metals

MDPI

Article

Prediction Model of Strip Crown in Hot Rolling Process Based on Machine Learning and Industrial Data

Zhenhua Wang [1,2,3,*], Yu Huang [2], Yuanming Liu [1,2,3] and Tao Wang [1,2,3]

[1] College of Mechanical and Vehicle Engineering, Taiyuan University of Technology, Taiyuan 030024, China
[2] National Key Laboratory of Metal Forming Technology and Heavy Equipment, Xi'an 710018, China
[3] Engineering Research Center of Advanced Metal Composites Forming Technology and Equipment, Ministry of Education, Taiyuan 030024, China
* Correspondence: wangzhenhua@tyut.edu.cn; Tel.: +86-0351-6014008

Abstract: The strip crown in hot rolling has the characteristics of multivariablity, strong coupling and, nonlinearity. It is difficult to describe accurately using a traditional mechanism model. In this paper, based on the industrial data of a hot continuous rolling field, the modeling dataset of a strip steel prediction model is constructed through the collection and collation of the on-site data. According to the classical strip crown control theory, the important process parameters that affect the strip crown are determined as input variables for the data-driven model. Some new intelligent strip crown prediction models integrating the shape control mechanism model, artificial intelligence algorithm, and production data are constructed using four machine learning algorithms, including XGBoost, Random Forest (RF), Support Vector Machine (SVM), and Multi-Layer Perceptron (MLP). The overall performance of the models is evaluated using error indicators, such as Mean Absolute Percentage Error (MAPE), Root Mean square Error (RMSE), and Mean Absolute Error (MAE). The research results showed that, for the test set, the determination coefficient (R^2) of the predicted value of the strip crown model based on the XGBoost algorithm reached 0.971, and the three error indexes are at the lowest level, meaning that the overall model has the optimal generalization performance, which can realize the accurate prediction of the outlet strip crown in the hot rolling process. The research results can promote the application of industrial data and machine learning modeling to the actual strip shape control process of hot rolling, and also have important practical value for the intelligent preparation of the whole process of steel.

Keywords: strip crown prediction; industrial data; machine learning; hot strip rolling; XGBoost algorithm

check for updates

Citation: Wang, Z.; Huang, Y.; Liu, Y.; Wang, T. Prediction Model of Strip Crown in Hot Rolling Process Based on Machine Learning and Industrial Data. *Metals* **2023**, *13*, 900. https://doi.org/10.3390/met13050900

Academic Editors: João Manuel R. S. Tavares and Emin Bayraktar

Received: 31 March 2023
Revised: 21 April 2023
Accepted: 4 May 2023
Published: 6 May 2023

Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

The production of hot-rolled strip occupies an important position in the modern iron and steel industry system. Strip shape, mainly including crown and flatness, is a key quality indicator in the production process of hot-rolled strip, and strip shape control is also a challenge in the production process of strip steel [1,2]. Although scholars and enterprises have conducted a large amount of research on this issue, up to now, there are still prominent issues of poor strip shape in the production process of hot-rolled strip. A large number of thin strip coils are rolled with product defects, such as mid wave, side wave, quarter wave, wedge shape, and unqualified crown hit, which seriously affect the yield of hot-rolled strip and bring incalculable economic losses to the enterprise. Poor shape quality of hot-rolled strip not only affects the smooth progress of the rolling process, but also adversely affects the smooth progress of subsequent processes, such as cold rolling, shearing, and other processes [3]. This inevitably requires accurate control of the shape accuracy of hot-rolled strip. The preset model in the shape control model calculates the adjustment amount of each relevant shape adjustment mechanism using a mathematical model based on the relevant parameters, such as the thickness, width, material, reduction,

and equipment parameters of the strip rolling mill, and outputs it to the actuator. Therefore, the shape preset model accuracy is not only related to the yield of each coil of strip steel, but also affects the convergence speed and accuracy of the feedback control module in adjusting the shape to reach the target value. In short, the accuracy of shape presetting directly affects the shape quality and rolling stability of the hot strip rolling.

Since the 1990s, the widespread application of computer control technology has led to the rapid development of strip shape setting and dynamic control models in hot strip rolling and the increasingly mature strip shape detection technology. Hydraulic bending, roll shifting, roll crossing, and roll sectional cooling technologies have become more advanced, and these factors together promote the continuous development of strip shape control systems and mathematical models. Pina et al. [4] proposed a new flatness control method for continuous cold rolling mills. They decomposed the measured data of strip elongation recorded by traditional flatness meters into orthogonal components perpendicular to the direction of the main actuator, thereby reducing the dimension of the problem, and calculated control variables by solving a quadratic optimization problem with priority constraints. The effectiveness and ease of configuration of this method have been fully demonstrated in the application process of many actuators. Moazeni et al. [5] established a three-dimensional model of a two-high cold rolling mill and analyzed the material flow and stress changes of non-uniform thickness slabs during steady-state broadband cold rolling using the finite element method. Tran et al. [6] proposed an experimental method for the interaction between residual stress and buckling when there is a flatness deviation at the edge of strips and verified it using finite element experiments. Linghu et al. [7] used three-dimensional elastoplastic finite element simulation of the cold rolling process of a six-high Continuously Variable Crown (CVC) mill to study the effect of a shape actuator on the strip crown and edge drop. Wang et al. [8] obtained an accurate flatness regulation efficiency coefficient matrix by establishing a self-learning model of flatness regulation efficiency coefficient, and then applied it to practical production. The above research mainly uses the mathematical analysis method and finite element method to study the strip shape problem in rolling processes, and has achieved good results, promoting the improvement of the accuracy of strip shape control. Due to the limitations of a large number of simplified conditions in the solving process of mathematical analytical methods and finite element methods, it becomes increasingly difficult to improve the accuracy of traditional mathematical models of strip shape in rolling process.

In order to continuously improve the accuracy of prediction models, artificial intelligence technology with data and algorithms as the core has attracted more and more attention from scholars. Machine learning algorithms are widely used in new material development, alloy properties prediction, process parameter optimization, and other fields. Yan et al. [9] used linear regression analysis to establish several predictive models to study the relationship between the microstructure and mechanical properties of Al7Si1.5Cu0.4Mg alloy and proposed a new quantitative relationship between tensile and fatigue properties. Li et al. [10] proposed a high-precision convolutional network prediction model for the mechanical properties of hot-rolled strip steel based on historical data of rolling production lines, and then conducted experimental verification. Opěla et al. [11] compared the shallow learning and deep learning algorithms of neural networks for predicting the hot flow stress of micro-alloyed medium carbon steel, and the results showed that the deep learning model based on RBM can achieve more reliable network performance under certain conditions. Churyumov et al. [12] established a true stress at hot deformation prediction model of high manganese steel using artificial neural networks (ANNs), the research results indicate that the established model can achieve high accuracy in prediction errors between 5.4% and 6.3%. Churyumov et al. [13] also established a flow stress model for steel using an ANN. The ANN-based model showed high accuracy for both training (the error was 6.6%) and approvement (11.5%) datasets. Li et al. [14] used neural network models and regression models to predict the mechanical properties of cast aluminum alloy ZL702A. By comparing

the two prediction models, it was found that the neural network model was more effective and accurate based on microstructure and casting temperature parameters.

In addition, machine learning also has a wide range of applications in the prediction of key parameters in the rolling process and process optimization. Nandan et al. [15] proposed a multi-objective optimal control strategy based on genetic algorithms and applied it to the identification of hot-rolled strip shape parameters and the setting of rolling schedules, which successfully obtained the optimal setting values for the crown and flatness, and also improved the control accuracy of strip shape. Chakraborti et al. [16] used genetic algorithms and ant colony algorithms to optimize the crown of hot-rolled strip, demonstrating the practicality of evolutionary algorithms in optimizing rolling process parameters. John et al. [17] directly used a combination of neural networks and genetic algorithms to establish a relationship model between input parameters and strip shape to predict the minimum strip shape value of hot-rolled strip. Liu et al. [18] established a transfer matrix between the flatness error characteristic parameter and the flatness adjustment parameter using a genetic algorithm to optimize the BP neural network method, and successfully applied the transfer matrix to the flatness adjustment mechanism of a 900 mm six-high HC rolling mill. Peng et al. [19] proposed a new method for recognizing the shape pattern of cold-rolled strip and verified it on an 8000 kN HC mill. The results show that the method can effectively reduce the shape deviation of the strip. Yang et al. [20] established an intelligent collaborative control model for cold-rolled strip flatness control mechanisms based on the combination of a flatness control matrix and differential algorithm optimization extreme learning machine (ELM). Zhang et al. have, respectively, used support vector machines (SVMs) [21], T-S cloud inference neural networks [22], PID neural networks [23], and radial basis networks [24] to study the problem of strip shape pattern recognition in the cold rolling process. Simulation results under various methods show that the various models proposed can identify common defects in strip shape with high accuracy. Deng et al. [25,26] constructed a hybrid model based on the combination of hot strip rolling production data and deep learning networks to predict strip outlet crown, achieving an absolute error of less than 5 μm for 97.04% of the predicted data in the modeling data. Song et al. [27] used machine learning algorithms to establish an accurate prediction model for the strip crown of hot rolling. The model prediction results showed that 97.83% of the data had a difference of less than 4 μm between the actual value and the predicted value. Through the above analysis, it can be found that combining big data technology with artificial intelligence modeling methods is a new trend in studying how to further improve the accuracy of shape control in rolling process [28,29]. Based on the massive industrial data accumulated in the process of hot strip rolling, this paper establishes an accurate prediction model for strip crown through in-depth mining of industrial data, combined with advanced machine learning algorithms, and then constructs a new intelligent shape control preset model that deeply integrates the crown control mechanism model, artificial intelligence algorithms, and production data. As such, it solves the problem that it is difficult to further improve the shape control accuracy when relying on traditional mechanism models, achieves the goal of effectively improving the flatness control accuracy, and obtains high-quality strip steel with good flatness.

The paper is organized as follows. The definition of the strip crown in hot rolling and the influence factors of the strip crown are briefly introduced in Section 2. In Section 3, the process of collecting and processing modeling datasets and the selection of model input variables are described in detail. Section 4 briefly introduces the four machine learning algorithms used for modeling and determines the main parameters of each model. The discussion of the strip crown forecasting results is described explicitly in Section 5, and Section 6 concludes this paper.

2. Theory of Strip Crown Control

2.1. Definition of Strip Crown

Strip shape intuitively refers to the degree of warpage of the strip, and essentially refers to the distribution of residual stress inside the strip. The measurement of strip shape usually includes both longitudinal and transverse indicators. The longitudinal direction is represented by flatness, which refers to the flat degree of the strip along the length direction. The transverse direction is expressed by crown, referring to the cross-sectional shape of the strip steel. The standard definition of strip crown is the difference between the center thickness of the strip and the specified edge thickness. To eliminate the effect of strip edge thinning, the edge reference point is usually located 40 mm from the strip edge. The schematic diagram of crown definition is shown in Figure 1. Equation (1) is as follows:

$$C_{40} = \frac{(h_c - h_i) + (h_c - h'_i)}{2} \quad (1)$$

where h_c is the thickness at the center of the strip, h_i is the thickness at the distance operation side at 40 mm, and h'_i is the thickness at the distance drive side at 40 mm.

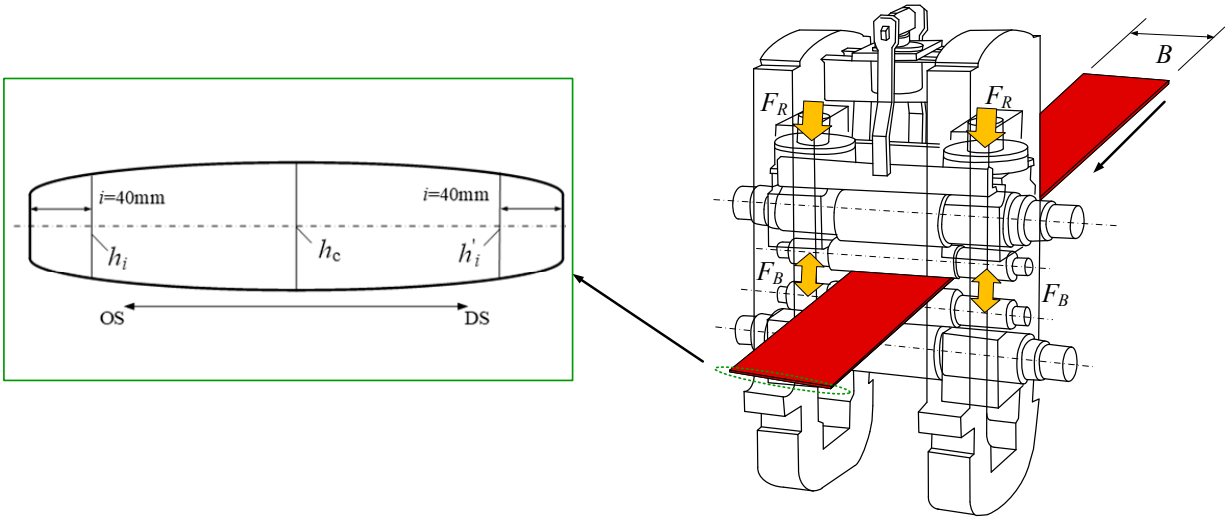


Figure 1. Thickness variation in the strip cross-section.

2.2. Influence Factors of Strip Crown

The crown of the strip can be seen as an image of the cross-sectional shape of the roll gap at the exit of the deformation zone. Therefore, all factors that can affect the cross-sectional shape of the roll gap at the exit of the deformation zone are factors that affect the strip crown. The strip width, rolling force, work roll diameter, backup roll diameter, crown of work roll, entrance strip crown, and roll bending force are listed as the main factors affecting the strip crown [30,31]. The interrelationship of various influencing factors is shown in Figure 2. The impact effects of these factors are different, and under certain conditions, a certain influencing factor may have a relatively significant effect. As can be seen from Figure 2, there is a non-linear and strong coupling characteristic between the influencing factors affecting the strip outlet crown.

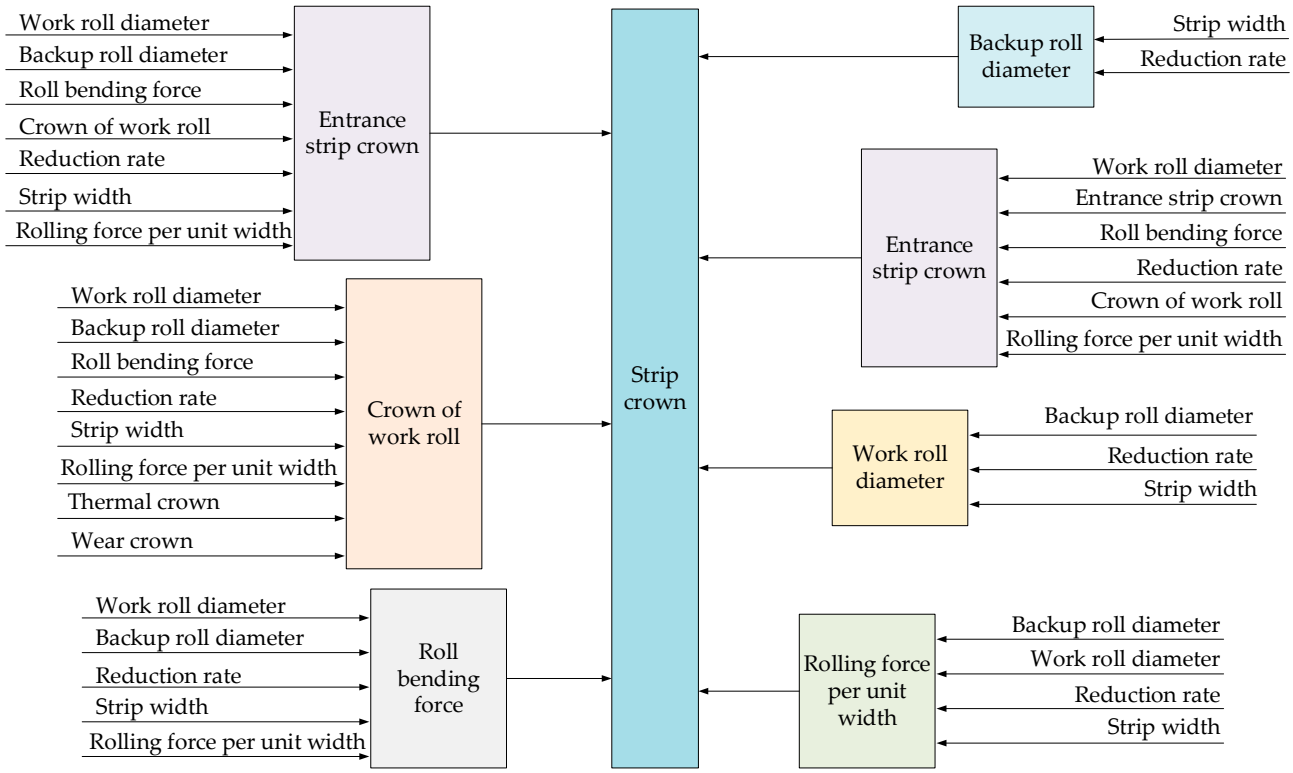


Figure 2. Influencing factors of center crown of hot-rolled strip.

3. Collection and Processing of Modeling Data

3.1. Data Collection

The process layout of the hot continuous rolling production line is shown in Figure 3. Generally, the hot continuous rolling production line consists of a heating furnace, roughing mill, finishing mill, laminar cooling device, and coiling. The roughing mill unit includes a high-pressure water descaler, a vertical mill (E1), and a roughing mill (R1). The vertical mill (E1) controls the width of the strip, and the roughing mill controls the thickness of the strip when it enters the finishing mill unit. The semi-finished products from the rough rolling area are sent to the finishing mill area through the roller table. The finishing mill unit includes a high-pressure water descaler, a flying shear, and a seven-stand finishing mill (F1~F7). After the strip comes out of the finishing mill, it enters the laminar cooling device, which controls the cooling speed and final coiling temperature of the strip. The strip is rolled into coils by a coiler.

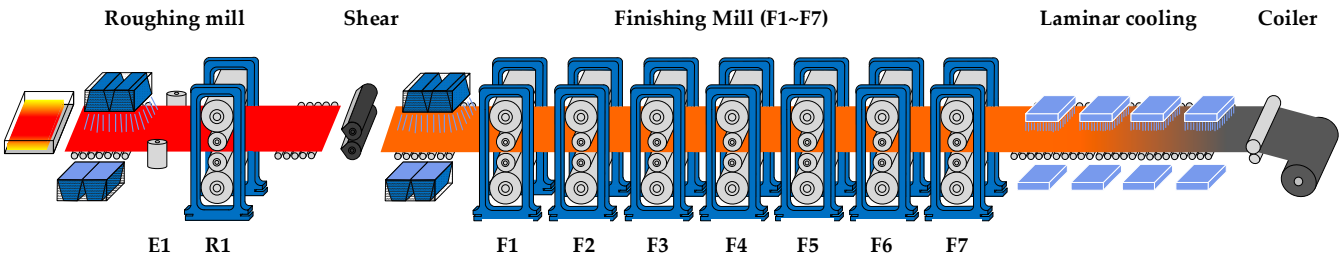


Figure 3. Schematic layout of the hot strip rolling mill.

During the actual production process, the automatic control system monitors and records the relevant data of each equipment and strip steel in real time. The data recorded in the database is divided into original data, calculated value data, and actual value data. The original data includes strip steel grade, material, slab size, and finished product size.

These data are necessary parameters for setting calculation and automatic control. The calculated value data includes various process parameters and model calculation data during the production process, which are important for the accuracy and quality of the finished product and for the analysis of the process. The actual value data includes the measured data of various pieces of testing equipment during the production process, which is an important basis for adaptive learning and the dynamic correction of the rolling mathematical model. The data collection process is shown in Figure 4.

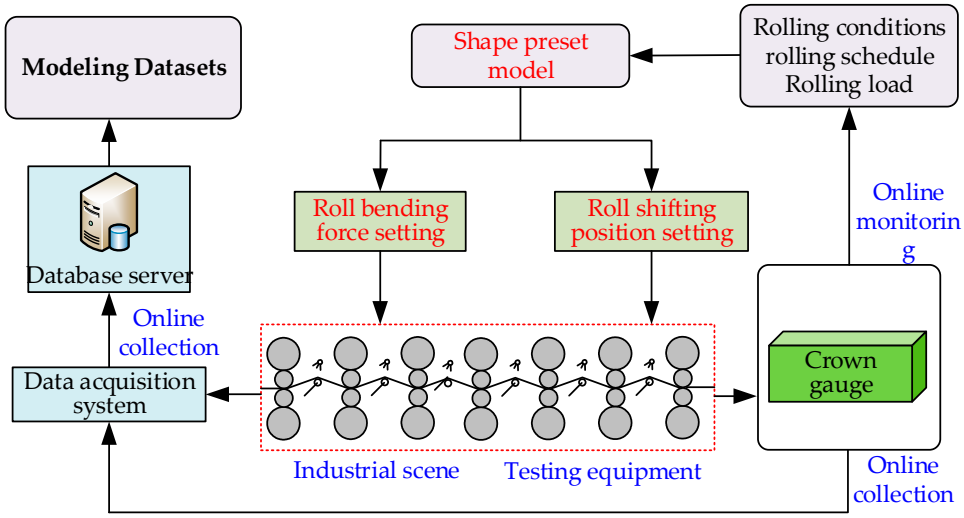


Figure 4. Field data collection process.

3.2. Data Preprocessing

The data were collected on a real 1780 mm hot strip mill process production line of HBIS Group Co., Ltd., which is located in Hebei Province, China. Pre-processing of collected data included deleting sample data with missing values and eliminating outlier data. After data processing, 1809 pieces of strip steel sample data were finally obtained to form a modeling dataset. Part of the modeling data is shown in Table 1. The modeling dataset is divided into the training set and test set based on the sampling balance. The training set proportion was 70%, and the test set proportion was 30%. In order to eliminate the impact of large differences in magnitude in different dimensions of sample data on model accuracy during modeling, the data of all input variables were standardized [32,33]. The formula for standardization processing is as follows:

$$x_i' = 2 \times \frac{x_i - \min(x_i)}{\max(x_i) - \min(x_i)} + (-1), i = 1, 2, 3, \ldots, n \quad (2)$$

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

3.3. Determination of the Input and Output Parameters of the Models

Based on the analysis of the influencing factors of strip outlet crown in the theory of strip shape control, the input variables of the machine learning model were finally determined according to the plate shape control theory. These input variables included strip width (W), slab thickness (H), exit thickness (H1~H7), entrance temperature (T1~T7), exit temperature (t1~t7), rolling force (F1~F7), rolling speed (V1~V7), strip yield strength (Q1~Q7), bending force (W1~W7), rolling shifting (S1~S7), roll diameter (D1~D7), roll thermal expansion (C1~C7), and roll wear (M1~M7). The specific input variables are shown in Figure 5. The strip crown is used as the model output variable.

Table 1. Part of the experimental data.

No	W/mm	H/mm	H_1/mm	T_1/°C	t_1/°C	F_1/MN	V_1/m/s	Q_1/MPa	W_1/MN	S_1/mm	D_1/m	C_1/mm	W_1/mm	C/μm
1	1260.1	34.54	21.98	944.8	947.2	27.49	1.21	194.03	1.42	89	0.8458	0.0845	0	40
2	1279.2	34.53	23.13	943.3	944.3	23.59	1.40	175.70	2.30	100	0.8458	0.1065	−0.0005	105
3	1276.5	34.55	23.04	969.8	972.0	25.02	1.45	187.36	2.30	100	0.8458	0.1270	−0.0010	53
4	1277.9	34.54	22.41	970.6	973.4	26.03	1.40	187.66	1.52	100	0.8458	0.1448	−0.0014	40
5	1276.7	34.56	22.43	976.7	978.1	25.87	1.31	186.68	2.30	100	0.8458	0.1578	−0.0019	67
6	1280.8	34.56	22.44	972.5	973.6	25.84	1.30	184.90	2.30	100	0.8458	0.1700	−0.0024	44
7	1280.5	34.56	22.45	983.4	983.3	24.66	1.29	176.42	2.28	90	0.8458	0.1750	−0.0028	44
8	1280.3	34.56	22.47	978.4	978.2	24.37	1.26	174.55	2.28	90	0.8458	0.1878	−0.0033	39
…	…	…	…	…	…	…	…	…	…	…	…	…	…	…
1807	1232.7	32.56	14.58	1007.8	1005.7	28.44	0.89	149.16	1.00	30	0.8457	0.3419	−0.0365	39
1808	1234.6	32.56	14.34	1006.6	1005.8	29.03	0.91	150.03	1.00	10	0.8457	0.3527	−0.0374	38
1809	1234.6	32.56	14.37	1006.2	1005.1	28.89	0.91	149.62	0.50	10	0.8457	0.3576	−0.0382	40

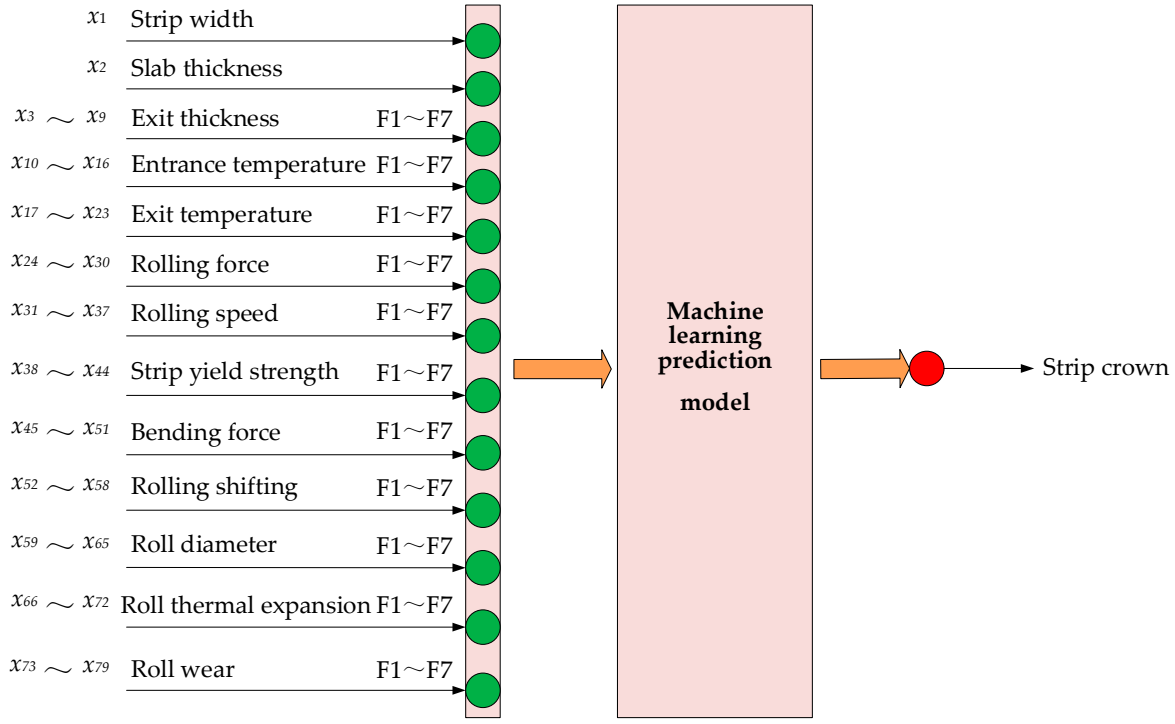


Figure 5. Input and output variables of the machine learning prediction model.

4. Experimental

ANNs are one of the most classic machine learning algorithms, widely used in various fields due to their strong nonlinear fitting ability and their ability to map any nonlinear relationship. Compared with ANNs, SVMs have a more solid mathematical theoretical foundation and can effectively solve the problem of constructing high-dimensional data models under limited sample conditions. Separate ANNs and SVMs are typical individual learners. The method of constructing and combining multiple individual learners to complete learning tasks is called ensemble learning. Random forest (RF) and XGBoost are two different ensemble learning strategies. RF is an ensemble learning method based on the idea of bagging, which obtains multiple parallel individual learners through effective resampling, and then calculates the model mean as the final model result. XGBoost is an ensemble learning method based on Boosting's idea, which continuously adjusts the weights of individual learners according to the error rate during the training process, so that individual learners with low error probability can obtain larger weights. The above four models are the most representative methods of machine learning. Therefore, these four methods are used to establish a strip crown prediction model based on the dataset in this article and to conduct comparative research to obtain the best strip crown prediction model.

4.1. MLP-Based Method

The multilayer perceptron model (MLP) is an ANN. An important feature of the MLP is that it has multiple neural layers. The MLP model consists of an input layer, hidden layer, and an output layer, with each layer consisting of multiple neurons [34,35]. According to the modeling dataset, the number of neurons in the input layer of the constructed MLP crown prediction model is equal to the number of input feature variables in the dataset. Therefore, the number of neurons in the input layer is 79, and the number of neurons in the output layer is equal to the number of output variables in the dataset, which means that the number of neurons in the output layer is 1. After testing the hidden layer's activation function of 'identity', 'logistic', 'tanh', and 'relu', it is found that the 'logistic' function has the best prediction accuracy. In addition, the number of neurons in the hidden layer [36] and the regularization term coefficients are important parameters that affect the accuracy of the MLP model. The MLP model adopts 'Python sklearn.neutral_network. MLPRegressor' generation, using 'GridSearchCV' to optimize the number of hidden layer neurons and regularization term coefficients (alpha). The optimal range of the number of neurons in the hidden layer is [80, 100, 200], and the optimal range of the 'alpha' is [0.01, 0.1, 0.5]. Under the dataset in this article, after optimizing the model parameters, the final parameters of the MLP regression model are shown in Table 2.

Table 2. The best configuration for the MLP model.

Parameters	Values
Number of hidden layer units	100
max_iter	5000
alpha	0.01

4.2. SVM-Based Method

The SVM is a new type of machine learning algorithm based on statistical learning theory, which is an approximate implementation of structural risk minimization methods [37,38]. It is suitable for small sample data. Through some nonlinear mapping, the input variables are dimensioned up to a high dimensional space. An SVM has the advantages of fast solution speed and strong generalization ability, and it is widely used in many fields. An SVM can ultimately be transformed into dual optimization forms, as follows:

$$\min \frac{1}{2}\sum_{i,j}^{n}(\alpha_i^* - \alpha_i)\left(\alpha_j^* - \alpha_j\right)K(x_i, x_j) + \varepsilon\sum_{i=1}^{n}(\alpha_i^* - \alpha_i) - \sum_{i=1}^{n} y_i(\alpha_i^* - \alpha_i) \tag{3}$$

The constraints are as follows:

$$\begin{cases} \sum\limits_{i=1}^{n} y_i(\alpha_i^* - \alpha_i) = 0 \\ \alpha_i, \alpha_i^* \geq 0, i = 1, 2, \cdots n \end{cases} \tag{4}$$

where n is the number of samples; α_i α_i^* are the Lagrange multipliers; $K(x_i, x_j)$ is the kernel function, and the commonly used kernel functions include linear kernel, polynomial kernel, and radial basis functions (RBF). Due to the strong nonlinear mapping ability of RBF, this paper uses polynomial kernel as the kernel function. Polynomial kernel is expressed as follows:

$$K(x_i, x_j + 1)^a \tag{5}$$

where a is kernel function parameter.

The number of polynomial kernel functions 'degree', penalty factor 'C', and the coefficient of the kernel function 'gamma' of the polynomial kernel function in SVM models have important impacts on the prediction accuracy of the model. The SVM model adopts 'Python sklearn.svm import SVR' generation, using 'GridSearchCV' to optimize the 'degree',

'C', and 'gamma'. The optimal range of the 'degree' is [1, 2, 3], the 'C' is [20, 30, 40], and the 'gamma' is [0.1, 0.2, 0.23]. After testing the kernel function type of 'linear', 'poly', 'rbf', and 'sigmoid', it is found that the 'poly' function has the best prediction accuracy. Under the dataset in this article, the parameters of the SVM finally determined after parameter optimization are shown in Table 3.

Table 3. The parameters for the SVM model.

Parameters	Values
Degree	3
penalty factor C	40
Gamma	0.23

4.3. RF-Based Method

RF is a parallel integrated learning algorithm based on decision trees. Compared to individual base learners, it achieves randomness in sample selection and feature selection through bagging and random subspace ideas [39,40], enhancing the generalization ability of models. According to the idea of bagging, the RF model obtains m sets of sampled training sets with the same capacity as the original training set through independent random sampling with m times of return, and then uses these sampled training sets to train in order to obtain corresponding m base learners. Due to the independence of sampling, the sampled training set obtained from each sampling is different from the original training set and other sampled training sets, which can effectively avoid the occurrence of local optimal solutions from the perspective of training sample selection and can also ensure a low correlation between each decision subtree.

The number of base learners 'n_estimators', the depth of decision tree 'max_depth', and the number of features to consider when limiting branching 'max_features' in the RF model have important impacts on the prediction accuracy of the model. The RF model adopts 'Python sklearn.ensemble import RandomForestRegressor' generation, using 'GridSearchCV' to optimize the 'n_estimators', 'max_depth', and 'max_features'. The optimal range of the 'n_estimators' is [60, 80, 100], the 'max_depth' is [10, 30, 50], and the 'max_features' is [6, 9, 12]. Under the dataset in this article, the parameters of the RF finally determined after parameter optimization are shown in Table 4.

Table 4. The parameters for the RF model.

Parameters	Values
n_estimators	100
max_depth	50
max_features	6

4.4. XGBoost-Based Method

XGBoost is a flexible, efficient, and convenient optimal distributed algorithm proposed by Chen et al. [41–43], based on the Gradient Boosting Decision Tree (GBDT). XGBoost adopts the idea of boosting ensemble learning to combine multiple decision trees to achieve better results and make the combined model more generalized. XGBoost is composed of multiple decision trees. Each decision tree learns the residual between the target value and the predicted value, where the predicted value is the sum of the predicted values of all previous decision trees. After the training of all decision trees is completed, a common decision is made. After obtaining the corresponding prediction values on each tree, the samples are accumulated as their final prediction results. During the training phase, each new tree is trained on the basis of the completed tree. Each decision tree is a weak learner. Boosting technology is used to upgrade all weak learners into strong learners. To avoid model overfitting and enhance generalization capabilities, XGBoost adds regularization

terms to the loss function of the GBDT model. The traditional GBDT calculates the loss function using a first order Taylor expansion, using a negative gradient value to replace the residual for fitting. XGBoost adds a second order Taylor expansion to the loss function, using second order derivatives to collect gradient direction information, thereby improving the accuracy of the model.

The loss function of the XGBoost algorithm is defined as follows:

$$L^{(t)} = \sum_{i=1}^{n} l\left(y_i, \hat{y}_i^{(t-1)} + f_t(X_i)\right) + \Omega(f_t) \tag{6}$$

$$\Omega(f_t) = \gamma T + \frac{1}{2}\lambda \sum_{j=1}^{T} \omega_j^2 \tag{7}$$

where $l\left(y_i, \hat{y}_i^{(t-1)} + f_t(X_i)\right)$ is the prediction residual of the ith sample to the tth iteration; X_i is the ith sample; $\Omega(f_t)$ is the regularization term; ω_j^2 is the score of leaf node; T is the number of leaf nodes; γ is the coefficient; λ is the coefficient of the sum of the weights regularized by the L2 regularization term of all leaf nodes.

The second-order Taylor expansion of the loss function of XGBoost algorithm is as follows:

$$L^{(t)} = \sum_{i=1}^{n} \left[l\left(y_i, \hat{y}_i^{(t-1)} + g_i f_i(X_i)\right) + \frac{1}{2} h_i f_i^2(X_i) \right] + \Omega(f_t) \tag{8}$$

where g_i is the first derivative term of the loss function; h_i is the second derivative term of the loss function.

The number of base learners 'n_estimators', the depth of decision tree 'max_depth', and the 'learning_rate' in XGBoost model have important impacts on the prediction accuracy of the model. The XGBoost model adopts 'Python XGBoost import XGBRegressor' generation, using 'GridSearchCV' to optimize the 'n_estimators', 'max_depth', and 'learning_rate'. The optimal range of the 'n_estimators' is [60, 80, 100], the 'max_depth' is [10, 30, 50], and the 'learning_rate' is [0.1, 0.2, 0.3]. Under the dataset in this article, the parameters of the XGBoost finally determined after parameter optimization are shown in Table 5.

Table 5. The parameters for XGBoost model.

Parameters	Parameters
n_estimators	100
max_depth	50
learning_rate	0.1

5. Results and Discussion

The experiments are carried out on a computer running 64-bit Windows 10 with a 3.0 GHz processor and 16 GB memory. The algorithms are implemented in the PYTHON language. This research adopted R^2, MAE, MAPE, and RMSE as criteria for assessing the prediction performance of machine learning models. The performance criteria calculation formulas are as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^{n} \left(y_i - y_i^*\right)^2}{\sum_{i=1}^{n} \left(y_i - \overline{y}\right)^2}, \left(\overline{y} = \frac{1}{n}\sum_{i=1}^{n} y_i\right) \tag{9}$$

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n} |y_i - y_i^*| \tag{10}$$

$$\mathrm{MAPE} = \frac{1}{n}\sum_{i=1}^{n} \left| \frac{y_i - y_i^*}{y_i} \right| \times 100\% \tag{11}$$

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{n} \left(y_i - y_i^*\right)^2}{n}} \tag{12}$$

where n denotes the number of sample data; y_i and y_i^* are the measured value and the predicted value of the ith sample, respectively.

The modeling dataset is divided into a training set and a test set. Four machine learning algorithms, namely MLP, SVM, RF, and XGBoost, are used to train the strip crown prediction model for hot strip rolling based on the same training set data, and the performance of the established models are tested using the same test set. The regression effects of the four prediction models are shown in Figure 6. According to Figure 6, under the same dataset, the regression prediction effect of the MLP model is the worst, with an R^2 value of 0.904 on the training set and 0.860 on the test set. Whether in the training set or the test set, there are a large number of samples distributed on both sides of the y = x line. In sharp contrast to the MLP model, the coefficient of determination of the predicted value of the XGBoost model reached 0.999 on the training set. All samples of the training set are distributed on a straight line, while in the test set, the R^2 value also reaches 0.971, and only a few scattered individual samples appear outside the straight line. These results demonstrate the excellent generalization performance of the XGBoost model. In addition, the overall regression effect of the SVM model and the RF model are between the MLP and XGBoost models, and the regression effect of the RF model is superior to the SVM model both in the training set and the test set. The R^2 values of the two models in the training set are 0.956 and 0.990, respectively, and the R^2 values in the test set are 0.909 and 0.945, respectively.

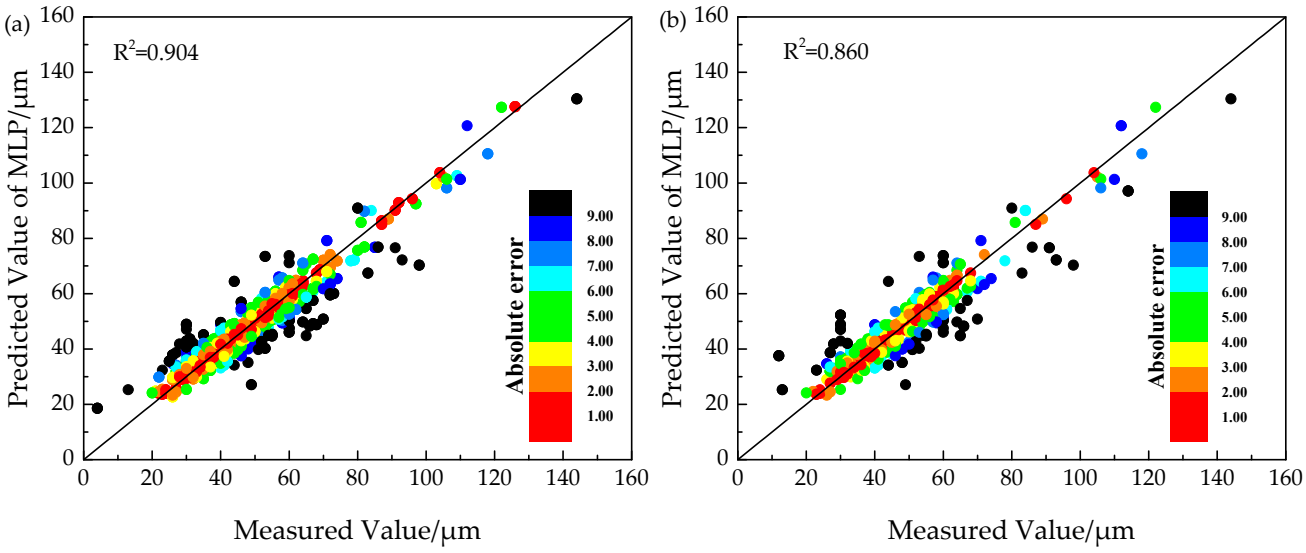


Figure 6. *Cont.*

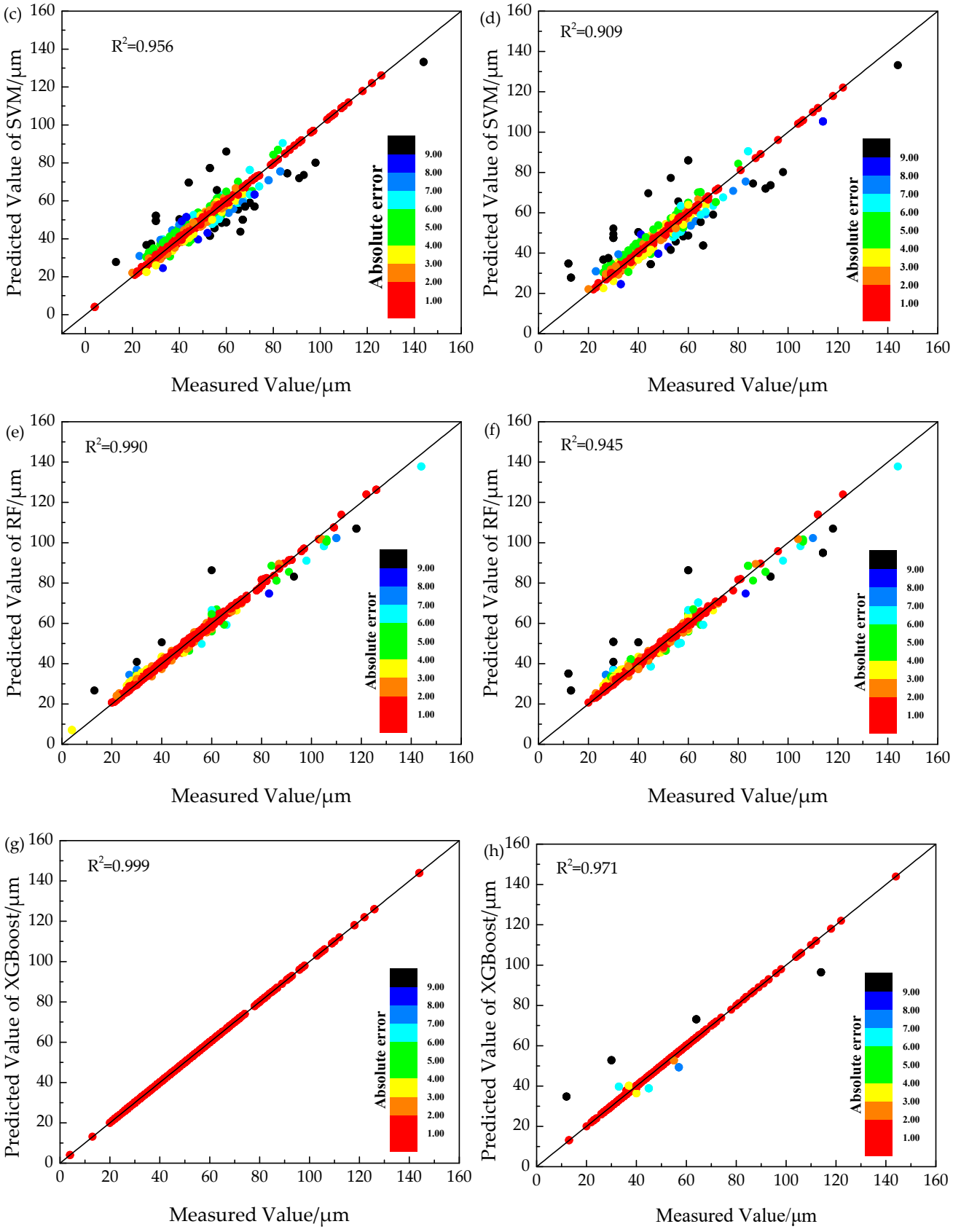


Figure 6. Models' regression effect on the training set and test set. (**a**,**c**,**e**,**g**) For training set and (**b**,**d**,**f**,**h**) for test set.

In order to visually display the comparison between the predicted values and the actual values of the four crown prediction models, Figure 7 is drawn to express the proximity of the predicted values to the actual values. Due to the large amount of modeling sample data, it is not possible to display the situation of all samples in one graph. Therefore, the first 100 samples from the training set and the test set are taken for plotting. As can be seen from Figure 7a, in the training set, the predicted values of each model have a greater degree of consistency with the actual values, because all the samples in the training set participated in the training of the model during the modeling process. On the contrary, there are many samples in the test set that have significant errors with the actual values, especially in the prediction results of the MLP and SVM models, where a large number of samples seriously deviate from the actual values of the corresponding crown. There are two reasons for this result. Firstly, compared to the training set, the test set samples did not participate in the model construction, resulting in a lower accuracy for its predicted values than the training set sample. In addition, because the generalization performance of MLP and SVM models is inferior to that of XGBoost and RF models, the prediction effect shown on the test set is that there are more samples with prediction results that seriously deviate from the actual values.

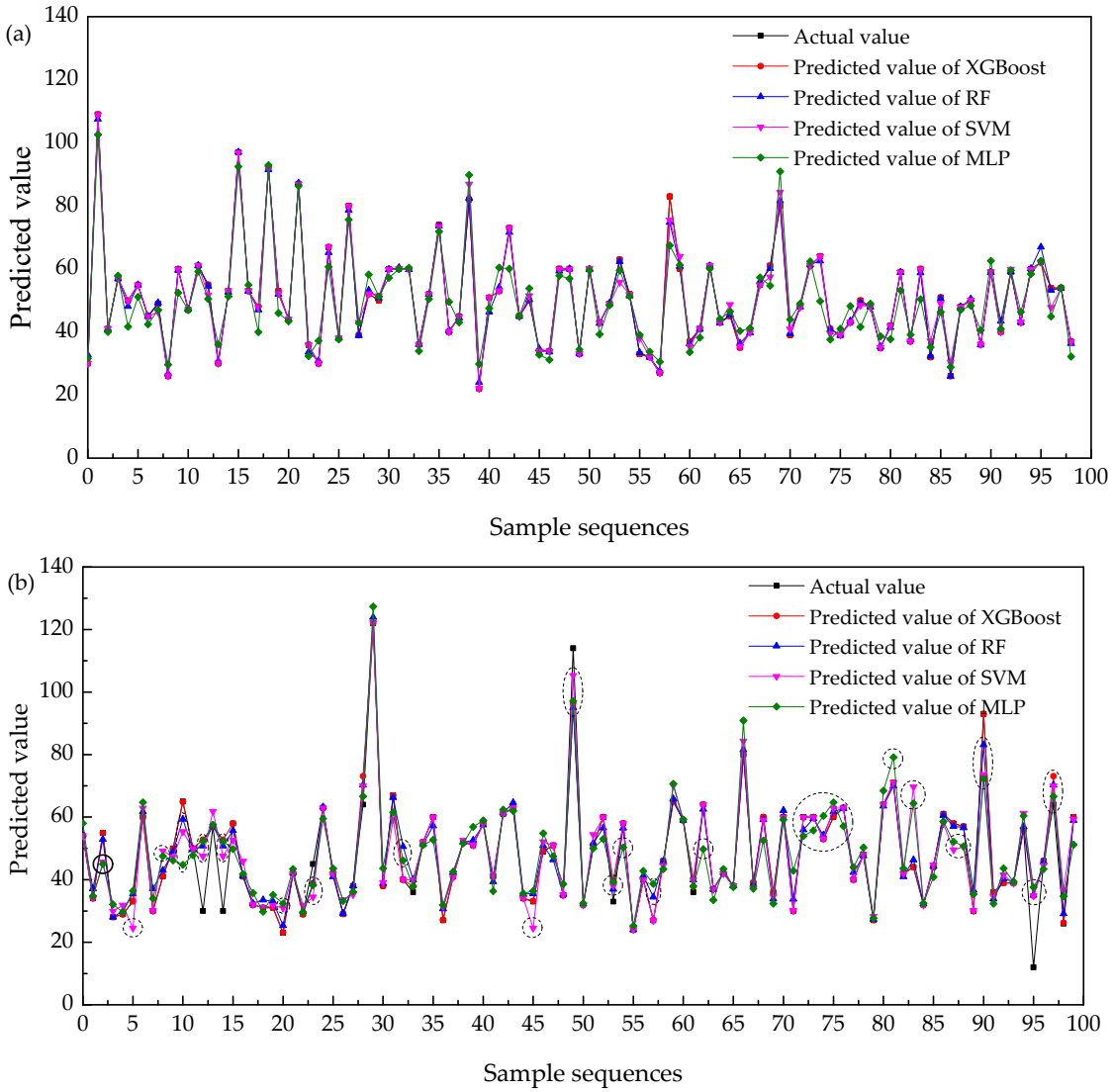


Figure 7. Comparison between the predicted and actual values of the models. (**a**) For training set and (**b**) for test set.

From the perspective of quantitative analysis, the absolute error between the predicted value and the actual value of each model in the training set and test set is statistically analyzed, and the statistical results are based on the reference standard that the absolute error of the predicted crown value is less than 4 μm. The results are shown in Table 6. From the table, it can be seen that the absolute error of XGBoost model's prediction results on the training set is all less than 4 μm. The proportion of samples with an absolute error of less than 4 μm on the test set reached 96.13%, with only a small percentage having an absolute error of greater than 4 μm. The sample proportions of other models within this error range are smaller than that of XGBoost model, and the order of the ratio from large to small is RF, SVM, and MLP. The absolute error frequency distribution histograms and corresponding normal distribution curves of the four models' predictions are shown in Figure 8. It can be seen from Figure 8 that compared with the other three prediction models, the prediction error of XGBoost model is more concentrated around the error of zero. With the increase in absolute error, the frequency gradually decreases and presents a normal distribution. It is fully proved that XGBoost prediction model has a good prediction effect.

Table 6. The absolute error of the predicted values for each model with less than 4 μm statistical results.

Models	Statistical Items	Training Set	Test Set
XGBoost	Number of samples	1266	522
	Proportion	100%	96.13%
RF	Number of samples	1228	474
	Proportion	97.00%	87.29%
SVM	Number of samples	1104	420
	Proportion	87.20%	77.35%
MLP	Number of samples	838	318
	Proportion	66.19%	60.92%

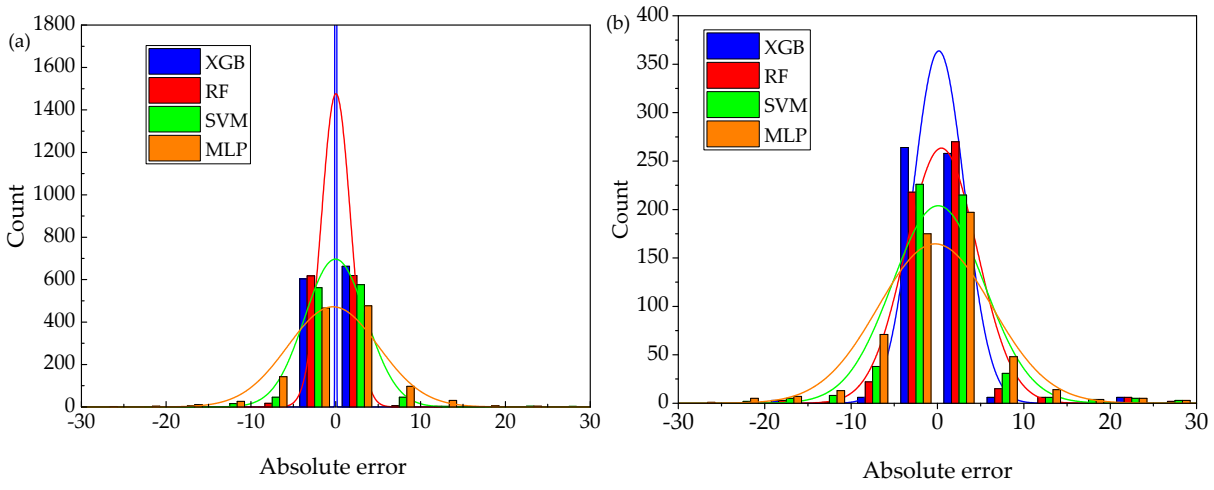


Figure 8. Distributions of the absolute error of the four models. (**a**) For training set and (**b**) for test set.

In order to evaluate the generalization performance more comprehensively and quantitatively, three errors, RMSE, MAE, and MAPE, are used as error indicators to analyze the above four models. Table 7 lists the calculated values of the three error indicators of each model. Figure 9 shows the error distribution histogram drawn according to the calculation results. When comparing the MLP model, RF model, SVM model, and XGBoost model, the result shows that the prediction accuracy of the XGBoost model is significantly better than the other models.

Table 7. Calculation results of the determination coefficient and error indicators.

Indices	Models	Training Set	Test Set
MAE	XGBoost	0.029	0.632
	RF	0.838	2.105
	SVM	1.531	2.727
	MLP	3.803	4.632
MAPE (%)	XGBoost	0.067	2.184
	RF	1.984	5.712
	SVM	3.283	6.975
	MLP	9.055	11.115
RMSE	XGBoost	0.042	2.980
	RF	1.729	4.134
	SVM	3.641	5.308
	MLP	5.378	6.582

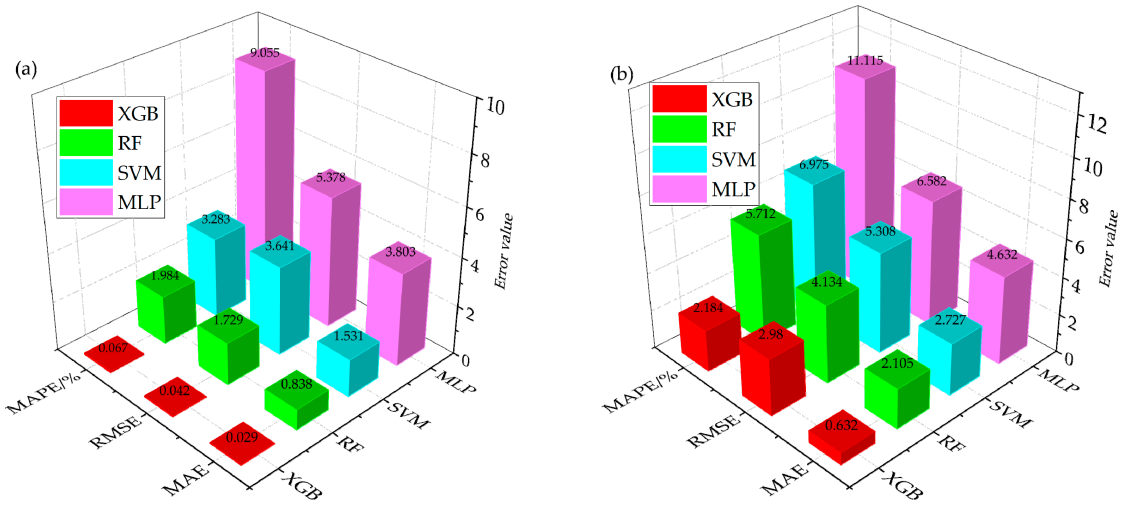


Figure 9. Error distribution histogram of four models. (**a**) For training set and (**b**) for test set.

In the training set, with the results presented in Figure 9a, the XGBoost model is used to predict the hot strip crown, and the calculation results of the RMSE, MAE, and MAPE error indicators are smaller than the others. For the data-driven model, the error on the test set shows the generalization performance of the established model. Figure 9b represents the calculated results for each model on the test set, the MAPE error value of the XGBoost model is 2.184%. The MAPE error value of the RF model is 5.712%. The MAPE error values of the SVM and MLP models are 6.975% and 11.115%, respectively. The MAE error value of the XGBoost model is 0.632. The MAE error value of the RF model is 2.105. The MAE error values of the SVM and MLP models are 2.727 and 4.632, respectively. The RMSE error value of the XGBoost model is 2.980. The RMSE error value of the RF model is 4.134. The RMSE error values of the SVM and MLP models are 5.308 and 6.582, respectively. The calculation results of the three errors show the same rule, that is, that the XGBoost model shows the best prediction accuracy on the test set while maintaining the best training effect of the model. The reason for the above results is that the MLP model uses the BP algorithm for training. The traditional BP algorithm is a local search optimization method, and the weights of the network are gradually adjusted along the direction of local improvement, which can cause the algorithm to fall into local extremum [44], which often leads to the model falling into an over-fitting state, and the training time of the model will increase exponentially with the increase in the number of hidden layers and the number of neurons in each hidden layer. The time spent in model training is shown in Figure 10. Compared with the other three models, MLP has the longest training time. This seriously affects the training efficiency of the MLP model. The SVM algorithm uses quadratic programming to solve support

vectors, which involves designing the calculation of m-order matrices. Therefore, when the matrix order is large, it will cause a decrease in generalization performance and consume a large amount of machine memory and computing time. This also explains why the two algorithms in Figure 6b,d have large absolute errors on the test set. RF and XGBoost all belong to the ensemble learning model. Ensemble learning integrates multiple models through a certain strategy and improves the accuracy of decision-making through group decision-making. The training of the RF model can be highly parallelized, fast, and efficient, which can be proved by Figure 10. The XGBoost model shows better comprehensive performance than RF. Based on the above analysis, it can be concluded that the XGBoost model is most suitable for establishing a prediction model of strip crown in hot rolling under the dataset of this research.

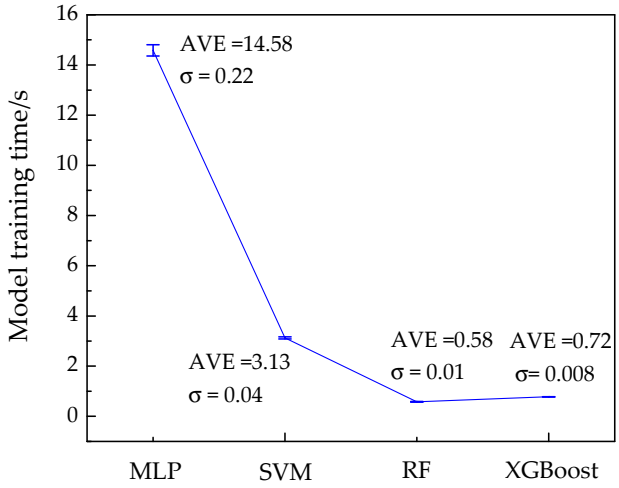


Figure 10. Training time of four models.

6. Conclusions

In this paper, some new data-driven strip crown prediction models integrating the shape control mechanism model, artificial intelligence algorithms, and production data are constructed using the XGBoost, RF, SVM, and MLP algorithms. Through the analysis and evaluation of the prediction results of strip crown for each model, the following main conclusions can be drawn. Using four machine learning algorithms and combining industrial data, the prediction models for hot-rolled strip crown are constructed. Under the same dataset, the XGBoost model has the highest coefficient of determination (R^2) for the prediction results, reaching 0.971 on the test set, and the MLP model has the lowest coefficient of determination for the prediction results, which is 0.860 on the test set. The RF and SVM models are between the XGBoost model and MLP model, with 0.945 and 0.909 on the test set, respectively. The comprehensive performance of four crown prediction models is evaluated using MAE, RMSE, and MAPE. The results show that the prediction model based on the XGBoost algorithm has the smallest errors under the same modeling dataset, showing the best prediction performance and the best generalization performance. Based on these advantages, the combination of the XGBoost algorithm and industrial data can be used to effectively predict the strip crown. By generating corresponding training data during the rolling process, those data-driven prediction methods can easily be extended to predict and optimize other parameters. The research in this paper provides a new method for solving complex industrial problems with multiple variables, strong coupling, and nonlinearity that cannot be handled by traditional mathematical models, and also provides technical support for the effective utilization of massive data and shape control in hot strip rolling.

Author Contributions: Z.W. contributed to the conception of the study, design of experiments, analysis, and manuscript preparation; Y.H. helped perform the analysis with constructive discussions; Y.L. contributed to write, review, and edit; T.W.'s contribution was investigation, supervision, and funding acquisition. All authors have read and agreed to the published version of the manuscript.

Funding: This study is financially supported by the National Natural Science Foundation of China (Nos.: 52105390, 52204397). National Key R&D Program of China (Nos.: 2018YFAO707300), as well as by the special fund for Science and Technology Innovation Teams of Shanxi Province (202204051002007).

Institutional Review Board Statement: This study does not involve any human or animal studies.

Informed Consent Statement: This study does not involve any human or animal studies.

Data Availability Statement: The authors confirm that the data and material supporting the findings of this work are available within the article.

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

References

1. Pittner, J.; Simaan, M.A. A useful control model for tandem hot metal strip rolling. *IEEE Trans. Ind. Appl.* **2010**, *46*, 2251–2258. [CrossRef]
2. Peng, K.; Zhang, K.; Dong, J.; You, B. Quality-relevant fault detection and diagnosis for hot strip mill process with multi-specification and multi-batch measurements. *J. Frankl. Inst.* **2015**, *352*, 987–1006. [CrossRef]
3. Peng, K.; Zhong, H.; Zhao, L.; Xue, K.; Ji, Y.D. Strip shape modeling and its setup strategy in hot strip mill process. *Int. J. Adv. Manuf. Technol.* **2014**, *72*, 589–605. [CrossRef]
4. Pin, G.; Francesconi, V.; Cuzzola, F.A.; Parisini, T. Adaptive task-space metal strip-flatness control in cold multi-roll mill stands. *J. Process Control.* **2013**, *23*, 108–119. [CrossRef]
5. Moazeni, B.; Salimi, M. Investigations on relations between shape defects and thickness profile variations in thin flat rolling. *Int. J. Adv. Manuf. Technol.* **2015**, *77*, 1315–1331. [CrossRef]
6. Tran, D.C.; Tardif, N.; Limam, A. Experimental and numerical modeling of flatness defects in strip cold rolling. *Int. J. Solids Struct.* **2015**, *69*, 343–349. [CrossRef]
7. Linghu, K.; Jiang, Z.; Zhao, J.; Li, F.; Wei, D.; Xu, J.; Zhang, X.; Zhao, X. 3D FEM analysis of strip shape during multi-pass rolling in a 6-high CVC cold rolling mill. *Int. J. Adv. Manuf. Technol.* **2014**, *74*, 1733–1745. [CrossRef]
8. Wang, P.; Peng, Y.; Liu, H.; Zhang, D.; Wang, J. Actuator efficiency adaptive flatness control model and its application in 1250 mm reversible cold strip mill. *J. Iron Steel Res. Int.* **2013**, *20*, 13–20. [CrossRef]
9. Yan, K.J.; Huang, W.Q.; Zuo, Z.X.; Ren, P.R.; Li, D.W.; Zhao, C.Z.; Song, J.N.; Zhang, X.B. Microstructure based analysis and predictive modeling of cast Al7Si1. *5Cu0. 4Mg alloy mechanical properties. Mater. Today. Commun.* **2022**, *30*, 103102.
10. Li, W.G.; Xie, L.X.; Zhao, Y.T.; Li, Z.X.; Wang, W.B. Prediction model for mechanical properties of hot-rolled strips by deep learning. *J. Iron. Steel. Res. Int.* **2020**, *27*, 1045–1053. [CrossRef]
11. Opěla, P.; Schindler, I.; Kawulok, P.; Kawulok, R.; Rusz, S.; Sauer, M. Shallow and deep learning of an artificial neural network model describing a hot flow stress Evolution: A comparative study. *Mater. Des.* **2022**, *220*, 110880. [CrossRef]
12. Churyumov, A.Y.; Kazakova, A.A. Prediction of true stress at hot deformation of high manganese steel by artificial neural network Modeling. *Materials* **2023**, *16*, 1083. [CrossRef] [PubMed]
13. Churyumov, A.; Kazakova, A.; Churyumova, T. Modelling of the steel high-temperature deformation behaviour using arti-ficial neural network. *Metals* **2022**, *12*, 447. [CrossRef]
14. Li, D.W.; Huang, W.Q.; Liu, J.X.; Yan, K.J.; Zhang, X.B. Prediction of mechanical properties of ZL702A based on neural network and regression analysis. *Mater. Today. Commun.* **2022**, *32*, 103679. [CrossRef]
15. Nandan, R.; Rai, R.; Jayakanth, R.; Moitra, S.; Chakraborti, N.; Mukhopadhyay, A. Regulating crown and flatness during hot rolling: A multi-objective optimization study using genetic algorithms. *Mater. Manuf. Process.* **2005**, *20*, 459–478. [CrossRef]
16. Chakraborti, N.; Kumar, B.S.; Babu, V.S.; Moitra, S.; Mukhopadhyay, A. Optimizing surface profiles during hot rolling: A genetic algorithms based multi-objective optimization. *Comput. Mater. Sci.* **2006**, *37*, 159–165. [CrossRef]
17. John, S.; Sikdar, S.; Swamy, P.K.; Das, S.; Maity, B. Hybrid neural-GA model to predict and minimise flatness value of hot rolled strips. *J. Mater. Process. Technol.* **2008**, *195*, 314–320. [CrossRef]
18. Liu, H.; Zhang, X.; Wang, Y. Transfer matrix method of flatness control for strip mills. *J. Mater. Process. Technol.* **2005**, *166*, 237–242. [CrossRef]
19. Peng, Y.; Liu, H.; Du, R. A neural network-based shape control system for cold rolling operations. *J. Mater. Process. Technol.* **2008**, *202*, 54–60. [CrossRef]
20. Yang, L.; Yu, H.; Wang, D.; Zhang, Z.; Jiang, Z. Intelligent shape regulation cooperative model of cold rolling strip and its ap-plication. *Steel Res. Int.* **2017**, *88*, 1600383. [CrossRef]

21. Zhang, X.; Zhang, S.; Tan, G.; Zhao, W. A novel method for flatness pattern recognition via least squares support vector re-gression. *J. Iron Steel Res. Int.* **2012**, *19*, 25–30. [CrossRef]
22. Zhang, X.L.; Zhao, L.; Zhao, W.B.; Xu, T. Novel method of flatness pattern recognition via cloud neural network. *Soft Comput.* **2015**, *19*, 2837–2843. [CrossRef]
23. Zhang, X.; Xu, T.; Zhao, L.; Fan, H.; Zhang, J. Research on flatness intelligent control via GA-PIDNN. *J. Intell. Manuf.* **2015**, *26*, 359–367. [CrossRef]
24. Zhang, X.L.; Cheng, L.; Hao, S.; Gao, W.Y.; Lai, Y.J. The new method of flatness pattern recognition based on GA-RBF-ARX and comparative research. *Nonlinear Dynam.* **2016**, *83*, 1535–1548. [CrossRef]
25. Zhang, X.L.; Cheng, L.; Hao, S.; Gao, W.Y.; Lai, Y.J. Optimization design of RBF-ARX model and application research on flatness control system. *Optim. Contr. Appl. Met.* **2017**, *38*, 19–35. [CrossRef]
26. Deng, J.; Sun, J.; Peng, W.; Hu, Y.; Zhang, D. Application of neural networks for predicting hot-rolled strip crown. *Appl. Soft Comput.* **2019**, *78*, 119–131. [CrossRef]
27. Song, L.; Xu, D.; Wang, X.; Yang, Q.; Ji, Y. Application of machine learning to predict and diagnose for hot-rolled strip crown. *Int. J. Adv. Manuf. Technol.* **2022**, *120*, 881–890. [CrossRef]
28. Xu, J.Z.; Gong, D.Y.; Zhang, W.C.; Chang, X.W.; Liu, X.H.; Wang, G.D. Model building of the initial crown effect rate in 4-high mill. *J. Mater. Sci. Technol.* **2005**, *21*, 165–169.
29. Li, G.; Gong, D.; Lu, X.; Zhang, D. Ensemble learning based methods for crown prediction of hot-rolled strip. *ISIJ Int.* **2021**, *61*, 1603–1613. [CrossRef]
30. Wang, Z.; Liu, Y.; Wang, T.; Gong, D.; Zhang, D. Prediction model of hot strip crown based on industrial data and hybrid the PCA-SDWPSO-ELM approach. *Soft Comput.* **2023**, 1–17. [CrossRef]
31. Di, H.S.; Xu, J.Z.; Gong, D.Y.; Liu, X.H.; Wang, G.D.; He, X.M. Effect of load distribution on strip crown in hot strip rolling. *J. Mater. Sci. Technol.* **2004**, *20*, 330–334.
32. Han, F.; Yao, H.F.; Ling, Q.H. An improved evolutionary extreme learning machine based on particle swarm optimization. *Neurocomputing* **2013**, *116*, 87–93. [CrossRef]
33. Niu, P.; Ma, Y.; Li, M.; Yan, S.; Li, G. A kind of parameters self-adjusting extreme learning machine. *Neural Process Lett.* **2016**, *44*, 813–830. [CrossRef]
34. Karataş, C.; Sozen, A.; Dulek, E. Modelling of residual stresses in the shot peened material C-1020 by artificial neural network. *Expert Syst. Appl.* **2009**, *36*, 3514–3521. [CrossRef]
35. Rafei, M.; Sorkhabi, S.E.; Mosavi, M.R. Multi-objective optimization by means of multi-dimensional MLP neural networks. *Neural Netw. World* **2014**, *24*, 31–56. [CrossRef]
36. Mohammad, A.T.; Hussen, H.M.; Akeiber, H.J. Prediction of the output power of photovoltaic module using artificial neural networks model with optimizing the neurons number. *Int. J. Renew. Energy Dev.* **2023**, *12*, 478–487. [CrossRef]
37. Suykens, J.A.; Van Gestel, T.; De Brabanter, J. *Least Squares Support Vector Machines*; World Scientific: Singapore, 2002.
38. Vapnik, V. *Statistical Learning Theory*; Wiley: New York, NY, USA, 1998.
39. Breiman, L. Bagging predictors. *Mach. Learn.* **1996**, *24*, 123–140. [CrossRef]
40. Ho, T.K. The random subspace method for constructing decision forests. *IEEE Trans. Pattern Anal. Mach. Intell.* **1998**, *20*, 832–844.
41. Chen, T.; Guestrin, C. XGBoost: A scalable tree boosting system. In Proceedings of the 22nd Acm Sigkdd International conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 785–794.
42. Dong, J.; Chen, Y.; Yao, B.; Zhang, X.; Zeng, N. A neural network boosting regression model based on XGBoost. *Appl. Soft Comput.* **2022**, *125*, 109067. [CrossRef]
43. Song, K.; Yan, F.; Ding, T.; Gao, L.; Lu, S. A steel property optimization model based on the XGBoost algorithm and improved PSO. *Comp. Mater Sci.* **2020**, *174*, 109472. [CrossRef]
44. Gori, M.; Tesi, A. On the problem of local minima in backpropagation. *IEEE Trans. Pattern Anal. Mach. Intell.* **1992**, *14*, 76–86. [CrossRef]

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