



# Vibration prediction and analysis of strip rolling mill based on XGBoost and Bayesian optimization

Yang Zhang<sup>1</sup> · Ranmeng Lin<sup>1</sup> · Huan Zhang<sup>1</sup> · Yan Peng<sup>2</sup>

Received: 7 March 2022 / Accepted: 27 May 2022 / Published online: 23 June 2022  
© The Author(s) 2022

## Abstract

The stable operation of strip rolling mill is the key factor to ensure the stability of product quality. The design capability of existing domestic imported and self-developed strip rolling mills cannot be fully developed, and the frequent occurrence of mill vibration and operation instability problems seriously restrict the equipment capacity and the production of high-end strip products. The vibration prediction analysis method for hot strip mill based on eXtreme gradient boosting (XGBoost) and Bayesian optimization (BO) is proposed. First, an XGBoost prediction model is developed based on a self-built data set to construct a complex functional relationship between process parameters and rolling mill vibration. Second, the important hyperparameters and parameters of XGBoost are optimized using Bayesian optimization algorithm to improve the prediction accuracy, computational efficiency, and stability of the model. Third, a comprehensive comparison is made between the prediction model in this paper and other well-known machine learning benchmark models. Finally, the prediction results of the model are interpreted using the SHapley Additive exPlanations (SHAP) method. The proposed model outperforms existing models in terms of prediction accuracy, computational speed and stability. At the same time, the degree of influence of each feature on rolling mill vibration is also obtained.

**Keywords** Strip rolling mill · Vibration prediction · eXtreme gradient boosting · Bayesian optimization · SHapley additive exPlanations

## Introduction

With the significant improvement of the process and equipment technical level in the current iron and steel industry, rolling equipment technology has become a bottleneck limiting the overall level of iron and steel industry. The stable operation of the strip rolling mill is a key link to guarantee the stability of the production process and the quality of the strip. Domestic introduced and independently developed strip rolling mills, whose design capacity cannot fully play because of the frequent equipment vibration and instability operation problems. The first crucial reason which caused this phenomenon is the low optimization level of parameters in the rolling process. Considering the strong dynamic

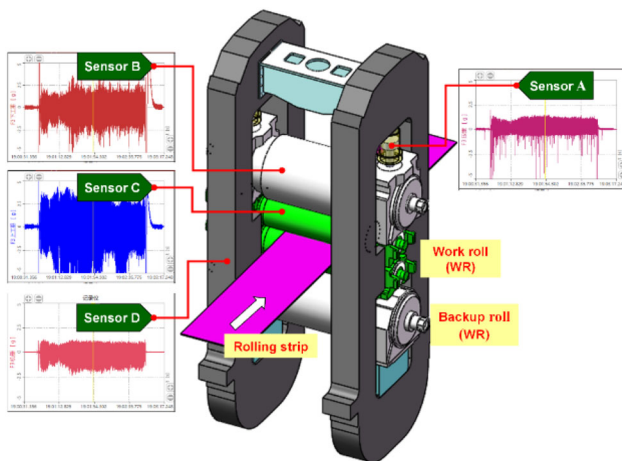
coupling effect between the components of the rolling mill system and the large fluctuation of the equipment state, it is difficult to accurately establish the coupling mechanism model between equipment vibration and rolling process, and realize the deep optimization of process parameters. The second main reason is the lack of rolling mill operation condition monitoring means. Through the field investigation of different enterprises, it is found that most strip rolling mills lack vibration detection method, and it is difficult to get equipment running status. As a result, equipment operation state can only be judged by monitoring the signal fluctuation of process parameters, and the rolling mill vibration trend cannot be accurately predicted.

The field vibration tests show that the low matching degree between rolling mill equipment and process parameters is the main factor causing the instability of rolling mill operation. However, the traditional rolling process control system rarely contains the relationship model between process parameters and vibration state of rolling mill, which leads to the mismatch between equipment status and process parameters in the long-term operation of strip rolling mill. The vibration of

✉ Yang Zhang  
gyangzhang@tyust.edu.cn

<sup>1</sup> School of Mechanical Engineering, Taiyuan University of Science and Technology, Taiyuan 030024, Shanxi, China

<sup>2</sup> National Engineering Research Center for Equipment and Technology of Cold Rolling Strip, Yanshan University, Qinghuangdao 066004, Hebei, China



**Fig. 1** The dynamic properties of strip rolling mill

strip rolling mill is a fusion of multi-source vibration signals caused by the coupling effect of complex internal structure and external load fluctuation of the system, as shown in Fig. 1. To accurately analyze the operation instability mechanism of strip rolling mill and reveal the coupling mechanism between complex dynamic responses in the rolling mill system, it is necessary to study the influence of rolling process parameters on the stability of strip rolling mill system.

Aiming at the problem of rolling mill vibration, domestic and foreign researchers have studied the mechanism of rolling mill vibration [1–3], vibration signal analysis [4, 5], and vibration suppression [6, 7] using the methods of theoretical analysis, numerical calculation, experimental simulation, and engineering verification. Gao et al. [8] summarize in detail the research on rolling mill vibration. However, strip rolling mill is a complex multi-body mechanical system composed of frame, depress hydraulic device, roll system, and other systems. It is difficult to establish an accurate mechanism model between equipment dynamic characteristics and process parameters to predict and analyze the trend of rolling mill vibration. With the rapid development of information technology, data mining technology is widely used in the predictive maintenance of rolling equipment, improvement of product quality, in-depth optimization of process parameters, and so on. Therefore, to quantitatively analyze the influence law of rolling process parameters on the operation stability of strip rolling mill, it is necessary to study the action law between the fluctuation of process parameters and the operation state of rolling mill, and deeply excavate the correlation between mill vibration data and rolling process parameters.

Bagheripour et al. [9] applied the artificial neural network algorithm to hot strip rolling mill to improve the prediction ability of the rolling force and rolling torque prediction model in the rolling process. Ma et al. [10] proposed a data-based quality related fault diagnosis scheme for hot strip mill

process' equipment fault diagnosis and fault cause analysis. Liu et al. [11] constructed an intelligent prediction model of rolling mill chatter based on long short-term memory (LSTM) recurrent neural network, and predicted the vibration using the historical information data of rolling piece specification, roll condition, rolling process, and rolling mill vibration state. Lu et al. [12] developed data-driven vibration prediction of cold rolling mill, and proposed the XGBoost model to predict rolling mill vibration. Chen et al. [13] and Pan et al. [14] designed a data-driven condition monitoring system to detect mechanical faults of bearings in the main driven system of hot tandem rolling mill. Dong et al. [15] used DBN and GA-BP algorithm to establish the rolling mill vibration prediction model to predict the rolling mill vibration. Deng et al. [16] established a data-based neural network model for prediction of strip crown in hot strip rolling mill. Song et al. [17] built a steel property optimization model based on the XGBoost algorithm and improved particle swarm optimization algorithm to improve and optimize the mechanical properties of steel. Shi et al. [18] established a train arrival delay prediction model based on the XGBoost and BO, and analyzed the prediction efficiency and accuracy of models with different stations. Zhou et al. [19] combined XGBoost with BO to estimate the advance rate of tunnel boring machine under hard rock conditions. Liang et al. [20] used GBDT, XGBoost, and LightGBM algorithms to predict the stability of hard rock columns. Mongan et al. [21] used particle swarm optimization artificial neural network to predict the quality of ultrasonic welded joints.

The prediction model of industrial data combined with machine learning algorithm has been widely used in industry [22–26]. In this paper, a vibration prediction model for hot strip mills based on the XGBoost and BO has been proposed. The main innovations of this paper are illustrated as follows.

- (1) Based on the self-built dataset, an XGBoost prediction model was developed, which took process parameters as model input variables and rolling mill vibration as model output variables, and accurately constructed complex nonlinear relationships between input and output variables.
- (2) The XGBoost hyperparameters and parameters were optimized using the BO algorithm to address the problems of slow computation speed, model stability, and prediction accuracy of the existing model using the GS algorithm.
- (3) The prediction results were interpreted using the SHAP method which fulfilling the technical gap of the lack of interpretability of machine learning models in the field of mill vibration prediction.

This paper is organized as follows: “**Methodology**” briefly describes the XGBoost, BO, and SHAP algorithms, and

provides performance metrics for the experimental study. “Data set and data pre-processing” describes the self-created dataset used in the study, including data collection, data pre-processing, and feature engineering. In “Results and discussion”, different algorithms are comprehensively evaluated, and the model in this paper is compared with other machine learning models. In “Model interpretation”, the prediction model in this paper is explained as a whole and locally, and the influence of different process parameters on rolling mill vibration is deeply analyzed. In “Conclusion”, the experimental conclusions have been summarized.

## Methodology

### XGBoost

XGBoost, proposed by Chen and Guestrin [27] in 2016, is an extensible end-to-end machine learning model. This model has the advantages of efficient tree pruning, regularization, and parallel processing. It has been used in many engineering fields to solve the industrial application problems [28–30]. XGBoost model optimizes the predicted value before each iteration through the residual. To solve the problem of over fitting in the optimization process, the objective function is regularized which can be described as

$$\mathcal{I}(\theta) = L(\theta) + \Omega(\theta), \tag{1}$$

where  $\theta$  is the parameter trained from the given data;  $\Omega$  indicates regularization;  $L$  denotes the training loss function (LOF), which is used to reveal the match degree the model and training data. Equation (2) is the prediction function; according to the Decision Tree (DT) theory, the output of the model  $\hat{y}_i$  depends on the mean of the votes or sets

$$\hat{y}_i = \sum_{k=1}^k f_k(x_i), f_k \in F. \tag{2}$$

The objective function of  $t$  time iteration can be showed with a mathematical model as follows:

$$\mathcal{I}(t) = \sum_{i=1}^n L(y_i, \hat{y}_i) + \sum_{k=1}^t \Omega(f_k), \tag{3}$$

where  $n$  is the number of prediction and  $\hat{y}_i^{(t)}$  can be defined as

$$\hat{y}_i^{(t)} = \sum_{k=1}^t f_k(x_i) = \hat{y}_i^{(t-1)} + f_t(x_i). \tag{4}$$

As shown by Chen and Guestrin [27], the regularization term  $\Omega(f_k)$  of DT is represented as

$$\Omega(f_k) = \gamma T + \frac{1}{2} \lambda \sum_{j=1}^T \omega_j^2. \tag{5}$$

$\lambda$  represents the penalty factor and  $T$  delegates the number of leaves in the DT,  $\gamma$  denotes the complexity of each leaf, and  $\omega$  is a vector of scores on the leaves. A second-order (instead of first-order) Taylor expansion in general gradient boosting is applied to the LOF in XGBoost (Chen and Guestrin [27]). Assuming the mean square error (MSE) as the LOF, the objective function can be obtained from the following equation:

$$\mathcal{I}(t) \approx \sum_{i=1}^n \left[ g_i \omega_{q(x_i)} + \frac{1}{2} (h_i \omega_{q(x_i)}^2) \right] + \gamma T + \frac{1}{2} \lambda \sum_{j=1}^T \omega_j^2, \tag{6}$$

where  $g_i$  and  $h_i$  denote the first-order and second-order derivatives of the MSE loss function, respectively, and  $q$  is the function that assigns a data point to the corresponding leaf.

Obviously, the LOF in Eq. (6) depends on the sums of loss values for each data sample. Since each data sample corresponds to only one leaf node, the LOF can also be used by the sums of the loss values of each leaf node, that is

$$\mathcal{I}(t) \approx \gamma T + \sum_{j=1}^T \left[ \left( \sum_{i \in I_j} g_i \right) \omega_j + \frac{1}{2} \left( \sum_{i \in I_j} h_i + \lambda \right) \omega_j^2 \right] \tag{7}$$

Therefore,  $G_j$  and  $H_j$  are defined as

$$G_j = \sum_{i \in I_j} g_i, H_j = \sum_{i \in I_j} h_i, \tag{8}$$

where  $I_j$  denotes all data samples in leaf node  $j$ .

As a result, the objective function can be optimized with translating into the process of finding the minimum of a quadratic function. Figure 2 shows the XGBoost training flowchart.

### Bayesian optimization (BO)

The choice of hyperparameters is critical to model performance [31–33]. BO has proven to be a very effective optimization algorithm for solving machine learning optimization problems. According to Bayes’ theorem, given the observation point  $E$ , the posterior probability  $P(M|E)$  of the model  $M$  is proportional to the likelihood ratio probability

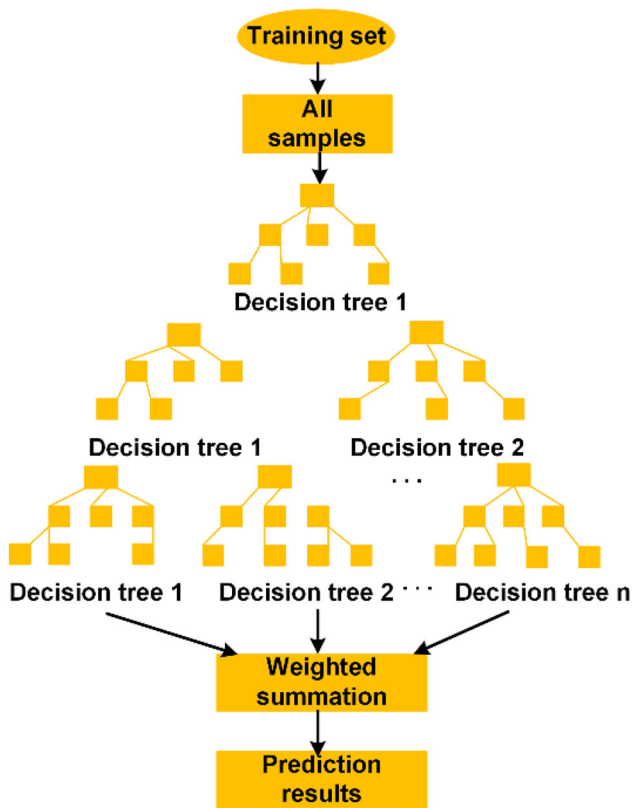


Fig. 2 XGBoost training flowchart

$P(E|M)$  of the observation point  $E$  multiplied by the prior probability  $P(M)$  of the model  $M$ , that is

$$P(M|E) \propto P(E|M)P(M). \quad (9)$$

The BO algorithm is based on the historical evaluation results of the objective function to build a proxy model of the objective function, which makes full use of the previous evaluation information when selecting the next set of hyperparameters, reduces the retrieval times of hyperparameters. As a result, the obtained hyperparameters are most likely to be optimal, thus improving the prediction accuracy and generalization ability of the model [34–39].

### SHapley Additive exPlanations (SHAP)

The interpretability of machine learning model is very important, because it can provide mechanism explanation of the machine learning model to make the best decision. SHAP is a method which is used to explain “black box” of machine learning model. SHAP is derived from the ideas of Shapley’s game theory and was first proposed by Lundberg and Lee [40]. SHAP attempts to evaluate the contribution of each input feature to the model output, and analyze whether the



Fig. 3 1580 mm hot strip rolling mill

contribution of each feature is negative or positive. Meanwhile, SHAP can calculate the contribution of each feature for each predicted output.

### Model evaluation indicators

To effectively evaluate the reliability of the vibration prediction model and carry out comparative experiments between different algorithms, the relationship between the predicted value and the real value of the model is evaluated with the coefficient ( $R^2$ ), mean square error (MSE), and mean absolute error (MAE) as the evaluation index. The calculation formula of the evaluation index is as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i - \bar{y})^2} \quad (10)$$

$$\text{MSE} = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2 \quad (11)$$

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i|, \quad (12)$$

where  $y_i$  denotes the true value,  $\hat{y}_i$  denotes the predicted value,  $\bar{y}$  denotes the sample mean, and  $N$  denotes the data sample size.

### Data set and data pre-processing

#### Data collection

Figure 3 shows the 1580 mm hot tandem strip rolling mill production line, which mainly consists of seven four-high mills. Figure 4 shows the vibration and process data acquisition system.

The measuring points of the sensors were arranged in the mill stand, depress hydraulic (DH) cylinder, backup roll (BUR) bearing seat, and work roll (WR) bearing seat, which

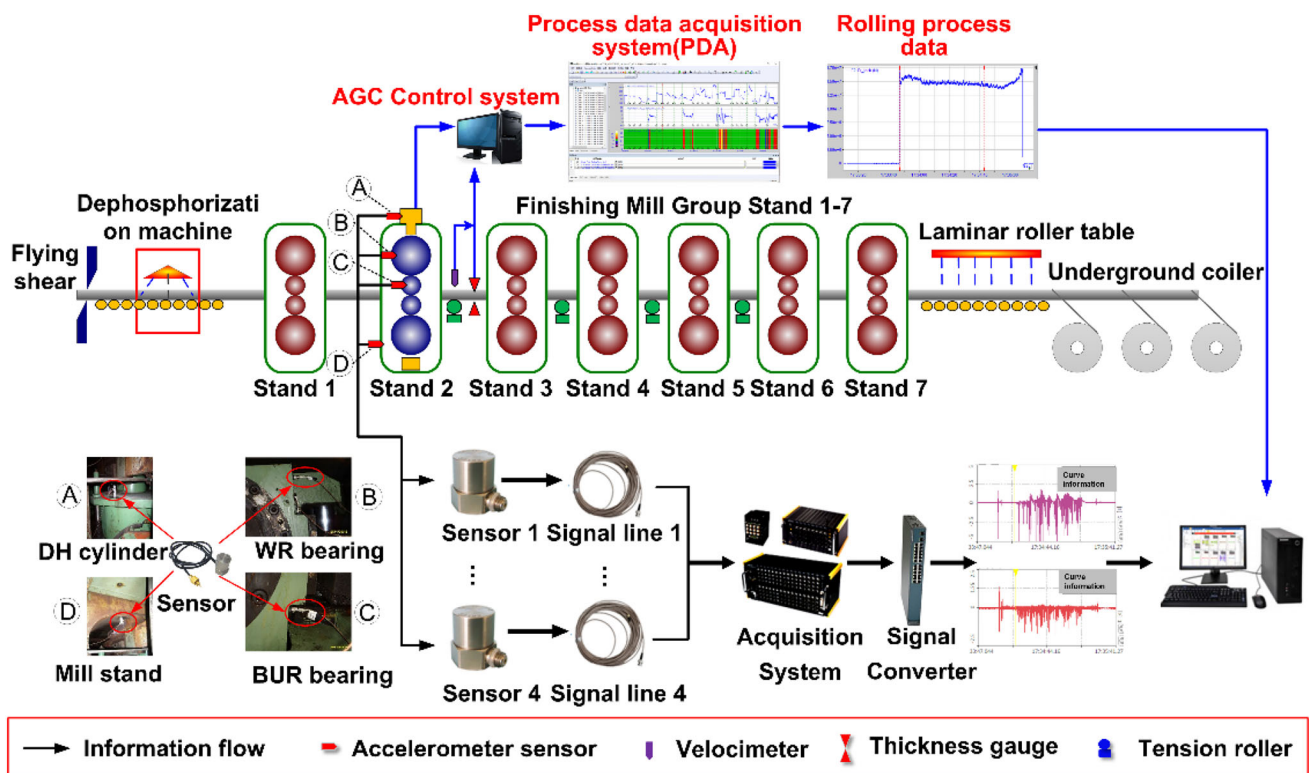


Fig. 4 Vibration and process data acquisition system

Table 1 Raw data table

Rolling speed (m/s)	Entrance thickness (mm)	Rolling force (N)	Outlet thickness (mm)	Back tension (KN)	Vibration acceleration (m/s <sup>2</sup> )
2.29204	0.0208033	26148.9	0.0120832	127.507	- 2.2364
2.29204	0.0208033	26148.9	0.0120832	127.507	- 2.20771
2.29335	0.0208033	26312.3	0.0120832	126.914	- 2.1839
2.29434	0.0208033	24826.3	0.0120832	126.426	- 2.12409
2.29368	0.0208033	21638.6	0.0120832	126.578	- 2.04779
2.29368	0.0208043	26,087	0.0120831	127.204	- 1.91534
2.29401	0.0208033	21834.2	0.0120831	126.979	- 1.84698
2.29401	0.0208033	21834.2	0.0120831	126.979	- 1.80425
2.29368	0.0208033	28153.8	0.0120841	126.307	- 1.74383
2.29368	0.0208033	28153.8	0.0120841	126.307	- 1.68645
2.29368	0.0208033	21518.2	0.0120831	126.68	- 1.62847
2.29303	0.0208033	25,811	0.0120831	127.024	- 1.5772
2.29171	0.0208033	22,899	0.0120831	126.934	- 1.55156

are prone to serious vibration in the rolling process. Position A is the depress hydraulic cylinder, position B is the backup roll bearing seat, position C is the work roll bearing seat, and position D is the mill stand. Based on the professional field knowledge, the process parameters, such as back tension, entrance thickness, outlet thickness, rolling force, and rolling speed, are closely related to rolling mill vibration. Based on

the field vibration test, it is known that horizontal vibration of the upper work roll has a significant impact on the product quality. Therefore, the process parameters of back tension, entrance thickness, outlet thickness, rolling force, and rolling speed are selected as the model input variables. At the same time, the upper work roll horizontal vibration is selected as the model output variable.

**Table 2** The statistics of the original data set

Samples	Rolling speed	Entrance thickness	Rolling force	Outlet thickness	Back tension	Vibration
Count	14,016	14,016	14,016	14,016	14,016	14,016
Mean	2.1618	0.0204	8,758,916	0.0121	138.74	0.0826
Std	0.4836	0.0007	8,078,373	0.0003	12.11	1.2151
Min	0.4450	0.0191	– 919,154	0.0108	62.21	– 7.219
25%	2.2901	0.0206	31,370	0.0120	127.13	– 0.3363
50%	2.2943	0.0210	15,998,900	0.0121	141.58	0.4358
75%	2.3184	0.0208	16,168,600	0.0123	150.41	0.7312

Excluding the anomalous data of the moment of steel biting and steel throwing, a total of 14,016 sets of valid data were collected. The whole data set was randomly divided into two sets, of which 80% was used as the training set to train the prediction model, and the other 20% was used as the test set to verify the prediction model. When dividing the data set, the random number seed is set to ensure that the data used to train and test different machine learning models are consistent every time. Under the condition that the training set is consistent, different optimization algorithms are used to optimize the important hyperparameters and parameters of the model, to ensure that the algorithm is as fair as possible. Table 1 shows some original data, Table 2 provides the statistics of the original data set.

### Data description and pre-processing

As shown in Fig. 5, the violin diagram describes the data distribution, and outliers' analyses of five input variables and one predicted output variable are described by violin diagram. Among them, the data distribution of rolling speed is shown in Fig. 5a. The rolling speed is relatively stable in the whole rolling process, mainly in the range of 2.20–2.35 m/s. During the early stage of rolling process, there are some abnormal values which should be eliminated. The data distribution of entrance thickness is shown in Fig. 5b. It can be seen from Fig. 5b that the entrance thickness is fluctuating obviously in the range of 0.0206–0.0210 m, which may be the main reason for the vibration of the rolling mill. As illustrated in Fig. 5c, the data distribution of rolling force is mainly distributed below 4000 KN in the rolling mill start-up stage and 14,000–20000 KN in the stable rolling stage. Figure 5d shows the data distribution of outlet thickness, which was concentrated near 0.012 mm. As shown in Fig. 5e, the data distribution of post-tension is mainly distributed near 130kN and 160kN. Figure 5f represents the data distribution of vibration acceleration. It can be seen from the Fig. 5f that the data distribution range is large, and there are many abnormal values at the same time. In addition, considering that the data set used in this paper has no missing values (as shown in Table

2) and there are only a few outliers, the elimination method, which would not affect the accuracy of model, is used in this paper to delete outliers.

### Feature engineering

Figure 6 illustrates the correlation analysis between input variables and output variable. According to the professional field knowledge, the best input variable is selected by setting the threshold to 0.15. It can be seen from Fig. 6 that the correlation between the five input variables and the output variable was weak, which was in line with the features of complex nonlinear dynamic coupling characteristics of the strip rolling mill system. In addition, the correlation coefficient between the five process parameters and vibration is greater than 0.15. Therefore, the five process parameters are selected as input variables.

### Results and discussion

Based on the production data collected by the strip rolling mill vibration and process data acquisition system (Fig. 4), a strip rolling mill vibration prediction model based on the XGBoost and BO is established through preprocessed data, as shown in Fig. 7.

### Hyperparameters and parameters setting

To improve the prediction accuracy of the prediction model, GS, RS, and BO were used to optimize the important hyperparameters and parameters of XGBoost model, and the hyperparameter configuration with higher prediction accuracy than the default XGBoost model was obtained. It should be noted that the number of estimators (recorded as  $\alpha$ ) is an important hyperparameter of the model. If the value is too small, the model will be under fitted. If the value is too large, the model will be over fitted. The max depth (recorded as  $\beta$ ) is another significant parameter of the model, which represents the complexity of the model. Generally, the tree model

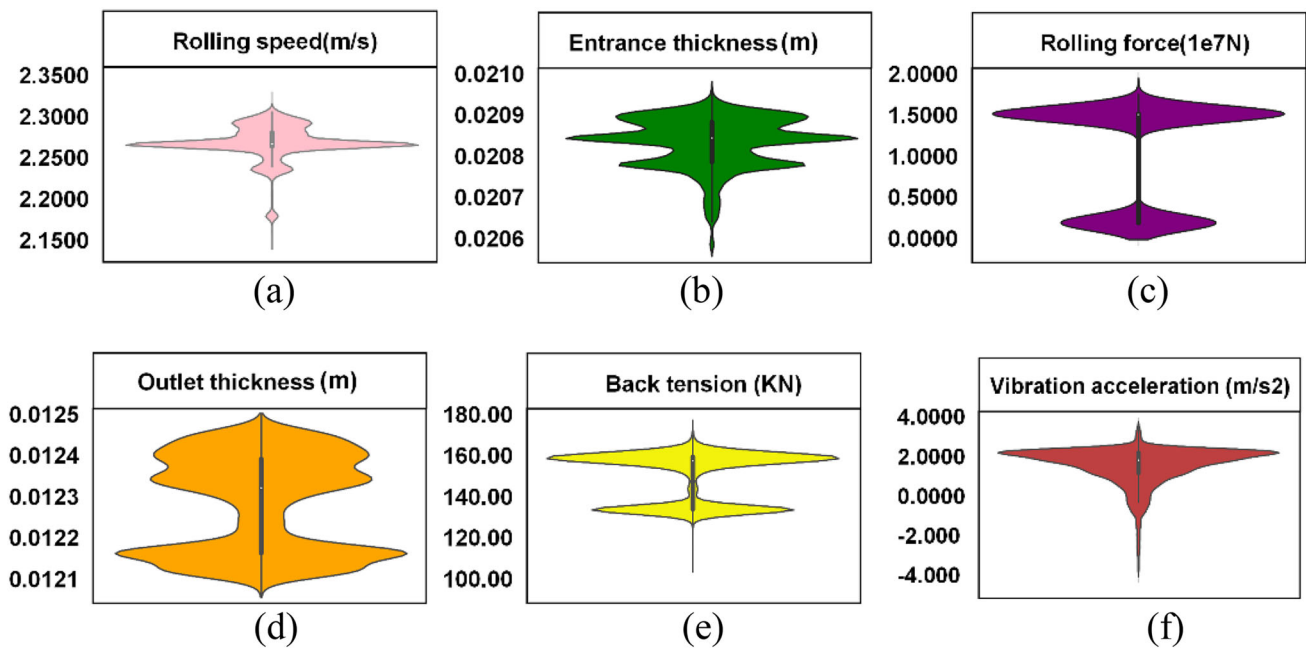
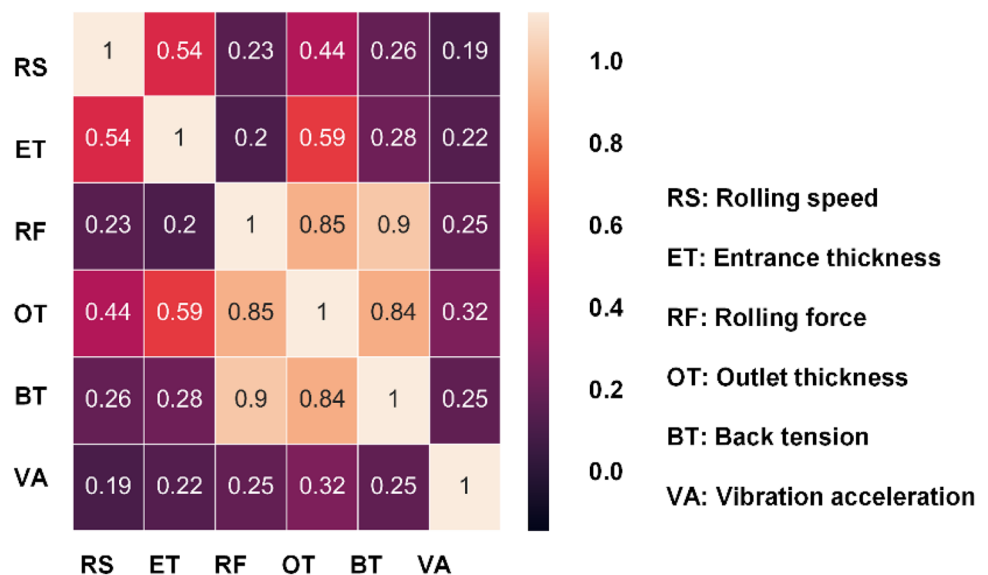


Fig. 5 Violin diagram of input and output data distribution

Fig. 6 Input–output correlation analysis



is pruned by setting appropriate values to avoid the problem of overfitting the model. The learning rate (recorded as  $\gamma$ ) is also a critical parameter of the model, whose value indicates the learning speed of the model. Furthermore, the searching range of the same hyperparameters should be consistent to ensure the fairness of the comparison results for different optimization algorithms. The hyperparameters searching space and optimal values of different optimization algorithms are shown in Table 3.

### Comparison and analysis of results before and after optimization

Figure 8 shows the optimization results of three algorithms: GS, RS, and BO. It can be seen from Fig. 8 that the prediction performance of the optimized model was significantly improved compared with the XGBoost model under the default hyperparameters. Among them, the XGBoost prediction model optimized by BO possessed the best indexes, and the  $R^2$  reached 0.9131, which was better than the other two optimization algorithms. The results revealed that vibration prediction model based on the XGBoost and Bo could

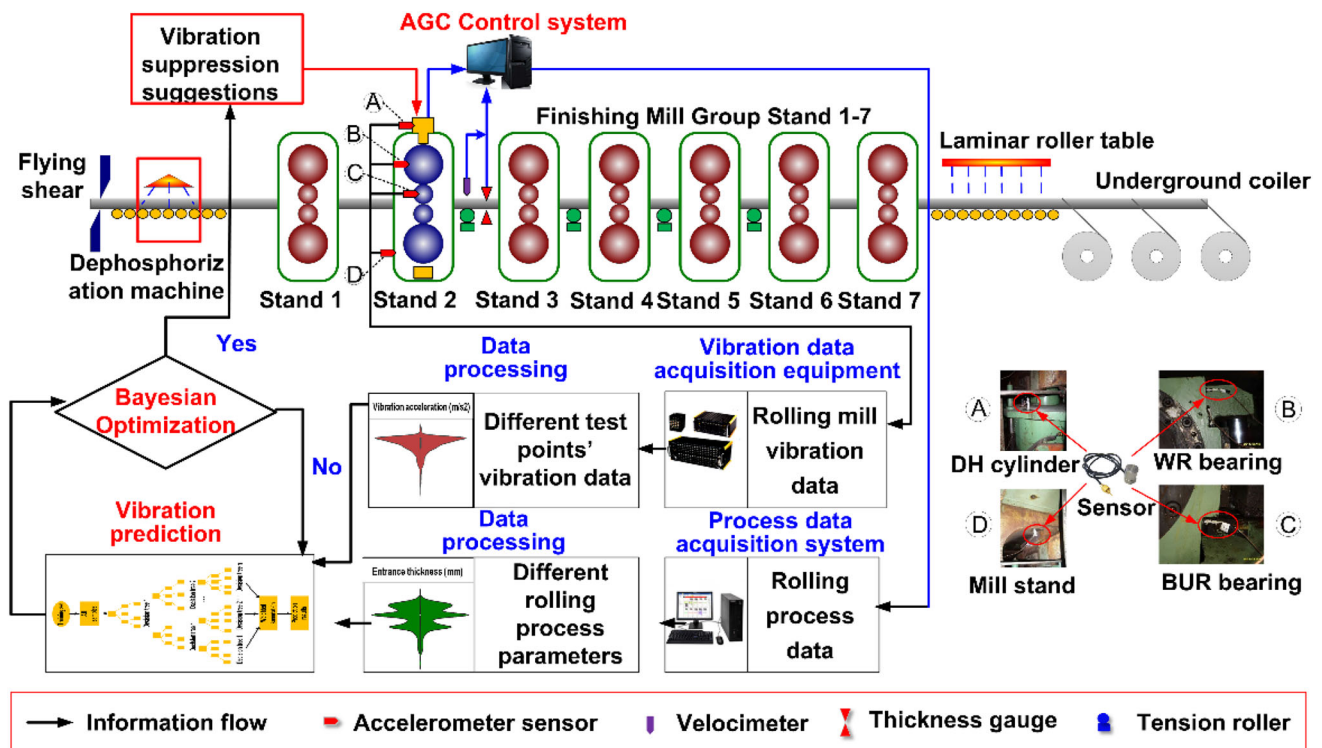


Fig. 7 Flowchart of data-driven vibration prediction method for strip rolling mill

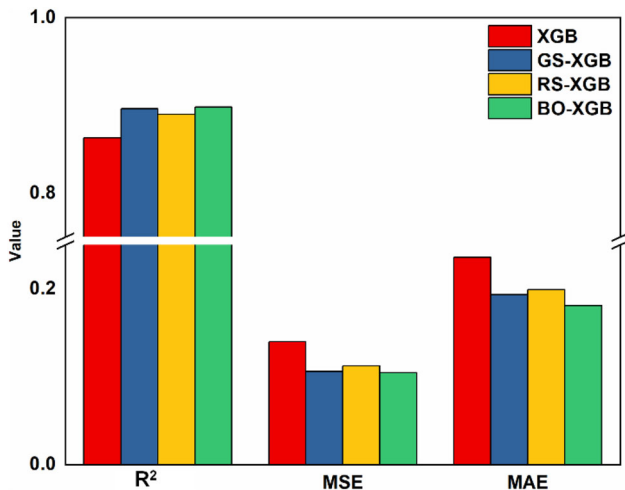


Fig. 8 Comparison of results before and after optimization

better fit the complex relationship between input process parameters and output vibration variables, and obtained better prediction results. As a consequence, the prediction model proposed in this paper was suitable for vibration prediction of hot strip rolling mill system.

To evaluate the stability of the three optimization algorithms, GS adopted cross validation technology, and RS and BO realized the optimization of model hyperparameters by setting a fixed number of iterations. The  $R^2$  distribution in the process of hyperparameters optimization is shown in

Table 3 Searching space and optimal hyperparameters of XGBoost model

Parameter	Low limit	Upper limit	Optimization		
			BO	GS	RS
$\alpha$	50	300	172	130	168
$\beta$	5	11	8	8	8
$\gamma$	0.01	1	0.25	0.2	0.15

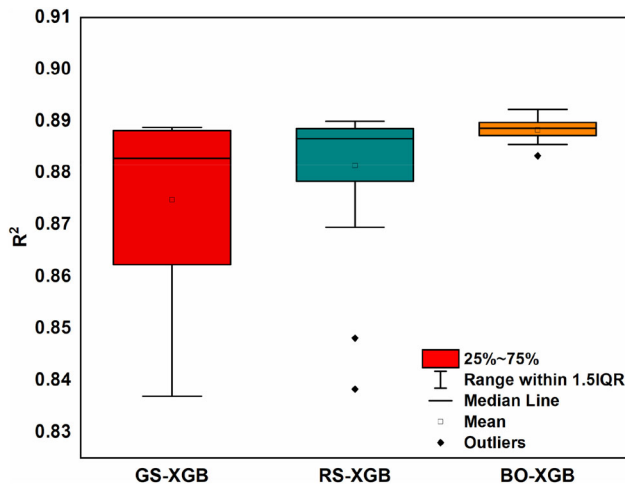
Fig. 9. From the distribution picture, the best prediction performance of the three optimization algorithms was very close, but the prediction performance stability of the model BO-XGB was the best. Therefore, the vibration prediction model based on the XGBoost and BO has stronger stability.

The running time of three optimization algorithms was tested using different proportions of original data set (10%, 30%, 50%, 70%, and 100%). The running time, under different proportions, of three optimization algorithms is shown in Table 4. Due to the enumeration method, the running time of GS is the longest, and the defects became more and more obvious with the increase of data sample size. RS had the fastest computing speed, but it was easy to miss the optimal value due to the low stability of the algorithm. As illustrated in Table 4, the considering the prediction performance and calculation speed, BO algorithm was selected to optimize the hyperparameters and parameters of the model.

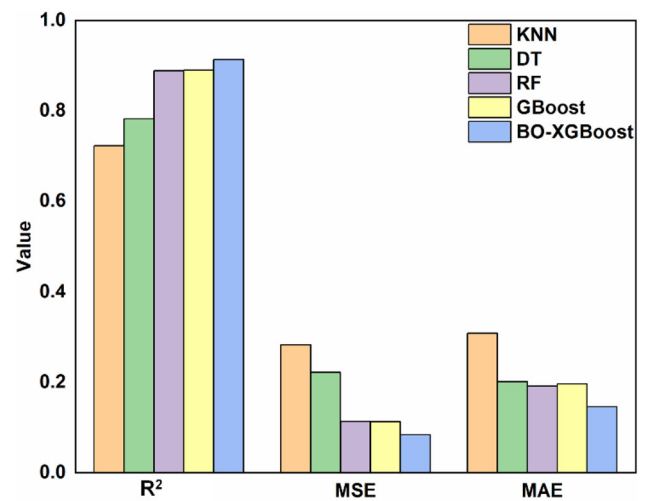


**Table 4** Model performance at different scales

Relative ratio (%)	Parameter dimension	BO		GS		RS	
		$R^2$	Time (s)	$R^2$	Time (s)	$R^2$	Time (s)
10 ( $n = 1401$ )	2	0.778	42	0.777	128	0.778	25
30 ( $n = 4203$ )	2	0.789	75	0.787	258	0.739	46
50 ( $n = 7008$ )	2	0.802	95	0.800	353	0.789	64
70 ( $n = 9811$ )	2	0.872	113	0.874	434	0.871	79
100 ( $n = 14,016$ )	2	0.913	136	0.907	560	0.901	101



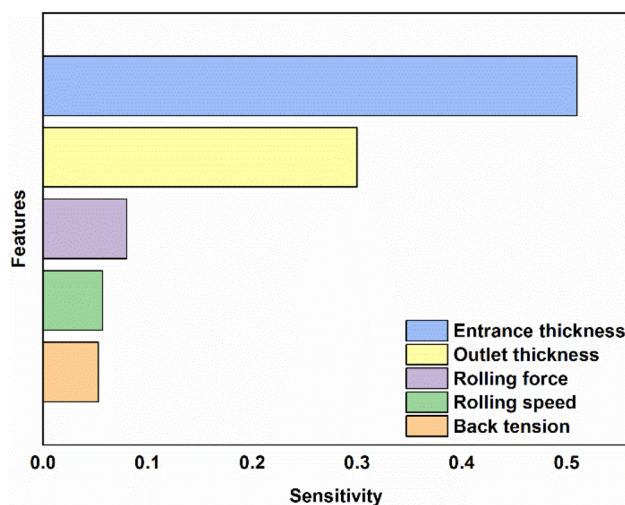
**Fig. 9**  $R^2$  distribution in the optimization process of three optimization algorithms



**Fig. 10** Comparison of BO-XGBoost with other machine learning models

**Model comparison of prediction performance**

To verify the prediction performance of the vibration prediction model of hot strip rolling mill based on the XGBoost and BO which is proposed in this paper, four machine learning models, including K-Nearest Neighbor (KNN), Decision Tree (DT), Random Forest (RF), and Gradient Boosting (GBoost), are selected for comparison and verification. These models have been applied to a variety of industrial fields and achieved good results. To ensure the fairness of model comparison, all the models were trained with the same training set and tested with the same test set. As shown in Fig. 10, the calculation effect of the KNN model is the worst effect, because the KNN model is not good at predicting data sets with large sample size. The prediction accuracy of RF model is higher than that of DT model, because RF is a parallel integrated learning algorithm, which generates multiple trees for weighted summation. At the same time, Gboost is also an integrated algorithm based on DT, which can promote weak learners to strong learners. Moreover, because the algorithm attaches importance to deviation, the prediction



**Fig. 11** Sensitivity analysis of process parameters based on BO-XGBoost

Fig. 12 SHAP summary plot

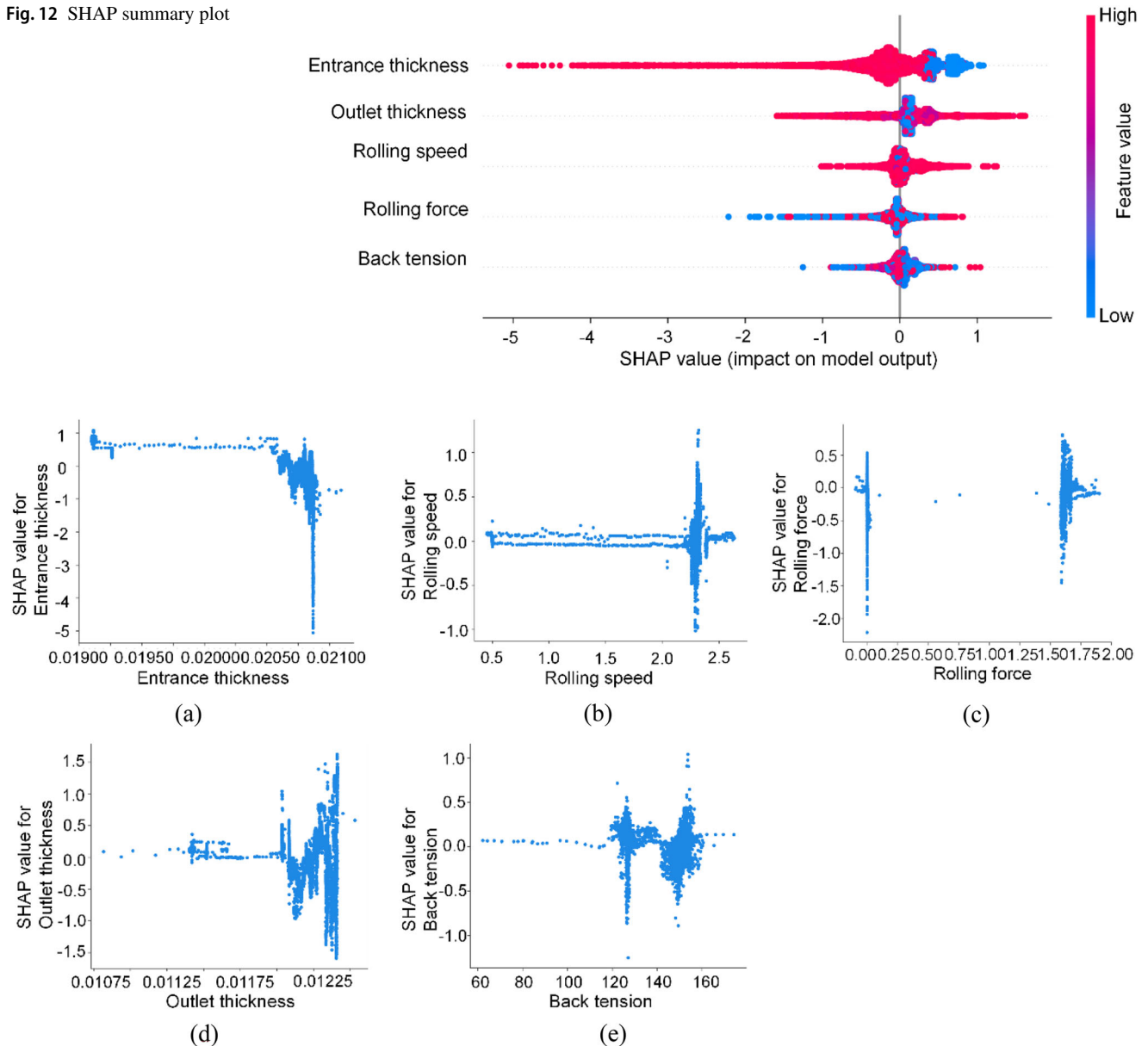


Fig. 13 SHAP feature dependence plots

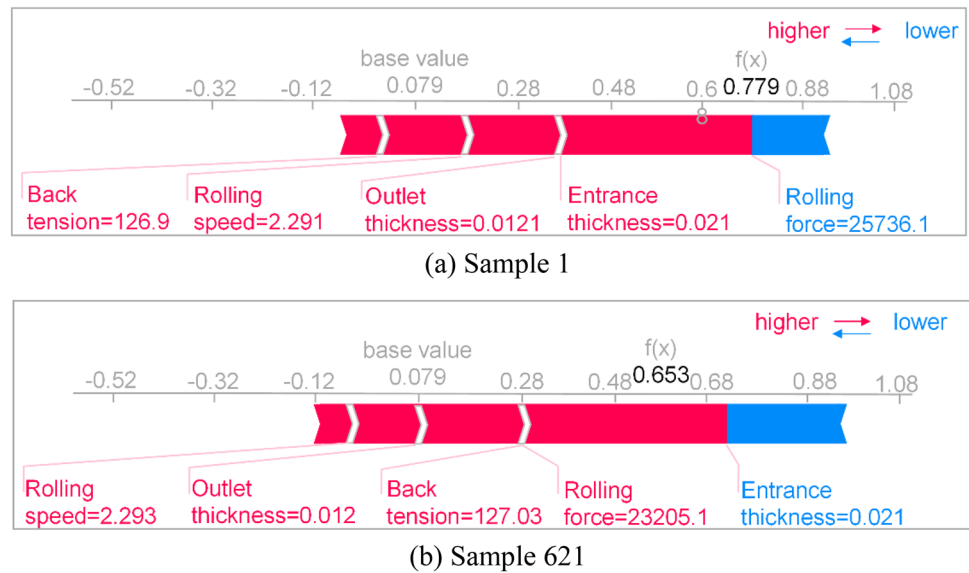
performance is better. In particular, the BO-XGBoost vibration prediction model established in this paper optimizes the loss function and improves the prediction accuracy by removing the constant term in the second-order Taylor expansion of the objective function. Therefore, the BO-XGBoost has better vibration prediction performance than GBoost.

## Model interpretation

Figure 11 shows the importance of the features based on the BO-XGBoost model. The figure ranks the features according to the magnitude of their contribution to the computational

process. Figure 11 is obtained by calling the “xgb. feature\_importances” function in the XGBoost model. From Fig. 11, the entrance thickness is the most important feature affecting the model prediction results, while the outlet thickness, rolling force, rolling speed, and back tension contribute relatively little to the model prediction results.

The traditional XGBoost model cannot explain the influence law of each feature on the prediction results, and cannot evaluate the contribution of each feature to the prediction results. SHAP emphasizes the contribution of each feature to the corresponding prediction model and to the global and local behavior by assigning an SHAP value to each input

**Fig. 14** SHAP feature force plot

variable to indicate its contribution to the result. Global interpretation aims to provide an overview of the SHAP values for input features of all samples. Figure 12 provides a global SHAP summary plot for the entire dataset, where the input features are placed on the y-axis according to their contribution to the rolling mill vibration prediction. The features are sorted from top to bottom based on the magnitude of their contribution, and the SHAP values are on the x-axis. Feature values are represented by colors, where blue-to-pink represent values from low to high. It can be seen from Fig. 12 that the entrance thickness has the greatest impact on the rolling mill vibration. The increase of entrance thickness can reduce the SHAP value, which shows that increasing entrance thickness can reduce rolling mill vibration. In addition, the contribution values of exit thickness, rolling speed, rolling force, and post-tension decrease in turn. It is worth noting that the smaller the rolling force, the smaller the corresponding SHAP value, which shows that reducing the rolling force can reduce the rolling mill vibration.

The variation law between the SHAP value and input features is shown in Fig. 13. As shown in Fig. 13a, when the entrance thickness increases to 0.0205 m, the SHAP value decreases rapidly. Therefore, a reasonable entrance thickness is helpful to reduce the rolling mill vibration in the process of formulating the rolling schedule. As shown in Fig. 13b, d, e, when the rolling speed, outlet thickness, and back tension are small, the corresponding SHAP value remains around 0. Currently, the features do not affect the prediction result of the model. With the increase of feature value, the SHAP value begins to fluctuate greatly, which shows that the prediction result of the model fluctuates greatly and affects the stability of rolling mill operation.

The local interpretation aims to interpret the predictions of each individual sample. In this paper, two samples are

selected for local interpretation of the BO-XGBoost model. The first sample in the dataset is illustrated in Fig. 14a and the 621st sample in the dataset is shown in Fig. 14b. As shown in Fig. 14, the red arrow indicates the positive shake value and feature, which increases the predicted value of the model, and the blue arrow indicates the negative shake value and feature, which decreases the predicted value of the model. As can be seen from Fig. 14a, the predicted value of rolling mill vibration of the first sample is  $0.779 \text{ m/s}^2$ . The SHAP values of back tension, rolling speed, outlet thickness, and entrance thickness are positive, which is the feature of improving the rolling mill vibration, while the rolling force is the feature of reducing the rolling mill vibration. As shown in Fig. 14b, the predicted value of rolling mill vibration of the 621st sample is  $0.653 \text{ m/s}^2$ . At the same time, the SHAP values of rolling force, rolling speed, outlet thickness, and back tension are positive, which is the feature of improving the rolling mill vibration, while the entrance thickness is the feature of reducing the rolling mill vibration. Moreover, the specific SHAP values of different features are provided by Fig. 15.

## Conclusion

To solve the problem of mismatching degree between the process parameters and the operation state of strip rolling mill, one prediction model was proposed and the conclusions of this paper were summarized as follows.

- (1) The complicated relationship between process parameters and rolling mill vibration was accurately established by the BO-XGBoost prediction model.

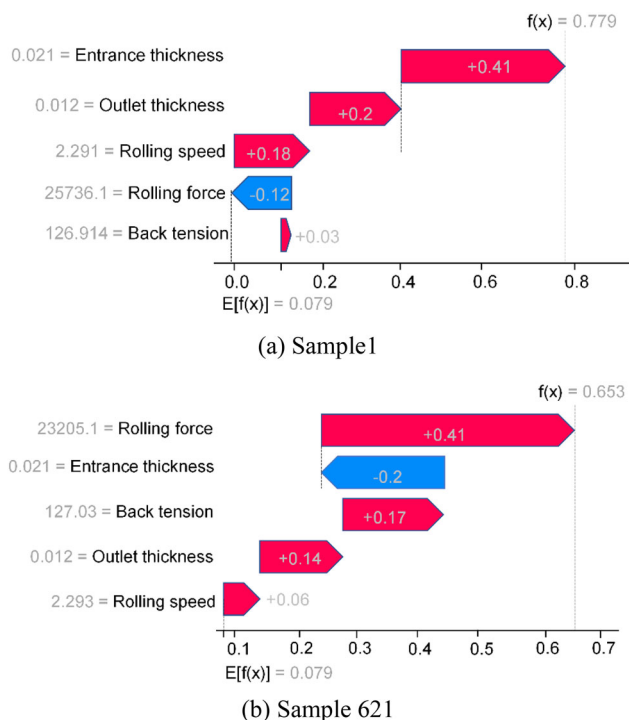


Fig. 15 SHAP feature waterfall plot

- (2) Compared with GS and RS, the prediction model optimized by BO algorithm with higher prediction accuracy, faster computational speed, and better stability.
- (3) The prediction results of the model were explained from a global perspective by introducing the SHAP method. As a result of the interpretation, the entrance thickness contributes the most to the output of BO-XGBoost prediction model.
- (4) Based on the collected data, the rolling speed of stand 2 rolling mill should not be greater than 2.3 m/s, the outlet thickness should not be greater than 0.01120 m and the rolling force should not be greater than 15000 KN to suppress the vibration of hot tandem strip rolling mill system.

By introducing Bayesian optimization algorithm and SHAP method, the problems of slow calculation speed, low prediction accuracy, and poor stability of the model were solved. At the same time, the proposed model will filling the technical gap of interpretable machine learning model in the field of rolling mill vibration prediction.

**Funding** The authors are grateful for the supports of the National Natural Science Foundation of China (Grant No. 51905365) and the grant from Shanxi Province Science and Technology Major Projects (No. 20181102015).

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

1. Thusty J, Chandra G, Critchley S, Paton D (1982) Chatter in cold rolling. *CIRP Ann* 31(1):195–199. [https://doi.org/10.1016/S0007-8506\(07\)63296-X](https://doi.org/10.1016/S0007-8506(07)63296-X)
2. Paton DL, Critchley S (1985) Tandem mill vibration: its cause and control. *Iron and Steel Making* 12(3):37–43
3. Yun IS, Wilson WRD, Ehmann KF (1998) Chatter in the strip rolling process. *J Manuf Sci Eng* 120(5):330–348. <https://doi.org/10.1115/1.2830132>
4. Sun ZH, Lu WL (2013) Single analysis of rolling mill vibration based on morphological undecimated wavelets and s-transform. *J Univ Sci Technol Beijing* 35(3):366–370
5. Ling QH, Yan XQ, Zhang YH (2016) Vibration feature extraction of hot continuous rolling based on s-transform. *J Vib Measure Diagn* 36(1):115–119+201–202
6. Yan XQ (2011) Machinery-electric-hydraulic coupling vibration control of hot continuous rolling mills. *J Mech Eng* 47(17):61–65
7. Yang JM, Zhang Q, Che HJ, Han XY (2010) Multi-objective optimization for tandem cold rolling schedule. *J Iron Steel Res Int* 17(11):39. [https://doi.org/10.1016/S1006-706X\(10\)60167-7](https://doi.org/10.1016/S1006-706X(10)60167-7)
8. Gao ZY, Zang Y, Zeng LQ (2015) Review of modeling and theoretical studies on chatter in the rolling mills. *J Mech Eng* 51(16):87–105
9. Bagheripoor M, Bisadi H (2013) Application of artificial neural networks for the prediction of roll force and roll torque in hot strip rolling process. *Appl Math Model* 37(7):4593–4607. <https://doi.org/10.1016/j.apm.2012.09.070>
10. Ma L, Dong J, Peng KX, Zhang K (2017) A novel data-based quality-related fault diagnosis scheme for fault detection and root cause diagnosis with application to hot strip mill process. *Control Eng Pract* 67:43–51. <https://doi.org/10.1016/j.conengprac.2017.07.005>
11. Liu Y, Gao ZY, Zhou XM, Zhang QD (2020) LSTM intelligent prediction of cold rolling chatter of thin plate driven by industrial data. *J Mech Eng* 56(11):121–131
12. Lu X, Sun J, Song ZX, Zhang DH (2020) Prediction and analysis of cold rolling mill vibration based on a data-driven method. *Appl Soft Comput J*. <https://doi.org/10.1016/j.asoc.2020.106706>
13. Chen JL, Wan ZG, Pan J, Zi YY, Wang Y, Chen BQ, Sun HL, Yuan J, He ZG (2016) Customized maximal-overlap multiwavelet denoising with data-driven group threshold for condition monitoring of rolling mill drivetrain. *Mech Syst Signal Process* 68–69:44–67. <https://doi.org/10.1016/j.ymsp.2015.07.022>
14. Pan J, Chen JL, Zi YY, Yuan J, Chen BQ, He ZG (2016) Data-driven mono-component feature identification via modified nonlocal means and MEWT for mechanical drivetrain fault diagnosis. *Mech Syst Signal Process* 80:533–552. <https://doi.org/10.1016/j.ymsp.2016.05.013>

15. Dong ZK, Liang PW, Chen CC, Sun JL, Zhao JY, Lu ML (2020) Research on vibration prediction of hot rolled high strength steel sheet mill based on DBN algorithm. *Min Metallurg Eng* 40(04):135–144
16. Deng JF, Sun J, Peng W, Hu YH, Zhang DH (2019) Application of neural networks for predicting hot-rolled strip crown. *Appl Soft Comput J* 78:119–131. <https://doi.org/10.1016/j.asoc.2019.02.030>
17. Song K, Yan F, Ding T, Gao L, Lu SB (2020) A steel property optimization model based on the XGBoost algorithm and improved PSO. *Comput Mater Sci*. <https://doi.org/10.1016/j.commatsci.2019.109472>
18. Shi R, Xu XY, Li JM, Yan YQ (2021) Prediction and analysis of train arrival delay based on XGBoost and Bayesian optimization. *Appl Soft Comput*. <https://doi.org/10.1016/J.ASOC.2021.107538>
19. Zhou J, Qiu YG, Zhu SL, Armaghani DJ, Khandelwal M, Mohamad ET (2020) Estimation of the TBM advance rate under hard rock conditions using XGBoost and Bayesian optimization. *Underground Space* 6(5):506–515. <https://doi.org/10.1016/j.undsp.2020.05.008>
20. Liang WZ, Luo SZ, Zhao GY, Wu H (2020) Predicting hard rock pillar stability using GBDT, XGBoost, and LightGBM algorithms. *Mathematics*. <https://doi.org/10.3390/math8050765>
21. Mongan PG, Hinchey EP, O'Dowd NP, McCarthy CT (2021) Quality prediction of ultrasonically welded joints using a hybrid machine learning model. *J Manuf Process* 71:571–579. <https://doi.org/10.1016/J.JMAPRO.2021.09.044>
22. Liang RH, Liu WF, Ma M, Liu WN (2020) An efficient model for predicting the train-induced ground-borne vibration and uncertainty quantification based on Bayesian neural network. *J Sound Vib*. <https://doi.org/10.1016/J.JSV.2020.115908>
23. Zhang WH, Yu JQ, Zhao AJ, Zhou XW (2021) Predictive model of cooling load for ice storage air-conditioning system by using GBDT. *Energy Rep* 7:1588–1597. <https://doi.org/10.1016/J.EGYR.2021.03.017>
24. Wang T, Zhang KF, Thé J, Yu HS (2022) Accurate prediction of band gap of materials using stacking machine learning model. *Comput Mater Sci*. <https://doi.org/10.1016/J.COMMATSCI.2021.110899>
25. Qu LC, Lyu J, Li W, Ma DF, Fan HW (2021) Features injected recurrent neural networks for short-term traffic speed prediction. *Neurocomputing* 451:290–304. <https://doi.org/10.1016/J.NEUCOM.2021.03.054>
26. Molin RMHD, Gomes DSR, Rodrigues MS, Cocco MV, Santos CLD (2022) Efficient bootstrap stacking ensemble learning model applied to wind power generation forecasting. *Int J Electr Power Energy Syst*. <https://doi.org/10.1016/J.IJEPES.2021.107712>
27. Chen TQ, Guestrin C (2016) XGBoost: a scalable tree boosting system. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ACM, 785–794
28. Zhang ZF, Huang YM, Qin R, Ren WJ, Wen GR (2021) XGBoost-based on-line prediction of seam tensile strength for Al-Li alloy in laser welding: experiment study and modelling. *J Manuf Process* 64:30–44. <https://doi.org/10.1016/J.JMAPRO.2020.12.004>
29. Nguyen-Sy T, Wakim J, To Q-D, Nguyen TT (2020) Predicting the compressive strength of concrete from its compositions and age using the extreme gradient boosting method. *Constr Build Mater*. <https://doi.org/10.1016/j.conbuildmat.2020.119757>
30. Zhang ZY, Liu ZC, Wu DZ (2020) Prediction of melt pool temperature in directed energy deposition using machine learning. *Addit Manuf*. <https://doi.org/10.1016/j.addma.2020.101692>
31. Jim B, Bob P, Bernd E, Patrick F (2021) Bayesian optimization of comprehensive two-dimensional liquid chromatography separations. *J Chromatogr A*. <https://doi.org/10.1016/J.CHROMA.2021.462628>
32. Verwaeren J, Weeën PVD, Baets BD (2015) A search grid for parameter optimization as a byproduct of model sensitivity analysis. *Appl Math Comput* 261:8–27. <https://doi.org/10.1016/j.amc.2015.03.064>
33. Valarmathi R, Sheela T (2021) Heart disease prediction using hyper parameter optimization (HPO) tuning. *Biomed Signal Process Control*. <https://doi.org/10.1016/J.BSPC.2021.103033>
34. Rao CJ, Liu M, Goh M, Wen JH (2020) 2-stage modified random forest model for credit risk assessment of P2P network lending to “Three Rurals” borrowers. *Appl Soft Comput J*. <https://doi.org/10.1016/j.asoc.2020.106570>
35. Alexander L, Cagatay C, Bedir T (2021) Analyzing the effectiveness of semi-supervised learning approaches for opinion spam classification. *Appl Soft Comput J*. <https://doi.org/10.1016/J.ASOC.2020.107023>
36. Bach D, Makoto O (2021) A random search for discrete robust design optimization of linear-elastic steel frames under interval parametric uncertainty. *Comput Struct*. <https://doi.org/10.1016/J.COMPSTRUC.2021.106506>
37. Ahmad M, Ahmad Z (2018) Random search based efficient chaotic substitution box design for image encryption. *Int J Rough Sets Data Anal (IJRSDA)* 5(2):131–147. <https://doi.org/10.4018/IJRSDA.2018040107>
38. Betrò B (1991) Bayesian methods in global optimization. *J Global Optim* 1(1):1–14. <https://doi.org/10.1007/BF00120661>
39. Kouziokas GN (2020) SVM kernel based on particle swarm optimized vector and Bayesian optimized SVM in atmospheric particulate matter forecasting. *Appl Soft Comput J*. <https://doi.org/10.1016/j.asoc.2020.106410>
40. Lundberg SM, Lee S-I (2017) A unified approach to interpreting model predictions. In: *Proceedings of the 31st International Conference on Neural Information Processing Systems*, Red Hook, NY, USA, Dec. 2017, pp. 4768–4777

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