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

Zhenhua Wang (wangzhenhuaneu@yeah.net)
Taiyuan University of Technology https://orcid.org/0000-0002-0934-3128
Yuanming Liu
Taiyuan University of Technology
Tao Wang
Taiyuan University of Technology
Dianyao Gong
Northeastern University
Dianhua Zhang
Northeastern University

Research Article

Keywords: strip crown prediction, extreme learning machine (ELM), industrial data, principal component analysis (PCA), regression model, hot strip rolling

Posted Date: September 8th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-433290/v1

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

Zhenhua Wang[1,2,4] Yuanming Liu[1,2,4], Tao Wang[1,2,4], Dianyao Gong[3], Dianhua Zhang[3]

[1] *College of Mechanical and Vehicle Engineering, Taiyuan University of Technology, Taiyuan 030024, Shanxi, PR China*

[2] *Engineering Research Center of Advanced Metal Composites Forming Technology and Equipment, Ministry of Education, Taiyuan 030024, Shanxi, PR China*

[3] *State Key Laboratory of Rolling and Automation, Northeastern University, Shenyang 110819, Liaoning, PR China*

[4] *TYUT-UOW Joint Research Centre, Taiyuan 030024, Shanxi, PR China*

* *Corresponding author. Tel.: +86-0351-6014008; E-mail address:* ***wangzhenhuaneu@yeah.net***

Abstract: The accurate prediction of strip crown is the precondition of the shape pre-set model in hot strip rolling. In this study, a new data-driven model of strip crown based on extreme learning machine (ELM) optimized by S-curve decreasing inertia weight PSO (SDWPSO) algorithm and industrial data is proposed. In order to simplify the model structure and save modeling time, principal component analysis (PCA) is used to reduce the dimension of the input data for modeling samples. The comprehensive performance of the proposed hybrid PCA-SDWPSO-ELM prediction model is evaluated by several error indexes. The superiority of the proposed model is also proved by comparing the prediction results with other three comparison models. The research shows that the hybrid PCA-SDWPSO-ELM method can solve the problem of nonlinear and strong coupling in the traditional engineering. It is suitable for the parameters prediction and optimization of 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 rolled strip products are widely used in the national economy (Pittner and Simaan 2011; Peng et al. 2015). Strip shape, including crown and flatness of strip, is one of the key indexes of product quality. The crown represents the difference thickness distribution between the two sides and the middle on the cross section of the strip and flatness represents the different elongation 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, shearing and so on. So, shape control is the core, frontier and high difficulty technology in hot rolling process. Naturally, the researches on shape prediction and control have great theoretical significance and practical value.

So far, many scholars have carried out intensive study in the field of shape control in rolling process. The common research methods of strip shape control are traditional mathematical analysis, finite element analysis and artificial intelligence. For strip crown control, it is generally effective to combine all kinds of factors affecting strip crown, such as detection device, mathematical model, load distribution, bending and shifting system (Pin et al. 2013), therefore, crown prediction has the characteristics of multivariable, nonlinear, strong coupling and so on. Its intrinsic mechanism is very complex and it is difficult to obtain accurate prediction model. It is the most basic method to establish the strip shape control model through the traditional mathematical analysis, the core idea of this method is to fully consider the characteristics of rolling mill model and metal flow law, and to calculate the elastic deformation and flexural deformation of rolling mill by using the influence function method (Li et al. 2010). Combined with bending roll and shifting roll strategy, the shape of roll gap profile is calculated to realize crown control (Peng et al. 2014). The finite element method can simulate the metal flow law and stress and strain of strip under various rolling conditions flexibly (Moazeni and Salimi 2015). Through the reasonable grid dividing and setting boundary conditions, it provides the most accurate representation for the roll system force and deformation of the whole rolling mill, so the calculation results of the shape parameters with high precision can be obtained, 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 more and more difficult to improve the precision of traditional shape mathematical model because of the many simplified conditions in the process of solving mathematical analytic method and finite element method.

In order to continuously improve the strip shape control accuracy, the artificial intelligence technology with data and algorithm as the core has attracted more and more scholars' attention. Data-driven modeling is generally realized by machine learning algorithm. The common machine learning algorithms mainly include artificial neural network (ANN), support vector machine (SVM), decision tree and random forest (RF), deep learning and so on. It is the earliest application of artificial intelligence in rolling field to establish rolling force prediction model of leveling rolling process using neural network and apply it to practical production line and good results have been obtained (Larkiola et al. 1998; Pican et al. 1996; Moussaoui and Abbassi 2006). The pattern recognition method is generally used in the shape control system, and the accurate shape standard pattern characteristic coefficient is the premise of the shape control, T-S cloud inference neural network (Zhang et al. 2015), PID neural network (Zhang et al. 2015) and radial basis function network (Zhang et al. 2016) can effectively identify common defects in cold rolling and improve the precision of cold rolled shape control. SVM is another common machine learning algorithm, which can realize model training under small sample size. SVM can be used to accurately predict the outlet crown of hot strip rolling (Wang et al. 2018). Data-driven modeling based on a single learner is prone to overfitting, and ensemble modeling methods can effectively solve this problem. RF is a typical integrated learning algorithm, which can also be used to establish outlet crown prediction model of the hot rolled strip with good generalization performance (Sun et al. 2021). Hybrid model is constructed based on hot rolling production data and deep learning network to predict strip exit crown, 97.04% absolute error of modeling samples is less than 5μm (Deng et al. 2019). Combining machine learning algorithm with heuristic intelligent optimization algorithm to establish data-driven model is also another research direction of intelligent modeling. Using genetic algorithm (GA) to establish multi-objective optimal control strategy and apply it to the identification of strip shape parameters and the setting of rolling schedule, the optimal values of strip crown and flatness setting values can be obtained successfully, and the precision of strip shape control can be improved (Nandan et al. 2005). GA and ant colony algorithm (ACA) are used to optimize the crown model of hot rolled strip, which proves that the evolutionary algorithm is practical in the optimization of rolling process parameters (Chakraborti et al. 2006). The relationship model between input parameters and strip shape can also be established by combining ANN and GA to predict the minimization flatness value of hot rolled strip (John et al. 2008). Besides, the transfer matrix between the characteristic parameters of flatness error and the parameters of flatness adjustment can be established by using GA to optimize the BP neural network. The transfer matrix is successfully applied to the strip shape adjustment mechanism of rolling mill to realize the accurate control of the strip shape (Liu et al. 2005; Peng et al. 2008). Combining the shape control matrix with the differential evolution algorithm (DE) optimization ELM, the intelligent cooperative control model of the shape control mechanism of cold rolled strip can be established and applied to the shape control process (Yang et al. 2017). Based on the above analysis, combination of big data technology and artificial intelligence modeling method is a new trend to study how to further improve the precision of shape control in rolling process.

In this study, a new hybrid PCA-SDWPSO-ELM forecasting model is proposed in combination with the artificial intelligence method and the industrial data to predict strip crown in hot rolling. The superiority of the proposed model is proved by contrast experiment. This paper is organized as follows: Section 2 introduces the basic theory of shape control. Sections 3 shows the collection and processing of modeling data and the related modeling process. The discussion of strip crown forecasting results is described explicitly in section 4 and Section 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. In order to eliminate the effect of strip edge thinning, the edge reference point is usually located at the 40 mm distance from the strip edge. The definition of strip crown is shown in **Fig.1**. The proportional crown is the ratio of the strip crown to the thickness of the strip center.

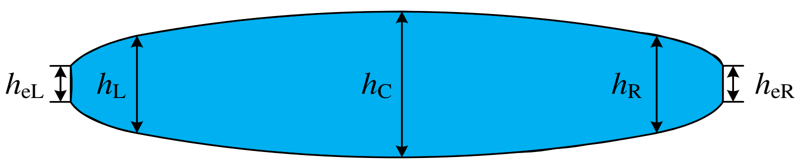

Fig.1. Thickness variation in strip cross section

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

$$Cp_h = \frac{C}{h_c} \tag{2}$$

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; Cp_{h} is proportional crown of the strip.

2.2 The unload roll gap crown model

The traditional crown control model consists of two parts: the unload roll gap crown model and the uniform load roll gap crown model. The unload roll gap crown is the roll gap crown of the rolling mill without workpiece and without adding force, which reflects the effect of roll crown on the shape of strip, and it is one of the important factors that affect the shape of load roll gap. As shown in **Fig.2**, the unload roll gap crown consists of two parts: roll gap between work rolls and roll gap between back-up roll and work roll.

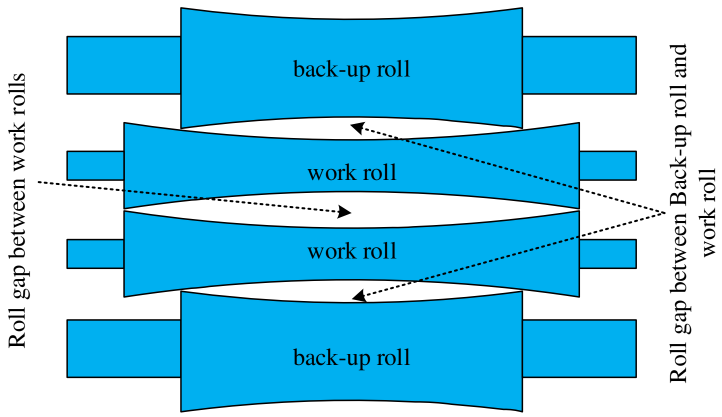

Fig.2. Schematic plan of unloaded roll gap

2.2.1 Roll crown model

The calculation of roll crown is the premise of unload roll gap crown calculation. 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. Roll thermal crown and roll wear crown are the crown formed by thermal expansion and wear during rolling process. The equivalent crown of roll is 0 for conventional mill, and for CVC mill, it can be calculated by interpolation of transverse position.

$$C_R = C_{\mathrm{grn}} + C_{\mathrm{eqv}} + C_{\mathrm{t}} + C_{\mathrm{w}} \tag{3}$$

where, C_R is roll crown, mm; C_{grn} is roll original grinding crown, mm; C_{eqv} is roll equivalent crown, mm; C_{t} is roll thermal crown, mm; C_{w} is roll wear crown, mm.

2.2.2 Roll gap crown between back-up roll and work roll

The gap crown between the back-up roll and the work roll is determined by the work roll crown and the back-up roll crown. Because it corresponds to the contact area between rollers, it is necessary to transform the crown of the work roll under the assumption that the roll crown curve is a conic distribution.

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

where, $C_{\mathrm{br\text{-}wr}}$ is the gap crown between the back-up roll and the work roll, mm; C_{br} is the back-up roll crown, mm; C_{wr} is the work roll crown, mm; L_{br} is the back-up roll length, mm. L_{wr} is the work roll length.

2.2.3 Roll gap crown between two work rolls

Work roll gap crown is the roll gap crown between two work rolls when no load is carried, and the size is equal to the work roll crown.

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

where, $C_{\mathrm{wr\text{-}wr}}$ is roll gap crown between two work rolls, mm; C_{wr} is work roll crown, mm.

2.3 The 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 uniform load roll gap depends on the unload roll gap crown, the roll system deflection and the elastic flattening deformation of the rolls caused by the rolling force and the bending force. The mathematical model of 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 back-up roll and work roll, and roll gap crown between two work rolls. The mathematical model is constructed as follows (Peng et al. 2014):

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

where, C_{ufd} is the uniform load roll gap crown, mm; p is the unit wide rolling force, kN/mm; F_{w} is bending force, kN; D_{wr} and D_{br} are the diameter of work roll and back-up roller, respectively, mm; E_{wr} is the elastic modulus of work roll, MPa; 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; $c_{i,j}$ are polynomial coefficients.

3 Methodology

3.1 ELM regression modeling

The ELM is a kind of single-hidden Layer Feedforward Neural Network (SLFN), which is proposed by in 2004 (Huang et al. 2004; Huang et al. 2006; Huang et al. 2012). The purpose is to simplify the learning parameters setting while overcoming the defect of the BP algorithm that is easy to fall into the local minimum and improve the learning efficiency (Lan et al. 2013).

$\boldsymbol{w}$ is defined as the connection weight of the input layer and the hidden layer, $\boldsymbol{\beta}$ is defined as the connection weight of the hidden layer and the output layer.

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

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

$\boldsymbol{b}$ is defined as the biases of the hidden layer neurons; n、l、m are number of neurons in input layer, hidden layer and output layer, respectively.

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

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

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

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

$g\left(x\right)$ is the activation function of the hidden layer neuron. $\boldsymbol{T}$ is the output of the network.

$$\boldsymbol{T}=\left[\boldsymbol{t}_1,\boldsymbol{t}_2,\boldsymbol{t}_3,\cdots,\boldsymbol{t}_Q\right]_{m\times Q},\boldsymbol{t}_j=\begin{bmatrix} t_{1j} \\ t_{2j} \\ \vdots \\ t_{mj} \end{bmatrix}_{m\times 1}=\begin{bmatrix} \sum_{i=1}^{l}\beta_{i1}g\left(\boldsymbol{w}_i\boldsymbol{x}_j+b_i\right) \\ \sum_{i=1}^{l}\beta_{i2}g\left(\boldsymbol{w}_i\boldsymbol{x}_j+b_i\right) \\ \vdots \\ \sum_{i=1}^{l}\beta_{im}g\left(\boldsymbol{w}_i\boldsymbol{x}_j+b_i\right) \end{bmatrix}_{m\times 1} \quad \left(j=1,2,\cdots,Q\right) \tag{13}$$

where, $\boldsymbol{w}_i=\left[w_{i1},w_{i2},\cdots,w_{in}\right]$; $\boldsymbol{x}_j=\left[x_{1j},x_{2j},\cdots,x_{nj}\right]^{\mathrm{T}}$.

$\boldsymbol{H}$ is the hidden layer output matrix, the specific forms are as follows:

$$\boldsymbol{H}\left(\boldsymbol{w}_1,\boldsymbol{w}_2,\cdots,\boldsymbol{w}_l,b_1,b_2,\cdots,b_l,\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_Q\right)=\begin{bmatrix} g\left(\boldsymbol{w}_1\cdot\boldsymbol{x}_1+b_1\right) & g\left(\boldsymbol{w}_2\cdot\boldsymbol{x}_1+b_2\right) & \cdots & g\left(\boldsymbol{w}_l\cdot\boldsymbol{x}_1+b_l\right) \\ g\left(\boldsymbol{w}_1\cdot\boldsymbol{x}_2+b_1\right) & g\left(\boldsymbol{w}_2\cdot\boldsymbol{x}_2+b_2\right) & \cdots & g\left(\boldsymbol{w}_l\cdot\boldsymbol{x}_2+b_l\right) \\ \vdots & \vdots & & \vdots \\ g\left(\boldsymbol{w}_1\cdot\boldsymbol{x}_Q+b_1\right) & g\left(\boldsymbol{w}_2\cdot\boldsymbol{x}_Q+b_2\right) & \cdots & g\left(\boldsymbol{w}_l\cdot\boldsymbol{x}_Q+b_l\right) \end{bmatrix}_{Q\times l} \tag{14}$$

Represented by a matrix:

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

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 $\boldsymbol{w}$ and $\boldsymbol{b}$ for SLFN, it means that

$$\sum_{j=1}^{Q}\left\|\boldsymbol{t}_j-\boldsymbol{y}_j\right\|=\mathbf{0} \tag{16}$$

$$\boldsymbol{y}_j=\left[y_{1j},y_{2j},\cdots,y_{mj}\right]^{\mathrm{T}}\left(j=\mathbf{1},\mathbf{2},\cdots,Q\right) \tag{17}$$

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

$$\sum_{j=1}^{Q}\left\|\boldsymbol{t}_j-\boldsymbol{y}_j\right\|<\varepsilon \tag{18}$$

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

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

Its solution is

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

Where $\boldsymbol{H}^{+}$ is the Moore-Penrose generalized inverse of the implicit layer output matrix $\boldsymbol{H}$.

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, which was originally proposed by Kennedy and Eberhart (Eberhart and Kennedy 1995). All the particles in the group adjust 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:

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

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

where the w is inertial factor, C_1 and C_2 are called the acceleration constant, $rand\left(0,1\right)$ is random number belong to 0-1, P_{id} represents the dth dimension of the ith variable individual extremum, P_{gd} represents the dth dimension of the global optimal solution and k represents the number of iterations.

3.2.2 Improvement of PSO algorithm

In the standard PSO algorithm, the default w is 1, which represents 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 that the region with the larger inertia weight accounts for only a small part of the total area. Therefore, in this paper, a S-curve decreasing inertia weight is constructed, which expands the area of large inertial weight. Its mathematical form is as follow:

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

where, n_{now} is the current number of iterations, n_{iter} is the maximum number of iterations, a and b are parameters of S-curve.

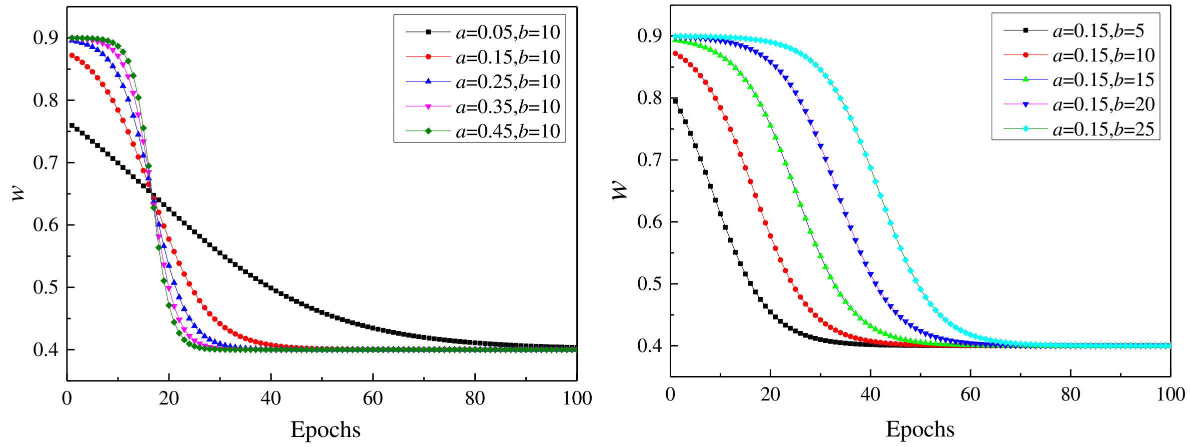

Fig.3. Graphs of S-curve function with different a and b

Table 1 The parameters used in PSO.

Parameters	Values
Size of population	100
Acceleration factors	C_1=2.4 C_2=1.6
Number of iterations	100
Parameters of S-curve	a=0.15 b=10

Different S-curves can be obtained by adjusting the value of a and b, and the process of decreasing inertia weight will be different. **Fig.3** shows the inertia weight decline curves under different combinations of a and b. The SDWPSO algorithm parameter settings as shown in **Table 1.**

3.3 Case study factory and modeling industrial data collection

3.3.1 Description of hot tandem rolling mill in Chengsteel

Hot tandem rolling mill in HBIS group 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 consists of furnace, vertical mill, roughing mill, flying shear, finishing mill, laminar cooling device, coiler and arrangement of each equipment is shown in **Fig.5.**

Fig.4. A real hot tandem rolling mill photo of Chengsteel

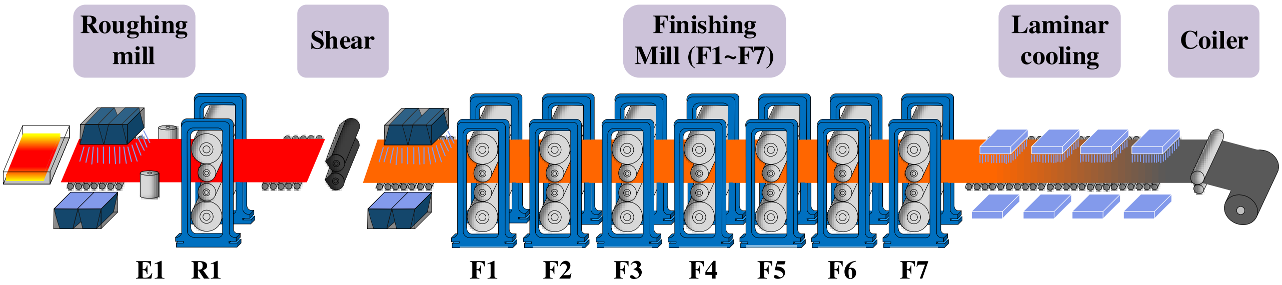

Fig.5. Schematic layout of the hot tandem rolling mill in Chengsteel

In the process of hot strip rolling, basic automation, process automation, man-machine interface, material tracking system and measurement instrument will produce a large number of real-time production data. Data exchange is realized through industrial ethernet, so that all functions of computer control system can be completed. The flow chart of the data communication is shown in **Fig.6**.

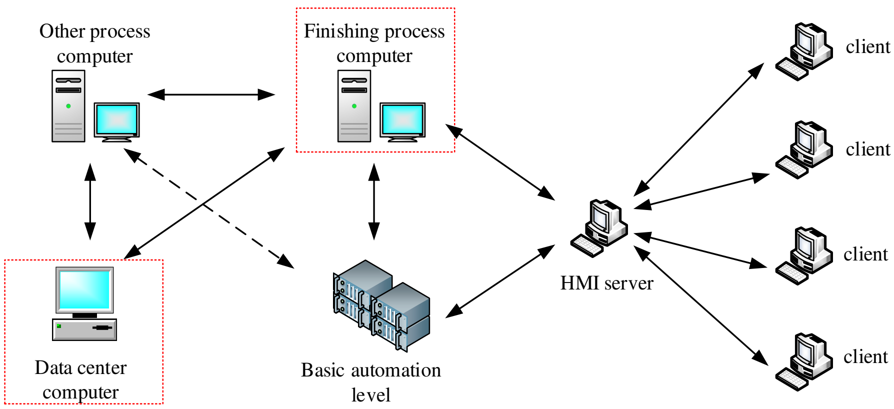

Fig.6. The flow chart of the data communication

3.3.2 Data collection

The main sources of data collected are as follows: the first part is the communication with other process computers. This part of rolling data mainly includes incoming data, product size data and performance requirements data. The incoming data include slab number, steel coil number, material, blank size, chemical composition, etc. The finished product size data generally include target thickness, target temperature and target width, etc. Performance requirements data include target yield strength, cooling rate and cooling temperature. The second part is the data from the instruments, which are mainly the actual rolling data measured in the rolling process, these data are the key data in the modeling process. It mainly includes the data related to the stands and the independent data of the stands. The relevant data of the stands mainly include the data of rolling force, the data of the roll gap and the speed of the motor, etc. Independent data of the stand mainly include pre-rolling and post-rolling meter data, such as measured thickness, measured width and measured temperature. The third part is the process intervention data of HMI operators. The hot rolling data flow is shown in **Fig.7.**

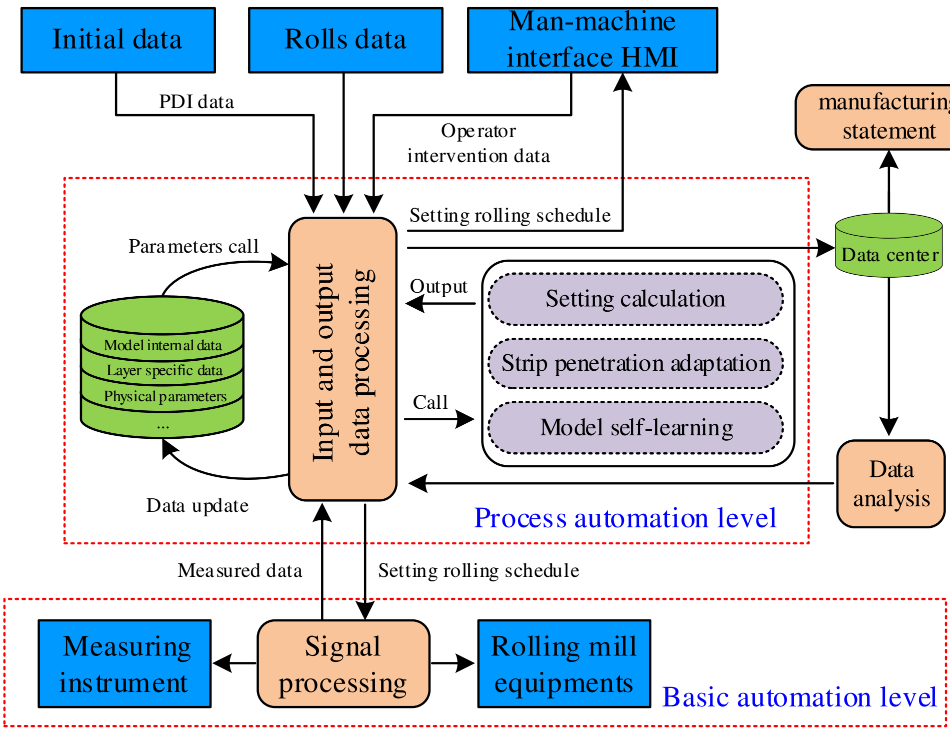

Fig.7. The data flow in hot strip rolling

3.3.3 Determine the input and output parameters of the model

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

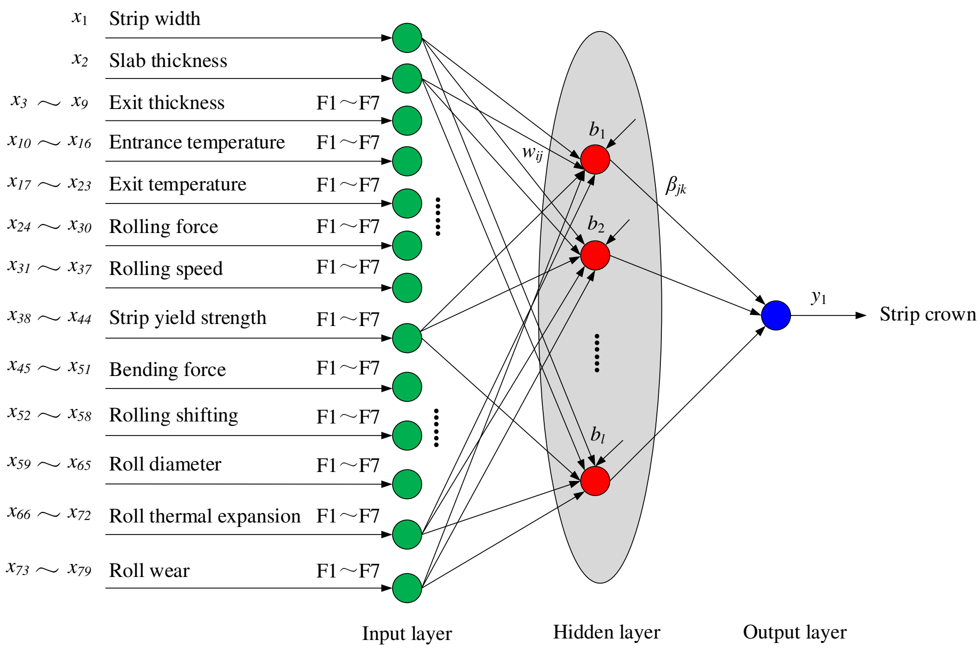

Fig.8. Input and output variables of the ELM models

- 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

A week of rolling data was collected from the data center. Due to the erroneous data and outliers in these raw data, it cannot be directly used in modeling. So, 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 which are extremely deviated from the mean.

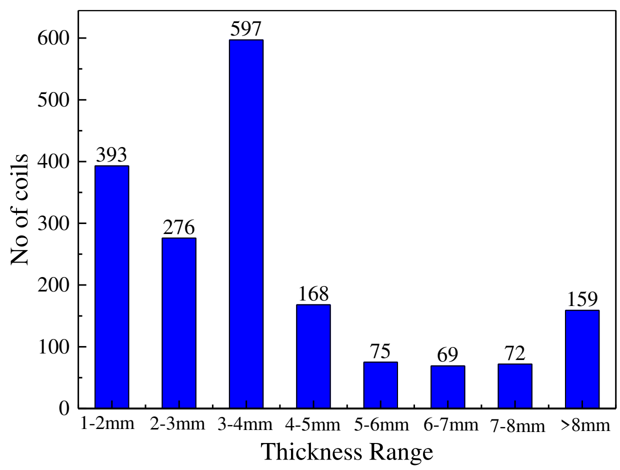

Fig.9. Distribution of strip thickness on sample data sets

A total of 1809 strip samples are used as modeling data sets 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 point of view of modeling, the whole data set can be divided into training set and test set. The sample data is normalized to [-1,1] (Han et al. 2013; Niu et al. 2016). Normalization formula as follows

$$x_i^{'} = 2 \times \frac{x_i - \min(x_i)}{\max(x_i) - \min(x_i)} + (-1), i = 1, 2, 3, ..., m \tag{24}$$

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

3.3.5 Dimension reduction of sample data by PCA

PCA is a multi-element statistical method for converting multiple indexes into several comprehensive indexes on the premise of low loss of information by using the thought of dimension reduction (Samarasinghe 2007). Usually, the comprehensive index generated by transformation is called as the principal component, in which each principal component is a linear combination of the original variables, and each principal component is independent of each other, which makes the principal component have some superior performance than the original variable. Refs. (Malvoni et al. 2016; Franceschi et al. 2018) show the implementation strategies of PCA. From the above analysis, the input variable that influences the crown of strip steel is a 79-dimensional data, which is catastrophic for the machine learning algorithm. This paper deals with dimension reduction under the premise of cumulative contribution rate of 0.95.

3.4 Models 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. Simple ELM without any optimization is named as single ELM model. Initial weights and biases of ELM optimized by SDWPSO algorithm is named as hybrid SDWPSO-ELM model. ELM optimized by PSO algorithm is named as hybrid PSO-ELM model. The hybrid SDWPSO-ELM model, which is modeled using reduction dimension of independent variables by PCA method, is named hybrid PCA-SDEPSO-ELM model.

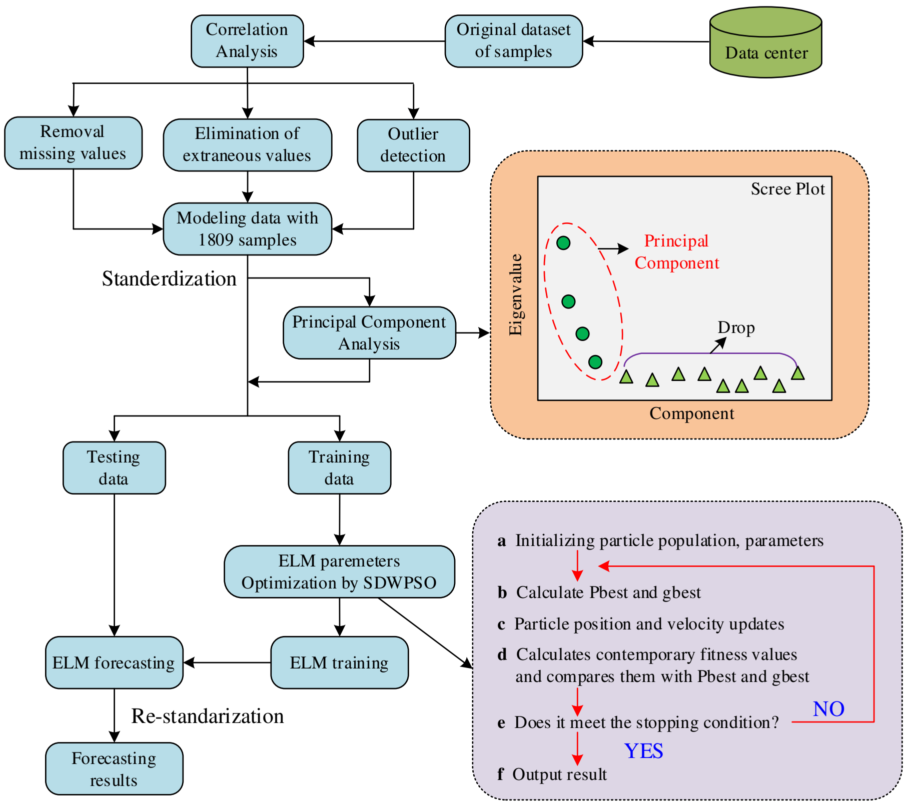

Fig.10. Schematic layout of the proposed model.

The four models are used to predict the exit crown of hot strip rolling and their comprehensive performance are evaluated. The **Fig.10** shows the main flow of the proposed hybrid SDWPSO-ELM models.

3.5 Model performance accuracy criteria

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

$$R^2 = 1 - \frac{\sum_{i=1}^{n}\left(y_i - y_i^*\right)^2}{\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2}, \left(\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i\right) \tag{25}$$

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}\left|y_i - y_i^*\right| \tag{26}$$

$$\mathrm{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - y_i^*}{y_i}\right| \cdot 100\% \tag{27}$$

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}\left(y_i - y_i^*\right)^2}{n}} \tag{28}$$

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

4 Strip crown prediction results and discussion

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

4.1 Comparison of search efficiency between PSO and SDWPSO

Under the same data set and network parameters, the standard PSO and the SDWPSO algorithm are used to optimize the initial weights and biases of the ELM. **Table 2** shows the performance of the two optimization algorithms on the test set under the conditions of the different number of hidden layers' neurons. Obviously, no matter the number of hidden layer neurons, the determination coefficient R^2 and MSE of the SDWPSO algorithm on the test set are better than that of the PSO algorithm, which fully proves the superiority of the SDWPSO algorithm.

Table 2 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

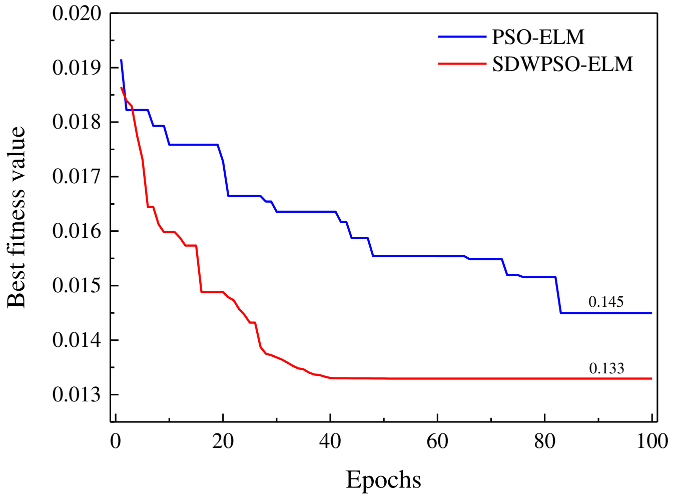

Fig.11. Comparison of optimization processes between PSO and SDWPSO algorithms

When the topology of ELM is 79-80-1, the variation of the best fitness value of the two algorithms is recorded. The variation rule of the best fitness value curves under the two algorithms are shown in **Fig.11**. It can be seen from the diagram 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 about 40 iterations. In addition, the best individual fitness obtained by the standard PSO algorithm is 0.0145, while the SDWPSO algorithm is 0.0133. The result represented by SDWPSO is more likely to be the global extremum of the solution space. After many tests, there is the same rule. Therefore, the SDWPSO algorithm proposed has more advantages because of its characteristic of finding global extremum accurately and quickly.

4.2 Comparison prediction accuracy between different models

In this section, the comprehensive performance of the hybrid PCA-SDWPSO-ELM, single ELM, hybrid PSO-ELM and hybrid SDWPSO-ELM models will be discussed in detail. The basic parameters of the models are as follows: the number of hidden layer neurons in the ELM is 80, the population size of the optimization algorithms is 100, number of iterations is 100, and the parameters of the S-curve *a* and *b* is 0.15 and 10, acceleration factors C_1 and C_2 is 2.4 and 1.6, respectively.

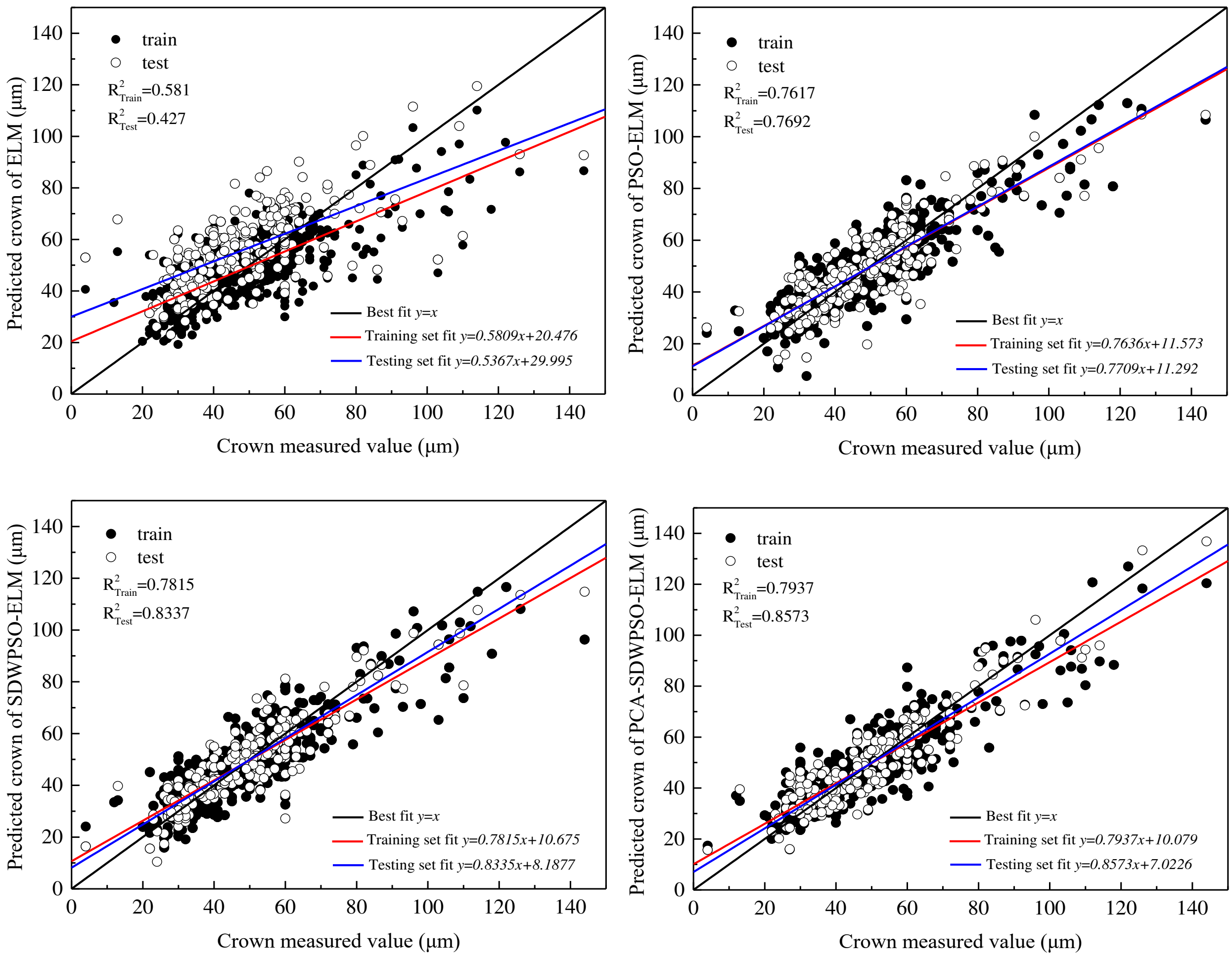

Fig.12. Models regression effect on training set and test set

The scatter plots and regression effect of the four models on the training set and the test set are depicted in **Fig.12**, respectively. 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 are difficult to achieve zero error, so the proximity of regression line and ideal line is one of the important indexes to characterize its performance. In the graph, the red line represents the regression line of the data on the training set used for modeling, while the blue represents the regression line of the data on the test set. The determination coefficient R^2 of models can evaluate 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 is. Clearly, the proposed hybrid PCA-SDWPSO-ELM model has more concentrated scatter distribution and presents higher R^2 values (R^2 values reached 0.7937 and 0.8573 on the training set and the test set, respectively).

Comparison 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. More intuitive and quantitative performances of models are characterized by MAE, MAPE and RMSE. **Table 3** shows the specific calculation results of each error index, and the histogram of error distribution is drawn according to **Table 3**, as shown in **Fig.14.**

In comparison to the hybrid PSO-ELM model and single ELM model which 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%, RMSE=8.428 for training set and R^2=0.769, MAE=6.441, MAPE=17.623%, RMSE=8.667 for test set) is far better than those of the single ELM model (R^2=0.581, MAE=7.972, MAPE=18.458%, RMSE=11.177 for training set and R^2=0.427, MAE=12.189, MAPE=35.434%, RMSE=15.725 for test set). This performance enhancement benefits from PSO algorithm selecting the optimal initial weights and biases for ELM.

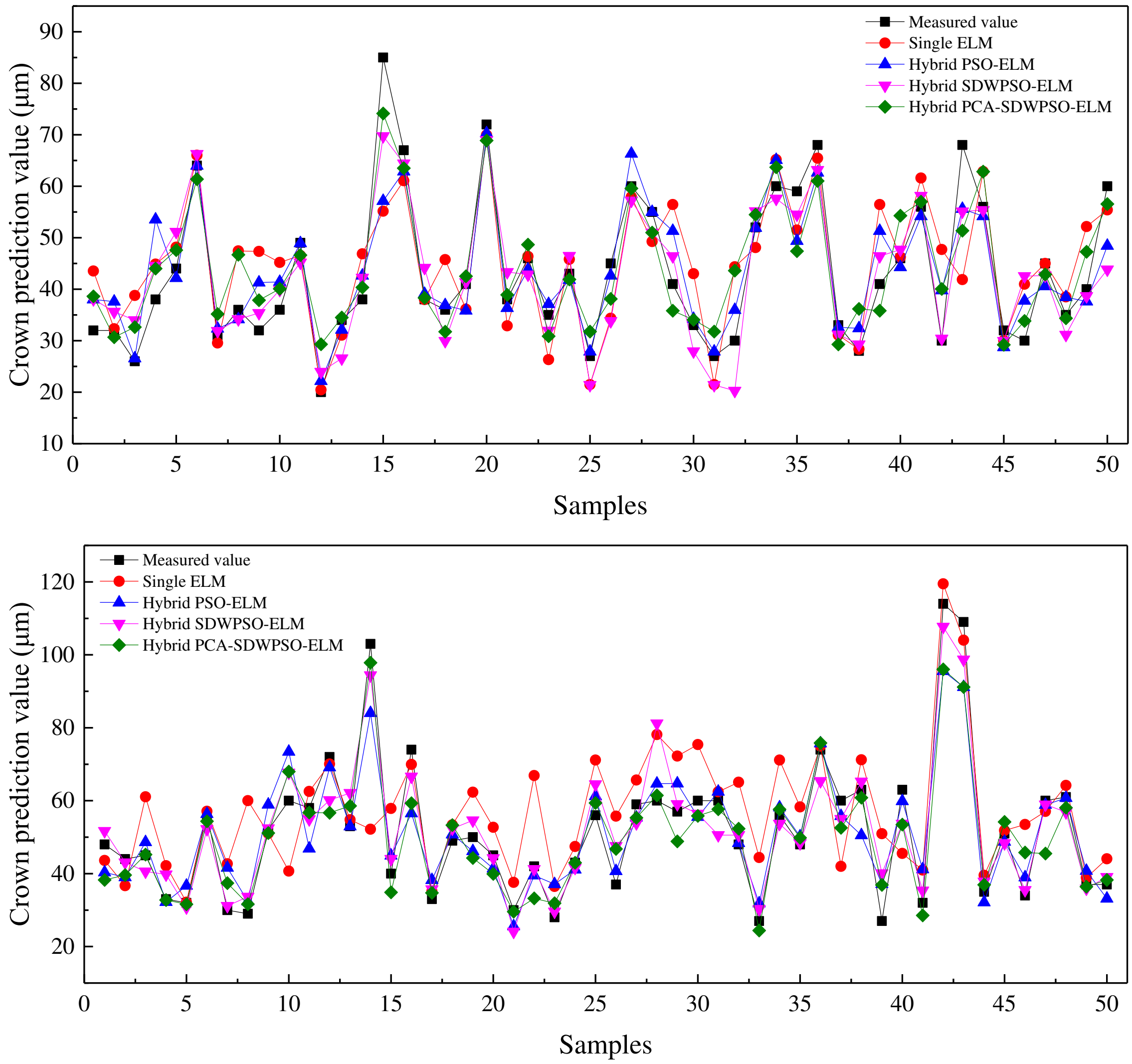

Fig.13. Comparison between the predicted and actual values of the models. (a) training set (b) testing set

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%, RMSE=8.071 for training set and R^2=0.834, MAE=5.358, MAPE=13.613%, RMSE=7.357 for test set) is better than hybrid PSO-ELM model. The inertia weight with S-curve decreases slowly in the initial stage of search, and the larger inertia weight tends to global search in the middle of search. So, the S-curve inertia weight makes the algorithm have better ergodicity in theory. At the end of iteration, because the small inertia weight is more inclined to local search, the S-curve inertia weight reaches the smaller value more quickly and it can obtain better local search performance.

Table 3 Calculation results of 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

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%, RMSE=7.842 for training set and R^2=0.857, MAE=5.165, MAPE=13.429%, RMSE=6.814 for test set) is better than that of the hybrid SDWPSO-ELM model. After PCA processing with a cumulative contribution of 0.98, the modeling independent variables are reduced from 79-dimensional to 14-dimensional. It has the advantages of fast training speed and simple topology structure to use the data set after dimension reduction. **Fig.15** shows the training time for each model. According to **Fig.15**, the training time of the hybrid PCA-SDWPSO-ELM model is less than that of the hybrid SDWPSO-ELM model, which confirms this conclusion. Fast response is of great significance for industrial on-line control. So, the hybrid PCA-SDWPSO-ELM model is more suitable for the crown prediction in the on-line control process of hot strip rolling than other models.

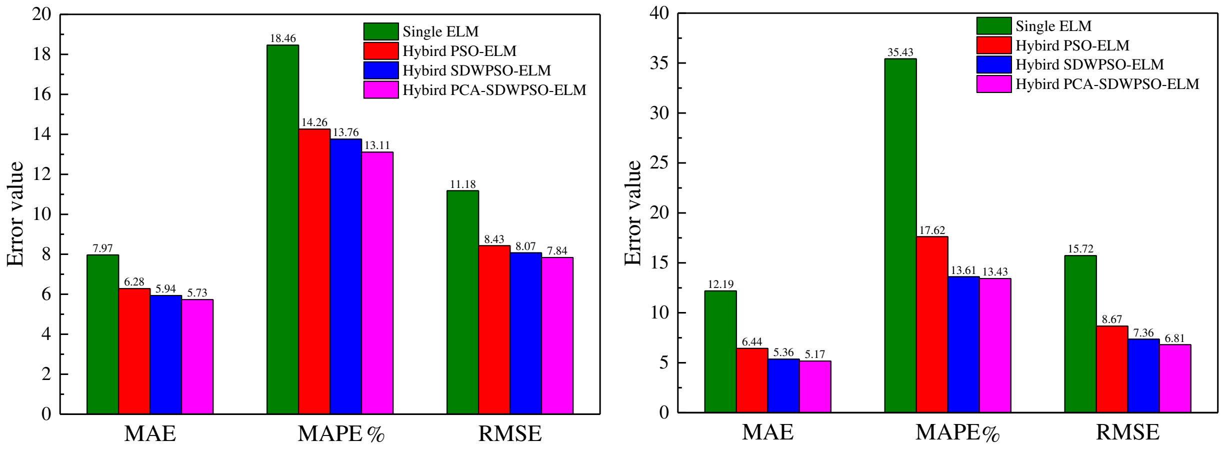

Fig.14. Error distribution histogram of four models. (a) for training set, (b) for test set

Synthesis of all above analysis, 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 hot rolling process. This study can greatly reduce the traditional mathematical calculation without losing the precision. More importantly, the research method proposed in this paper can be extended to other model parameters prediction and optimization process, and it can effectively solve other nonlinear and strong coupling problems in rolling process.

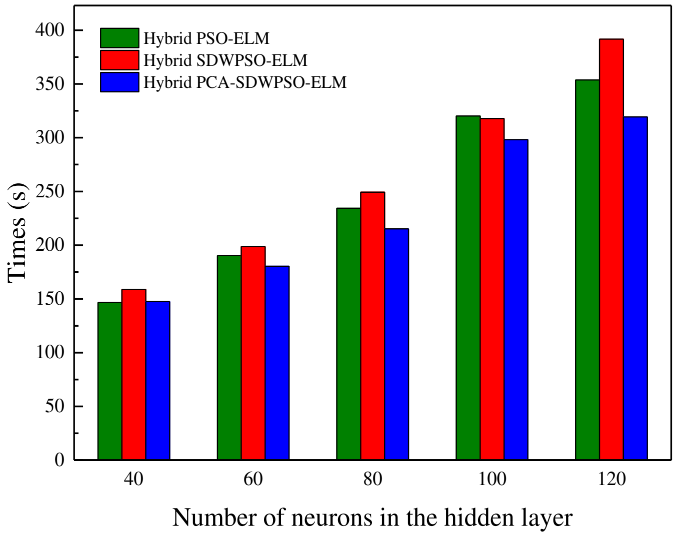

Fig.15. Comparison of training time required for each model

5 Conclusions

An efficient approach for prediction of strip crown in rolling process based on machine learning algorithm, called hybrid PCA-SDWPSO-ELM model is proposed in this research. The purpose of study is to establish a soft measurement method based on production data to realize accurate prediction of the strip exit crown, and then 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 the other three models.

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

Acknowledgments

This study is financially supported by the National Key R&D Program of China (Nos.: 2018YFAO707300), National Natural Science Foundation of China (Nos.: 51904206), 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.

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.

Author contributions

Zhenhua Wang contributed to the conception of the study, performed the experiment; Yuanming Liu contributed significantly to analysis and manuscript preparation; Zhenhua Wang performed the data analyses and wrote the manuscript; Tao Wang helped perform the analysis with constructive discussions; Dianyao Gong contributed to write, review and edit; Dianhua Zhang's contribution was supervision.

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. MHS'95. Proceedings of the Sixth International Symposium on Micro Machine and Human Science. https://doi: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

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

Han F, Yao HF, Ling QH (2013) An improved evolutionary extreme learning machine based on particle swarm optimization. Neurocomputing 116: 87-93

Huang GB, Zhu QY, Siew CK (2006) Extreme learning machine: Theory and applications. Neurocomputing 70(1): 489-501

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, 2004. Proceedings, vol 2. pp 985–990

Huang GB, Zhou H, Ding X, ZhangR (2012) Extreme learning machine for regression and multiclass classification. IEEE Trans Syst Man Cybern B Cybern 42(2):513–529

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 s80–81 (Suppl 5): 16-23

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 reducted 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

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

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 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

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 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

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 F. A. Parisini T. (2013) Adaptive task-space metal strip-flatness control in cold multi-roll mill stands. Journal of Process Control.23: 108-119

Pittner J, Simaan M (2011) A useful control model for tandem hot metal strip rolling. IEEE T 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. Proceedings of the IEEE international conference on evolutionary computation, IEEE Press, Piscataway

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

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

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, 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 Dynam 83(3): 1535-1548

Zhang XL, Zhao L, Zhao WB, Xu T (2015) 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 (2015) Research on flatness intelligent control via GA-PIDNN. J Intell Manuf 26(2): 359-367