established and compared. A simple ELM without any optimization is named the single ELM model. The initial weights and biases of the ELM optimized by the SDWPSO algorithm are named the hybrid SDWPSO-ELM model. The ELM optimized by the PSO algorithm is named the hybrid PSO-ELM model. The hybrid SDWPSO-ELM model, which is modeled using the dimension reduction of independent variables by the PCA method, is named the hybrid PCA-SDEPSO-ELM model.

The four models are used to predict the strip crown of hot strip rolling and their comprehensive performance is evaluated. Figure 10 shows the main flowchart of the proposed hybrid SDWPSO-ELM model.

3.5 Model performance accuracy criteria

R^2 , MAE, MAPE and RMSE metrics are used to evaluate the comprehensive performance of each model. The formula for calculating the four indicators is as follows:

Fig. 10 Flowchart of the proposed model



$$R^2 = 1 - \frac{\sum_{i=1}^Q (y_i - y_i^*)^2}{\sum_{i=1}^Q (y_i - \bar{y})^2}, \left(\bar{y} = \frac{1}{Q} \sum_{i=1}^Q y_i \right) \quad (35)$$

$$MAE = \frac{1}{Q} \sum_{i=1}^Q |y_i - y_i^*| \quad (36)$$

$$MAPE = \frac{1}{Q} \sum_{i=1}^Q \left| \frac{y_i - y_i^*}{y_i} \right| \cdot 100\% \quad (37)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^Q (y_i - y_i^*)^2}{Q}} \quad (38)$$

where Q is the sample size; y_i, y_i^* is the actual output and predicted output crown of the i th strip sample, respectively; and \bar{y} is the average actual crown of the strip samples.

4 Strip crown prediction results and discussion

In this section, to show the performance of the proposed hybrid PCA-SDWPSO-ELM model, MATLAB is used to implement the model calculation. The average value of three implementations for each model is taken as the comprehensive performance.

4.1 Comparison of the search efficiencies of PSO and SDWPSO

With the same dataset and network parameters, the standard PSO and the SDWPSO algorithm are used to optimize the initial weights and biases of the ELM. Table 1 shows the performance of the two optimization algorithms on the test set with different numbers of hidden layer neurons. Regardless of the number of hidden layer neurons, the determination coefficient R^2 and MSE of the SDWPSO algorithm on the test set are better than those of the PSO algorithm, which fully proves the superiority of the SDWPSO algorithm.

When the topology of the ELM is 79-80-1, the variation in the best fitness value of the two algorithms is recorded. The variation rule of the best fitness value curves for the two algorithms is shown in Fig. 11. The diagram shows that the convergence rate of the fitness curve of the standard PSO algorithm is slow during the iteration process, and the fitness value process decreases gradually during the whole iteration process. It is stable after 80 iterations. By comparison, the improved SDWPSO algorithm quickly reaches the best fitness after approximately 40 iterations. In addition, the best individual fitness obtained by the standard PSO algorithm is 0.0145, while that of the SDWPSO algorithm is 0.0133. The result represented by SDWPSO is

Table 1 Performance of the PSO and SDWPSO algorithms on the test set

Number of hidden layer neurons	PSO algorithm		SDWPSO algorithm	
	R^2	MSE	R^2	MSE
20	0.5993	130.4134	0.6529	112.9932
40	0.6272	121.3071	0.7103	94.2611
60	0.7428	83.6819	0.7709	74.5358
80	0.7692	80.0299	0.8337	61.2996
100	0.8092	62.0823	0.8318	54.7187

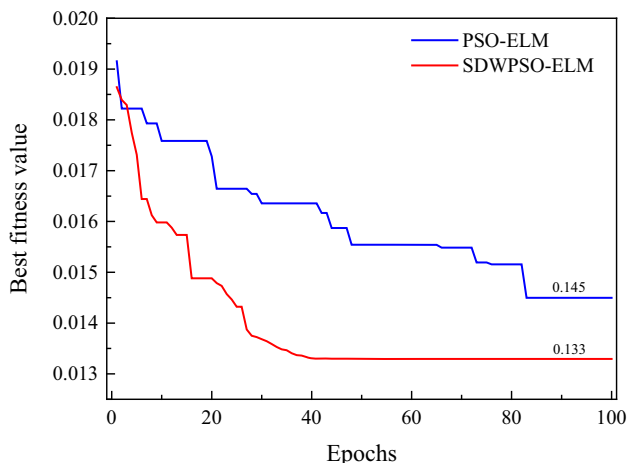


Fig. 11 Comparison of the optimization processes between of PSO and SDWPSO algorithms

more likely to be the global extremum of the solution space. After many tests, the same rule is observed. Therefore, the proposed SDWPSO algorithm has more advantages because of its characteristic of finding global extrema accurately and quickly.

4.2 Comparison of the prediction accuracies of different models

In this section, the comprehensive performance of the hybrid PCA-SDWPSO-ELM, single ELM, hybrid PSO-ELM and hybrid SDWPSO-ELM models is discussed in

Table 2 Parameters used in PSO

Parameters	Values
Hidden layer neurons	80
Activation function of the hidden layer neurons	Sigmoid
Population size	100
Acceleration factors	$C_1 = 2.4$ $C_2 = 1.6$
Number of iterations	100
Parameters of the S-curve	$a = 0.15$ $b = 10$

detail. The basic parameters of the models are shown in Table 2.

The scatter plots and regression effect of the four models on the training set and the test set are depicted in Fig. 12. The black straight line $y = x$ represents the ideal prediction model, which means that the predicted value is exactly the same as the actual value. In the actual process, the prediction models have difficulty achieving zero error, so the proximity of the regression line and ideal line is one of the important indices used to characterize its performance. In the graph, the red line represents the regression line of the data obtained on the training set used for modeling, while the blue line represents the regression line of the data obtained on the test set. The determination coefficient R^2 of models can be used to evaluate the proximity to the ideal case. $R^2 = 1$ means that the prediction is absolutely correct and there is no error. The smaller the value of R^2 is, the worse the performance of the model. Clearly, the proposed hybrid PCA-SDWPSO-ELM model has a more concentrated scatter distribution and presents higher R^2 values. (The R^2 values reached 0.7937 and 0.8573 on the training set and the test set, respectively.)

A comparison of the predicted values of the four models on the training set and the test set with the corresponding actual values is shown in Fig. 13. It can be roughly seen in the diagram that the predicted strip crown values of the samples are consistent with the actual values. The absolute error frequency distribution histograms and corresponding normal distribution curves of the four model predictions are shown in Fig. 14. Figure 14 shows that the absolute error distribution of the crown prediction value of the single ELM model is the most divergent. After PSO optimization of the ELM model, the concentration of the absolute error distribution is improved. After replacing PSO algorithm with SDWPSO optimization algorithm, the absolute error distribution of prediction results is more concentrated, and the most concentrated model is the proposed hybrid PCA-SDWPSO-ELM model. The dispersion of the normal curve can also be described by the standard deviation (SD). The smaller the SD represents the more concentrated the data distribution. The size of SD shown in Fig. 14 also proves that the hybrid PCA-SDWPSO-ELM model is the most superior both on the

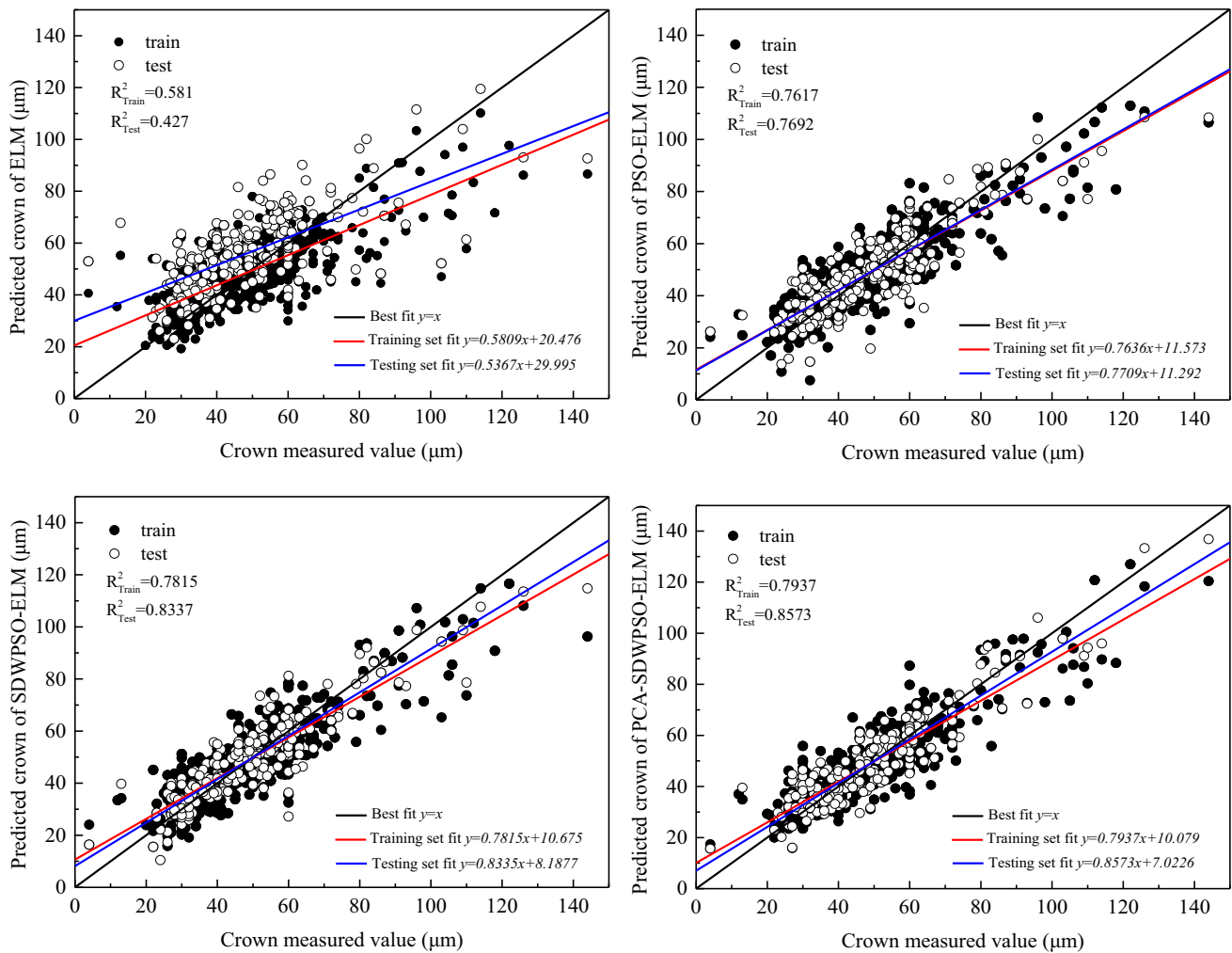


Fig. 12 Models regression effect on the training set and test set

training and testing set. Its prediction absolute error concentrated in the range of 0 and shape is high in the middle and low on both sides. The overall error distribution is approximately symmetrical, which means that the number of samples with large prediction error is smaller and the whole model is more stable and fault-tolerant.

More intuitive and quantitative performances of models are characterized by the MAE, MAPE and RMSE. Table 3 shows the specific calculation results of each error index, and the error distribution histogram is drawn according to Table 3, as shown in Fig. 15.

In comparison with that of the hybrid PSO-ELM model and single ELM model whose initial weights and biases are not optimized, the performance of the hybrid PSO-ELM model ($R^2 = 0.762$, MAE = 6.280, MAPE = 14.265% and RMSE = 8.428 for the training set and $R^2 = 0.769$, MAE = 6.441, MAPE = 17.623% and RMSE = 8.667 for the test set) is far better than that of the single ELM model ($R^2 = 0.581$, MAE = 7.972, MAPE = 18.458% and

RMSE = 11.177 for the training set and $R^2 = 0.427$, MAE = 12.189, MAPE = 35.434% and RMSE = 15.725 for the test set). This performance enhancement benefits from the PSO algorithm selecting the optimal initial weights and biases for the ELM.

When comparing the hybrid SDWPSO-ELM model with the hybrid PSO-ELM model, the performance of the hybrid SDWPSO-ELM model ($R^2 = 0.782$, MAE = 5.945, MAPE = 13.758% and RMSE = 8.071 for the training set and $R^2 = 0.834$, MAE = 5.358, MAPE = 13.613% and RMSE = 7.357 for the test set) is better than that of the hybrid PSO-ELM model. The inertia weight with the S-curve decreases slowly in the initial stage of the search, and the larger inertia weight tends ward a global search in the middle of the search. Therefore, the S-curve inertia weight makes the algorithm have better ergodicity in theory. At the end of the iteration, because the small inertia weight is more inclined to a local search, the S-curve

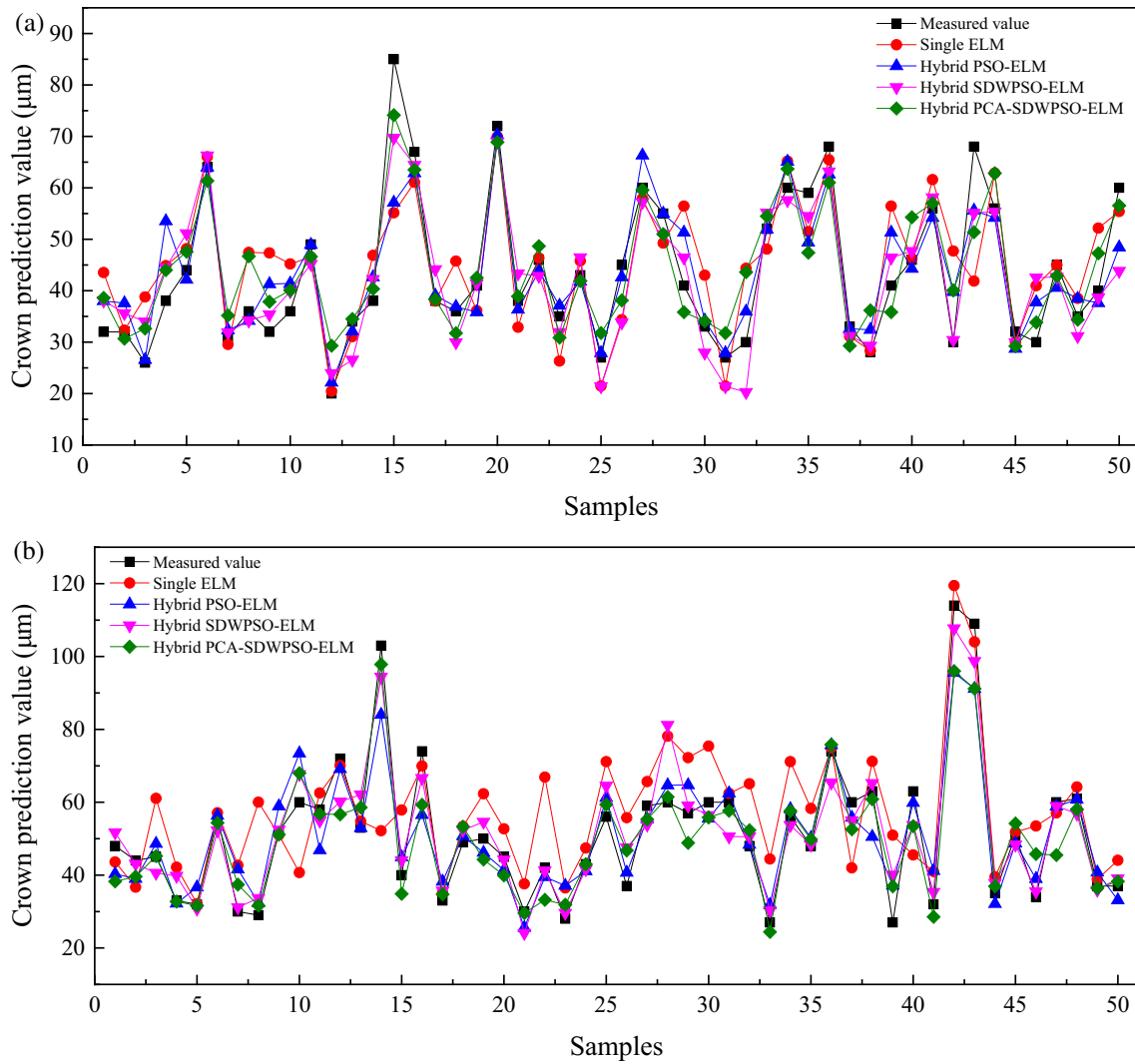


Fig. 13 Comparison between the predicted and actual values of the models. **a** Training set and **b** testing set

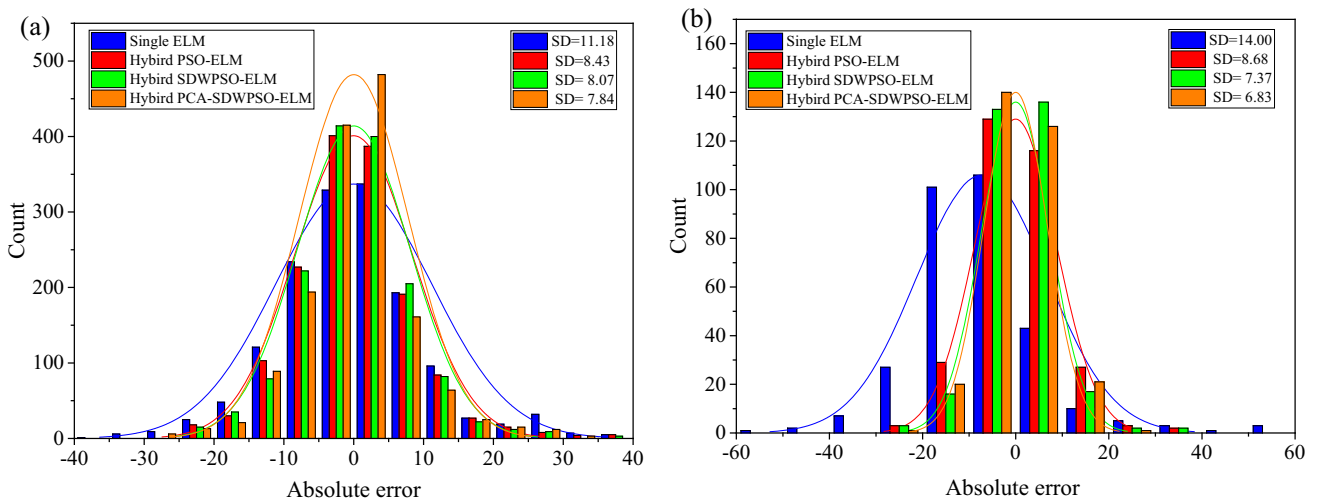


Fig. 14 Distributions of the absolute error of the four models. **a** Training set and **b** testing set

Table 3 Calculation results of the determination coefficient and error indicators

Indices	Models	Training set	Test set
R^2	Single ELM	0.581	0.427
	Hybrid PSO-ELM	0.762	0.769
	Hybrid SDWPSO-ELM	0.782	0.834
	Hybrid PCA-SDWPSO-ELM	0.794	0.857
MAE	Single ELM	7.972	12.189
	Hybrid PSO-ELM	6.280	6.441
	Hybrid SDWPSO-ELM	5.945	5.358
	Hybrid PCA-SDWPSO-ELM	5.726	5.165
MAPE (%)	Single ELM	18.458	35.434
	Hybrid PSO-ELM	14.265	17.623
	Hybrid SDWPSO-ELM	13.758	13.613
	Hybrid PCA-SDWPSO-ELM	13.108	13.429
RMSE	Single ELM	11.177	15.725
	Hybrid PSO-ELM	8.428	8.667
	Hybrid SDWPSO-ELM	8.071	7.357
	Hybrid PCA-SDWPSO-ELM	7.842	6.814

inertia weight reaches the smaller value more quickly and can obtain better local search performance.

When comparing the hybrid PCA-SDWPSO-ELM model with the hybrid SDWPSO-ELM model, the performance of the hybrid PCA-SDWPSO-ELM ($R^2 = 0.794$, MAE = 5.726, MAPE = 13.108% and RMSE = 7.842 for the training set and $R^2 = 0.857$, MAE = 5.165, MAPE = 13.429% and RMSE = 6.814 for the test set) is better than that of the hybrid SDWPSO-ELM model. After PCA processing with a cumulative contribution of 0.98, the independent modeling variables are reduced from 79-dimensional to 14-dimensional. It has the advantages of a high training speed and simple topology structure to use

the dataset after dimension reduction. Figure 16 shows the training time of each model. According to Fig. 16, the training time of the hybrid PCA-SDWPSO-ELM model is less than that of the hybrid SDWPSO-ELM model, which confirms this conclusion. A fast response is of great significance for industrial online control. Therefore, the hybrid PCA-SDWPSO-ELM model is more suitable for strip crown prediction in the online control process of hot strip rolling than other models.

With the synthesis of the above analyses, this study clearly demonstrates that it is effective to use industrial data and the hybrid PCA-SDWPSO-ELM approach to model the crown, which is one of the key strip shape parameters in the hot rolling process. This study can greatly reduce the traditional mathematical calculation without losing precision. More importantly, the research method proposed in this paper can be extended to other model parameter prediction and optimization processes, and it can effectively solve other nonlinear and strong coupling problems in the rolling processes.

5 Conclusions

An efficient approach for the prediction of strip crown in hot strip rolling processes based on a machine learning algorithm, called the hybrid PCA-SDWPSO-ELM model, is proposed in this research. The purpose of this study is to establish a soft measurement method based on production data for the accurate prediction of the strip crown and to improve the precision of shape control in hot strip rolling. The following conclusions can be drawn by comparing the comprehensive performance of the proposed model with that of the other three models.

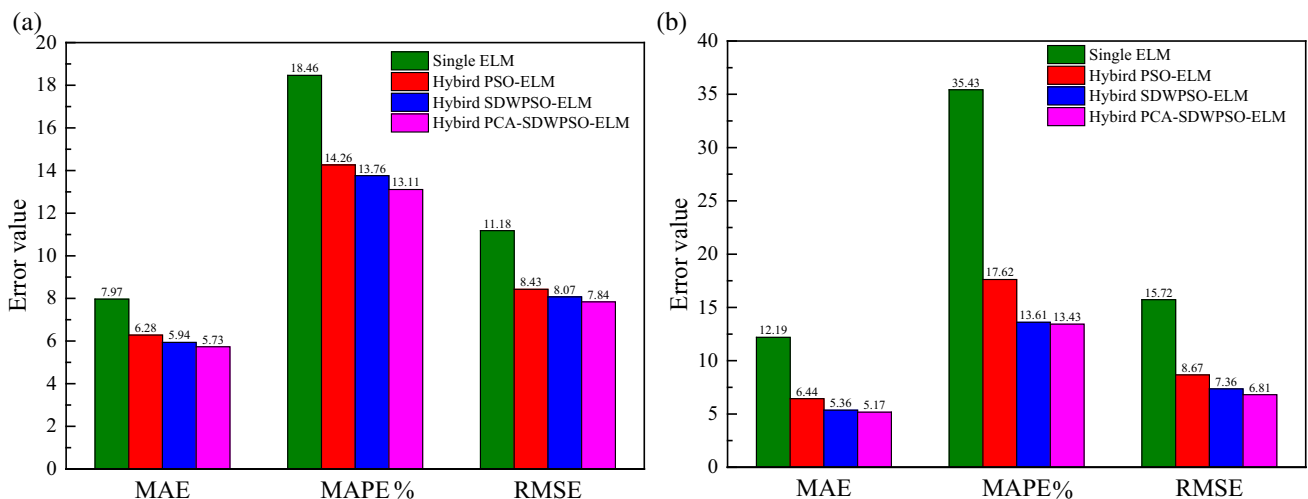


Fig. 15 Error distribution histogram of four models. **a** Training set and **b** test set

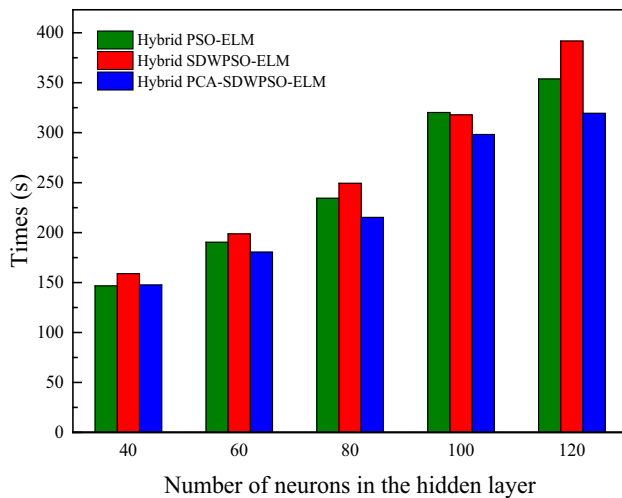


Fig. 16 Comparison of the training times required for each model

1. The proposed S-curve decreasing inertia weight PSO algorithm can greatly improve the search efficiency of the traditional PSO algorithm and overcome its shortcoming of easily falling into local minima. Based on this algorithm, the initial weights and biases of the ELM network are optimized and selected, and the accuracy of the hybrid SDWPSO-ELM model for predicting the strip crown in hot strip rolling is improved significantly.
2. The dimensionality reduction of independent variables is an effective method to deal with the modeling of industrial data. Training a model using a dataset after dimensionality reduction can not only improve the generalization performance of the model but also simplify the model structure and save modeling time. A less time consuming and quick response is beneficial to online real-time control systems in industry.
3. The combination of the ELM and industrial data can be used to predict the strip crown effectively. This data-driven prediction method can be easily extended to other parameter predictions and optimization by generating the corresponding training data in rolling processes. The research in this paper provides a new method to solve the multivariable, strong coupling and nonlinear complex industrial problems that cannot be handled by traditional mathematical models and provides technical support for the efficient utilization of massive data in hot strip rolling processes and the precision control of strip shape.

Author contributions ZW contributed to the conception of the study, performed the experiment; YL contributed significantly to analysis and manuscript preparation; ZW performed the data analyses and wrote the manuscript; TW helped perform the analysis with

constructive discussions; DG contributed to write, review and edit; DZ contribution was supervision.

Funding This study is financially supported by the National Natural Science Foundation of China (Nos.: 52105390, 51904206), National Key R&D Program of China (No.: 2018YFA0707300), Natural Science Foundation of Shanxi Province (No.: 201801D221130), Shanxi Province Science and Technology Major Projects (No.: 20181102015), Scientific and Technological Innovation Programs of Higher Education Institutions in Shanxi.

Data availability Enquiries about data availability should be directed to the authors.

Declarations

Conflict of interest The authors declare no conflict of interest.

Ethical approval This article does not contain any studies with human participants or animals performed by any of the authors.

Informed consent The authors declare that they do not need to obtain consent from any one for completing this research work.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1007/s00500-023-07895-6>.

References

- Chakraborti N, Kuamr BS, Babu VS, Moitra S, Mukhopadhyay A (2006) Optimizing surface profiles during hot rolling: a genetic algorithms based multi-objective optimization. *Comp Mater Sci* 37(1–2):159–165
- Deng J, Sun J, Peng W, Hu Y, Zhang D (2019) Application of neural networks for predicting hot-rolled strip crown. *Appl Soft Comput* 78:119–131
- Eberhart R, Kennedy J (1995) A new optimizer using particle swarm theory. In: *MHS'95. Proceedings of the sixth international symposium on micro machine and human science*. <https://doi.org/10.1109/MHS.1995.494215>
- Franceschi F, Cobo M, Figueredo M (2018) Discovering relationships and forecasting PM10 and PM2.5 concentrations in Bogotá, Colombia, using Artificial Neural Networks, Principal Component Analysis, and k-means clustering. *Atmos Pollut Res* 9(5):912–922
- Han F, Yao HF, Ling QH (2013) An improved evolutionary extreme learning machine based on particle swarm optimization. *Neurocomputing* 116:87–93
- Huang CM, Harris RW (1993) A comparison of several vector quantization codebook generation approaches. *IEEE T Image Process* 2:108–112
- Huang GB, Zhu QY, Siew CK (2004) Extreme learning machine: a new learning scheme of feedforward neural networks. In: *2004 IEEE international joint conference on neural networks, vol 2*, pp 985–990
- Huang GB, Zhu QY, Siew CK (2006) Extreme learning machine: theory and applications. *Neurocomputing* 70(1):489–501
- Huang GB, Zhou H, Ding X (2012) Extreme learning machine for regression and multiclass classification. *IEEE Trans Syst Man Cybern B Cybern* 42(2):513–529

- Ji YF, Song LB, Sun J, Peng W, Li HY, Ma LF (2021) Application of SVM and PCA-CS algorithms for prediction of strip crown in hot strip rolling. *J Cent South Univ* 28(8):2333–2344
- John S, Sikdar S, Swamy PK, Das S, Maity B (2008) Hybrid neural-GA model to predict and minimise flatness value of hot rolled strips. *J Mater Process Tech* 195(1–3):314–320
- Lan Y, Hu Z, Soh YC, Huang GB (2013) An extreme learning machine approach for speaker recognition. *Neural Comput Appl* 22(3–4):417–425
- Larkiola J, Myllykoski P, Korhonen AS, Cser L (1998) The role of neural networks in the optimisation of rolling processes. *J Mater Process Tech* 80–81(Suppl 5):16–23
- Li HJ, Xu JZ, Wang GD, Shi LJ, Xiao Y (2010) Development of strip flatness and crown control model for hot strip mills. *J Iron Steel Res Int* 17(3):21–27
- Li G, Gong D, Lu X, Zhang D (2021) Ensemble learning based methods for crown prediction of hot-rolled strip. *ISIJ Int* 61(5):1603–1613
- Linghu K, Zhao J, Li F, Wei D, Xu J, Zhang X, Zhao X (2014) 3D FEM analysis of strip shape during multi-pass rolling in a 6-high CVC cold rolling mill. *Int J Adv Manuf Tech* 74(9–12):1733–1745
- Liu HM, Zhang XL, Wang YR (2005) Transfer matrix method of flatness control for strip mills. *J Mater Process Tech* 166(2):237–242
- Malvoni M, De Giorgi MG, Congedo PM (2016) Photovoltaic forecast based on hybrid PCA-LSSVM using dimensionality reduced data. *Neurocomputing* 211:72–83
- Moazeni B, Salimi M (2015) Investigations on relations between shape defects and thickness profile variations in thin flat rolling. *Int J Adv Manuf Tech* 77(5–8):1315–1331
- Mohajer A, Daliri MS, Mirzaei A, Ziaeddini A, Nabipour M, Bavaghar M (2022a) Heterogeneous computational resource allocation for NOMA: toward green mobile edge-computing systems. *IEEE Trans Serv Comput* 2022:1–14
- Mohajer A, Sorouri F, Mirzaei A, Ziaeddini A, Rad KJ, Bavaghar M (2022b) Energy-aware hierarchical resource management and Backhaul traffic optimization in heterogeneous cellular networks. *IEEE Syst J* 2022:1–12
- Monteil J, Beghdadi A (1999) A new interpretation and improvement of the nonlinear anisotropic diffusion for image enhancement. *IEEE T Pattern Anal* 21(9):940–946
- Moussaoui A, Abbassi H (2006) Hybrid hot strip rolling force prediction using a bayesian trained artificial neural network and analytical models. *Am J Appl Sci* 3(6):1885–1889
- Nandan R, Rai RR, Jayakanth R, Moitra S, Chakraborti N, Mukhopadhyay A (2005) Regulating crown and flatness during hot rolling: a multiobjective optimization study using genetic algorithms. *Mater Manuf Process* 20(3):459–478
- Nikjoo F, Mirzaei A, Mohajer A (2018) A novel approach to efficient resource allocation in NOMA heterogeneous networks: Multi-criteria green resource management. *Appl Artif Intell* 32(7–8):583–612
- Niu P, Ma Y, Li M, Yan S, Li G (2016) A kind of parameters self-adjusting extreme learning machine. *Neural Process Lett* 44(3):813–830
- Peng Y, Liu H, Du R (2008) A neural network-based shape control system for cold rolling operations. *J Mater Process Tech* 202(1–3):54–60
- Peng K, Zhong H, Zhao L, Xue K, Ji Y (2014) Strip shape modeling and its setup strategy in hot strip mill process. *Int J Adv Manuf Tech* 72(5–8):589–605
- Peng K, Zhang K, Dong J, You B (2015) Quality-relevant fault detection and diagnosis for hot strip mill process with multi-specification and multi-batch measurements. *J Franklin I* 352(3):987–1006
- Pican N, Alexandre F, Bresson P (1996) Artificial neural networks for the presetting of a steel temper mill. *IEEE Expert* 11(1):22–27
- Pin G, Francesconi V, Cuzzola FA, Parisini T (2013) Adaptive task-space metal strip-flatness control in cold multi-roll mill stands. *J Process Control* 23:108–119
- Pittner J, Simaan M (2011) A useful control model for tandem hot metal strip rolling. *IEEE Trans Ind Appl* 46(6):2251–2258
- Samarasinghe S (2007) Neural networks for applied sciences and engineering: from fundamentals to complex pattern recognition. Auerbach publications, Boca Raton
- Shi Y, Eberhart R (1998) A Modified particle swarm optimizer. In: Proceedings of the IEEE international conference on evolutionary computation, IEEE Press, Piscataway
- Song L, Xu D, Wang X, Yang Q, Ji Y (2022) Application of machine learning to predict and diagnose for hot-rolled strip crown. *Int J Adv Manuf Tech* 120(1):881–890
- Sun J, Deng J, Peng W, Zhang D (2021) Strip crown prediction in hot rolling process using random forest. *Int J Precis Eng Man* 22:301–311
- Tran DC, Tardif N, Limam A (2015) Experimental and numerical modeling of flatness defects in strip cold rolling. *Int J Solids Struct* 69–70:343–349
- Tsai JT, Chou PY, Chou JH (2015) Performance comparisons between PCA-EA-LBG and PCA-LBG-EA approaches in VQ codebook generation for image compression. *Int J Electron* 102(11):1831–1851
- Wang ZH, Liu YM, Gong DY, Zhang DH (2018) A new predictive model for strip crown in hot rolling by using the hybrid AMPSO-SVR-Based approach. *Steel Res Int* 89(7):1800003
- Wang Z, Ma G, Gong D, Sun J, Zhang D (2019) Application of mind evolutionary algorithm and artificial neural networks for prediction of profile and flatness in hot strip rolling process. *Neural Process Lett* 50(3):2455–2479
- Yang L, Yu H, Wang D, Zhang Z (2017) Intelligent shape regulation cooperative model of cold rolling strip and its application. *Steel Res Int* 88(7):1600383
- Zhang XL, Zhao L, Zhao WB, Xu T (2015a) Novel method of flatness pattern recognition via cloud neural network. *Soft Comput* 19(10):2837–2843
- Zhang XL, Xu T, Zhao L, Fan H, Zang J (2015b) Research on flatness intelligent control via GA-PIDNN. *J Intell Manuf* 26(2):359–367
- Zhang XL, Cheng L, Hao S, Gao WY, Lai YJ (2016) The new method of flatness pattern recognition based on GA-RBF-ARX and comparative research. *Nonlinear Dyn* 83(3):1535–1548
- Zhou ZH (2012) Ensemble methods: foundations and algorithms. CRC Press, Boca Raton

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

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