IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received December 2, 2020, accepted December 19, 2020, date of publication December 22, 2020, date of current version January 4, 2021.

Digital Object Identifier 10.1109/ACCESS.2020.3046685

A Novel GPR-Based Prediction Model for Strip Crown in Hot Rolling by Using the Improved Local Outlier Factor

YAN WU[1], XU LI[2], FENG LUAN[3], AND YAODONG HE[2]
[1]School of Metallurgy, Northeastern University, Shenyang 110819, China
[2]State Key Laboratory of Rolling and Automation, Northeastern University, Shenyang 110819, China
[3]School of Computer Science and Engineering, Northeastern University, Shenyang 110819, China

Corresponding authors: Xu Li (lixu@ral.neu.edu.cn) and Feng Luan (luanfeng@mail.neu.edu.cn)

This work was supported in part by the National Key Research and Development Program of China under Grant 2017YFB0304100, in part by the National Natural Science Foundation of China under Grant U20A20187, and in part by the Fundamental Research Funds for the Central Universities under Grant N180708009 and Grant N2007006.

ABSTRACT In the hot rolling process, the prediction of strip crown is the key factor to improve the flatness quality of the strip. However, the traditional prediction method can only provide prediction values, but does not quantitatively evaluate the prediction error and stability. While Gaussian process regression (GPR) provides full probability prediction and estimates the uncertainty in the prediction. Therefore, for the first time, GPR is applied to predict strip crown. Furthermore, considering the negative influence of unavoidable outliers in measurement data, this article proposes an improved local outlier factor (LOF) algorithm to calculate the weights. And a novel Weight-GPR based on improved LOF prediction model is established. The proposed model not only retains the effective information of outlier values, but also avoids the negative influence brought by outlier values. The prediction experiments based on the real world production line data show that the proposed model can be successfully applied to the prediction of the strip crown in hot rolling process. Also, the performance of the proposed model is compared with typical GPR, ANN and SVM, and the results demonstrate that the Weight-GPR based on the improved LOF model provides better prediction accuracy and stability.

INDEX TERMS Gaussian process regression (GPR), outlier detection, strip crown, hot rolling process, local outlier factor (LOF).

I. INTRODUCTION

Hot rolling process is to use a series of rolls to progressively thin a cast or semi-finished steels to a desired thickness products such as the strip and sheet steel [1]. In deforming the strip, the rolling process causes variations in thickness across the width of the strip. These variations are commonly referred to as crown which is an important attributes of strip quality and has been paid more and more attention [2]. A good strip shape quality is that the strip has a desired crown and flatness, and it is also an important factor to determine the competitiveness of strip in the market. Therefore, it is an urgent problem to precisely predict the crown in hot rolling process for the early intervention of product quality [3], [4]. The traditional strip crown prediction method is a mechanism model established on the basis of some necessary simplifications and assumptions according to the mathematical and physical characteristics of the rolling process. For example, to obtain strip crown, a finite difference method [5] was introduced to imitate the roll temperature field and thermal crown. Pour *et al.* adopted a numerical model to predict many factors in the process of hot rolling like the rolling force and work roll wearing crown [6]. With the continuous improvement of computational ability, finite element method (FEM) has become a popular technology because it can simulate various rolling processes well, and has been applied to the analysis of symmetrical flatness actuator efficiency for rolling mill [7]. Although these traditional mechanism models have been widely used, they often behave poorly when confronted with the problems of strong coupling and non-linearity among parameters. Moreover, the mechanism models also have the shortcomings of slow improvement

The associate editor coordinating the review of this manuscript and approving it for publication was Eunil Park.

speed and high computational complexity. At the same time, in actual production, the mechanism models will also be limited by production conditions. Such as aging machines, intricate operation system, some parameters are difficult to measure and collect during rolling, changeable temperature or humidity and so on, which makes the mechanism models show very low efficiency and limitation in strip crown prediction [5]. With the development of big data and artificial intelligence (AI) technology, more and more prediction methods have been applied and studied in hot rolling process. The first AI method to be applied in hot rolling process is Artificial Neural Network (ANN), which uses a set of data to learn the implicit relationship between the input and output variables, but does not involve the basic knowledge of rolling. A mixed GA-ANN model [8] was established to predict and minimize the shape of hot rolled strip. Alaei *et al.* [9] developed an online prediction of work roll thermal expansion with ANN model. Shardt *et al.* [10] used ANN to predict the shape or crown of the strip, and then analyzed the sensitivity of the influencing factors of shape change. Wang *et al.* [11] proposed an ANN method optimized by intelligent algorithm to predict the profile and flatness. Deng *et al.* [4] proposed an ANN method for predicting strip crown based on reasonable data processing strategy. Although ANN has been applied in strip profile prediction, it also has some disadvantages, such as slow convergence speed, long training time, large randomness of network structure selection, easy to fall into local minimum and so on. For better modeling the strip thickness in hot strip rolling process, Wang *et al.* [12] proposed a strip crown prediction model based on support vector regression (SVR), and optimized the model parameter by using the improved adaptive mutation particle swarm optimization (AMPSO).

However, there are more powerful and effective prediction methods, such as Gaussian process regression (GPR). Different from ANN and SVM method, which can only provide prediction values, GPR can also captures the model uncertainty. It uses the mean prediction interval width (MPIW) to measure the stability of the model, thus providing more information about future possibilities and changes, which is more useful in practice [13]–[15]. GPR has become one of the most powerful probabilistic predicting methods. Its inference can be derived strictly in the function space by evaluating or approximating the posterior process, and it provide full probabilistic prediction with an estimation of uncertainty in the prediction [16], [17]. Moreover, GPR owns its unique advantages. For example, compared with SVM, the kernel hyperparameters, like length scale and noise level, can be simply learned via evidence maximization in GPR. In SVM, it is still a big issue, and not convincingly solved yet. Consequently, it is free and flexible to choose proper kernel functions in GPR [15]. As for the comparison with ANN, the training process of GPR is more efficient in smaller dataset due to its none-parametric characters such as well-tuned smoothing [13], [15]. In fact, GPR has been extensively used in various field such as short-term wind power prediction [18], fuel consumption prediction [19], some kinds of prediction on lithium-ion batteries such as remaining useful life [20], and state of health [21]. In industrial production, due to the uncertainty of machine operation and the complexity of production environment, it is necessary to evaluate the uncertainty in the prediction results. So it is imperative to apply GPR model to predict strip crown. In this article, for the first time, GPR model is applied to the rolling field and to predict the strip crown.

In order to improve the prediction performance of GPR, a Weight-GPR model is proposed in this article. In hot rolling process, the actual experimental data are affected by some negative factors, such as the measurement deviation caused by the machine and the inherent uncertainty of the working environment caused by changes in temperature and humidity [22], thus inevitably generating outliers. Due to their high heteroscedasticity and other side effects, outliers are partially responsible for irregularities in the model, deviations in parameter settings and incorrect results [23], [24]. Therefore, in order to better restore the authenticity of the data and obtain reliable analysis results, outliers need to be processed. Although outliers are usually considered as errors or noises, they may actually carry some important information. Removing them may lead to the risk that the potential correlation and inherent distribution of the entire data set may be discarded and partially destroyed [25]. Therefore, in the proposed Weight-GPR model, not only all data points are retained, but also each data sample is assigned a weight value related to outliers to reduce the negative impact of outliers. And the weight value is calculated by an improved local outlier factor (LOF) algorithm. This density-based LOF can be used to give a detailed mathematical value to represent the outlier level of the observation point, instead of using a binary attribute tag to represent whether it is an outlier [26], [27]. At the same time, traditional LOF algorithm is improved by feature weighting technology in Euclidean space. This also can avoid the problem of ''curse of dimensionality'' in Euclidean space for outlier detection of high-dimensional data [28]–[30]. Finally, a novel Weight-GPR based on improved LOF is established and applied to strip crown prediction in hot rolling process.

On the prediction of strip crown, the shortcomings of the existing work can be summarized as follows: 1) The traditional mechanism model has the disadvantages of difficult parameter acquisition and high computational complexity. 2) New data-driven machine learning algorithms, such as ANN, have the disadvantages of large randomness in network structure selection and easy to fall into local minimum. In this study, GPR model is applied to strip crown prediction for the first time and improved. Its significance lies in: 1) GPR can capture the uncertainty of the model, thus measuring the stability of the model, which is more useful in practice. 2) The improved GPR model adopts an improved LOF algorithm to analyze and process the outliers which eliminates the negative influence of the outliers while retaining the important information carried by the outliers. This article is organized as follows: Section 2 introduces the hot rolling process and description of the experimental data

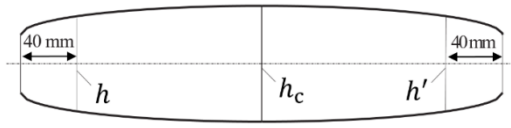

FIGURE 1. Sketch of strip profile [4].

in brief. Section 3 introduces some necessary information about the GP theory and the typical GPR model, then establishment and principle of the proposed Weight-GPR model is illustrated in detail. Furthermore, the weight calculating method based on the improved LOF is provided in Section 4. In Section 5, the prediction performance of the Weight-GPR based on improved LOF model is evaluated and compared with typical GPR, ANN and SVM models. Section 6 shows the conclusion.

II. HOT ROLLING TECHNOLOGY AND STRIP CROWN

A. HOT ROLLING TECHNOLOGY

Strip Crown is a vital performance index of strip flatness quality, which is mainly determined by the 3D deformation in the deformation zone. Crown is defined as the difference of thickness between the center and a point 40 mm from the edge of the strip, and this is schematically shown in Figure 1 [4]. For most applications steel mill clients require a strip with little crown or little variation in thickness. The mathematical formula of strip crown is shown as below:

$$C_{40} = h_c - \frac{h + h'}{2} \tag{1}$$

where C_{40} stands for the strip crown at the edge 40mm, h_c is the thickness at the center of the strip, h and h' respectively means the thickness at 40mm on the left and right sides.

B. STRIP CROWN INDUSTRIAL DATA

Figure 2 shows the complete rolling process in a typical hot rolling workshop. The process consists of 6 key parts: the reheating furnace, the roughing mill, the hot coil box and flying shear, the finishing mill, the laminar cooling, and the coiler. The finishing mill rolling is consists of 7 groups of stands. Each group stand is composed of a pair of working rolls and a pair of larger supporting rolls. A single batch consists of a coil of rough steel, which enters the reheating furnace to be reheated to the appropriate temperature. Next, the strip passes through the roughing mill, where its thickness and width are reduced to close to the desired value. Then, the strip enters the finishing mill section, where the strip is carefully milled to the required width and thickness. The profile of the strip can be controlled by changing the bending forces between the two work rolls [31]. After the last stand, the thickness distribution, profile and flatness of strip and the parameters of the rolling process are measured by corresponding equipment and sensors. Next, the strip is cooled by water to an appropriate final temperature. Finally, the strip is coiled and is ready for shipment.

There are many factors that affect the strip crown, which are mainly related to the roller, strip and rolling conditions in the rolling process. According to the measured parameters in the rolling process, nine important attributes are selected as input variables for strip crown prediction. They are Cooling water flow of rolling mill (%), Entrance temperature (°C), Exit temperature (°C), Strip width (m), Entrance thickness (mm), Exit thickness (mm), Bending force (kN), Rolling force (kN) and Entry profile (μm). Among them, the Cooling water flow of rolling mill can reflect the roll state. The Width, Thickness and Entry profile of the strip can reflect the state of the strip. Rolling force, Bending force and temperature can reflect rolling conditions [11], [12]. 474 samples of strip data is collected from a 1780mm hot rolling process production line in HBIS GROUP COMPANY, located in HeBei Province, China. Among them 67% (315) are used for training and establishing the prediction models and 33% (159) are used for testing and validating the effectiveness of the models. Furthermore, the measurement data should be processed with z-score normalization to the same scale to reduce the impact of different magnitudes and dimensions. Figure 3 is a fractal dimension visualization diagram of experimental data, showing the distribution of 9 input variables. As shown in figure 3, the input data varies greatly in different dimensions. In order to avoid the increase of model error caused by the large difference of data in different dimensions, it is necessary to z-score normalize.

III. THE NOVEL WEIGHT-GPR MODEL

Training data is defined as a set S of n observations $S = \{(x_i, y_i)|i = 1, 2, 3, \cdots, n\}$, where the $x_i \in R^d$ denotes hot rolling industrial data with nine attributes, and y_i is the predict crown value. The training inputs and outputs are aggregated into matrixes $X = [x_1, x_2, \ldots, x_n]$ and $Y = [y_1, y_2, \ldots, y_n]$ respectively. The testing inputs and outputs are also aggregated into matrixes $X^* = [x_1^*, x_2^*, \ldots, x_n^*]$ and $Y^* = [y_1^*, y_2^*, \ldots, y_n^*]$.

A. TYPICAL GPR MODEL

Gaussian Process (GP) is a collection of random variables, any finite number of which have a joint Gaussian distribution, introducing a distribution over functions [32]. It is precisely defined by the mean function and covariance function, expressed as follow:

$$f(x) \sim GP(m(x), k(x, x')) \tag{2}$$

where $m(x) = E(f(x))$ and $k(x, x') = E[(f(x) - m(x))(f(x') - m(x'))]$. In GP, the mean function $m(x)$ encodes the central tendency and is always set to be zero. And the covariance function $k(x, x')$, also called the kernel function, describes the information about the shape and structure which we expect the function to have.

Based on the GP theory, the typical GPR model is always assumed that the relationship between inputs and outputs has the following mathematical form:

$$y_i = f(x_i) + \varepsilon \tag{3}$$

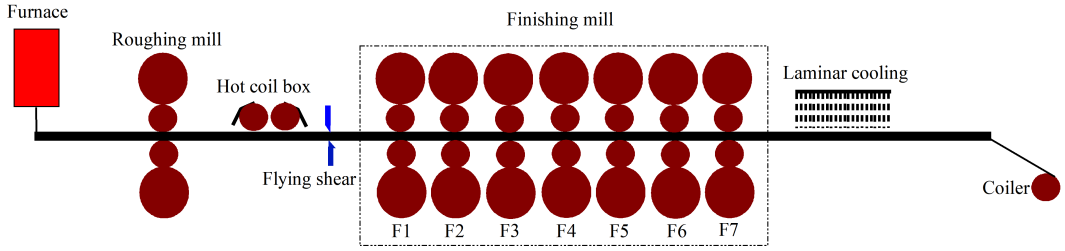

FIGURE 2. Schematic diagram of the hot rolling process.

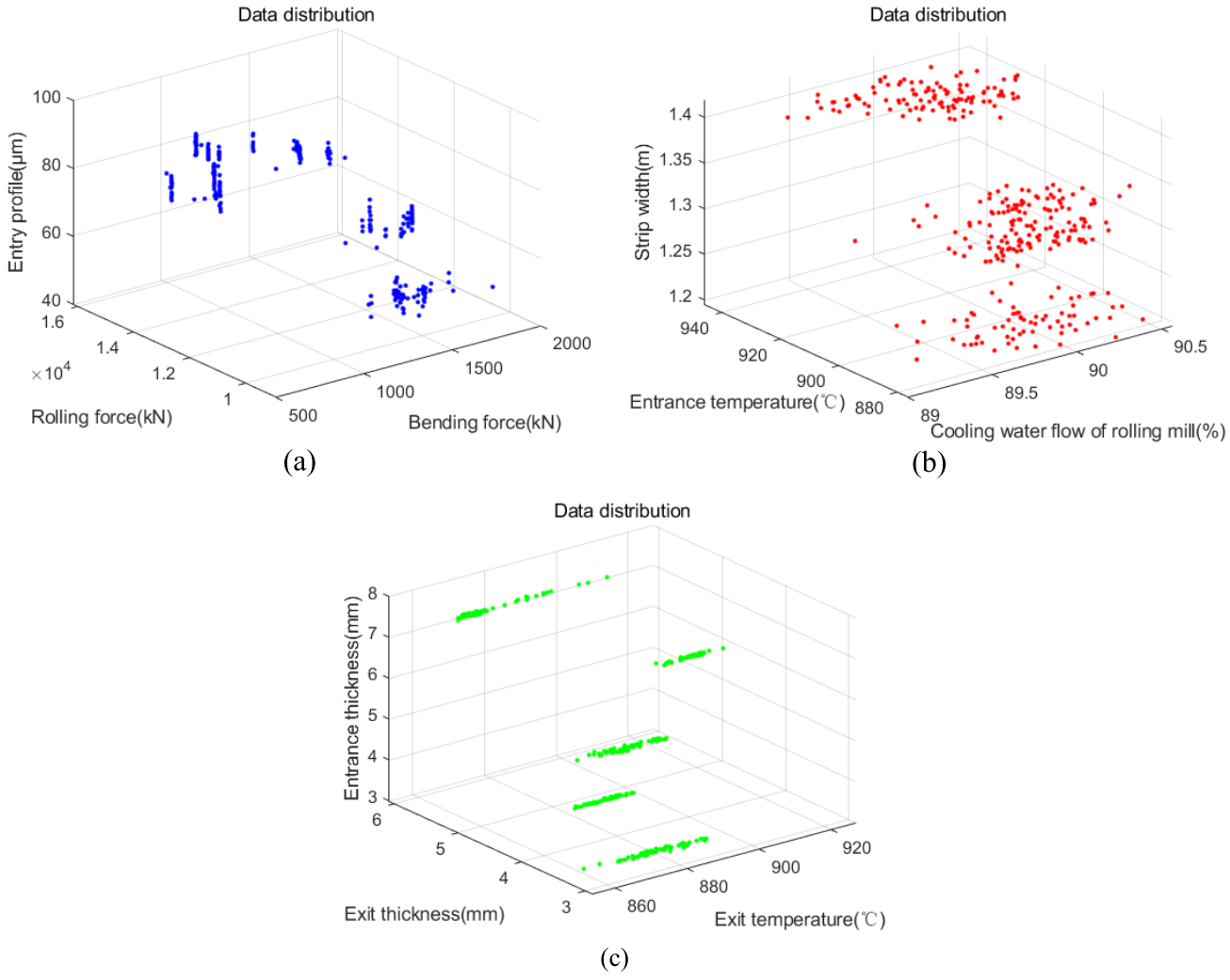

FIGURE 3. Fractal dimension visualization diagrams of different attributes. (a) Entry profile (μm), Rolling force (kN) and Bending force (kN). (b) Strip width(m), Entrance temperature (°C) and Cooling water flow of rolling mill (%). (c) Entrance thickness(mm), Exit thickness(mm) and Exit temperature (°C).

where the noise ε added to the $f(x)$ is regarded to follow an independent, identically Gaussian distribution, with zero mean and variance σ_n^2. Suppose that $F = [f(x_1), f(x_2), \ldots, f(x_n)]^T$ performs under the framework of GP, that is:

$$P(F|X) \sim N(0, K) \tag{4}$$

where K is the covariance matrix and $K_{ij} = k(x_i, x_j)$ according to GP theory [14]. The element K_{ij} is the covariance between values of the latent functions $f(x_i)$ and $f(x_j)$, which encodes about the prior knowledge of the nonlinear relation between them. The distribution of Y conditioned on F and X can be given as follow:

$$P(Y|F, X) = N(f, \sigma_n^2 I) \tag{5}$$

where I is the $N \times N$ identity matrix. The marginal distribution of Y is described in the following form:

$$P(Y|X) = \int P(Y|F,X)\,P(F|X)\,df = N(0, K + \sigma_n^2 I) \quad (6)$$

GPR is used to predict the distribution of the observed function values f^* at the test inputs X^*. The joint distribution over Y and f^* can be written as follow:

$$\begin{bmatrix} Y \\ f^* \end{bmatrix} \sim N(0, \begin{bmatrix} K(X,X) + \sigma_n^2 I & K(X,X^*) \\ K(X^*,X) & K(X^*,X^*) \end{bmatrix}) \quad (7)$$

By using the standard rules of typical GP, the prediction distribution of the observed function values is:

$$P\left(f^*|X,Y,X^*\right) \sim N(m, \sigma_s^2) \quad (8)$$

where $m = K(X^*,X)\left[K(X,X) + \sigma_n^2 I\right]^{-1} y$, $\sigma_s^2 = K(X^*,X^*) - K(X^*,X)\left[K(X,X) + \sigma_n^2 I\right]^{-1} K(X,X^*)$. Besides an expected prediction value (mean value) of the latent function f^*, the predicted output also includes a measurement of the uncertainty level according to the calculated variance.

B. ESTABLISHMENT OF THE WEIGHT-GPR MODEL

To make GPR model learn the potential correlation and inherent distribution of the dataset as much as possible, all of detected outliers were reserved instead of removing. However, the side effects, such as the model misspecification, biased parameter establishment and incorrect results, caused by outliers cannot be neglected. Therefore, each data sample including the norm ones is assigned with a specifically calculated weight value in [0], [1]. And in proposed Weight-GPR model, it is assumed that the relationship between inputs and outputs has a distinct mathematical form which differs from that of the typical GPR model. It has the following form:

$$w_i y_i = w_i f(x_i) + \varepsilon \quad (9)$$

$$f(x) = \emptyset(x)^T \beta \quad (10)$$

where $\emptyset(x)$ stands for the nonlinear function which maps the D-dimensional input vector x into an M dimensional feature space, β is a vector of weights (parameters) of the nonlinear model and the noise ε also follows an independent distributed Gaussian distribution: $\varepsilon \sim N(0, \sigma_n^2)$. Aggregate all the weight values into a matrix Ω, which is n-diagonal.

$$\Omega = \begin{bmatrix} w_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & w_n \end{bmatrix} \quad (11)$$

Based on the new proposed mathematical form and the GP theory, the establishment of the Weight-GPR model is derived and presented step by step in the following. Use the inference in Bayesian linear model for reference to predict the function values f^* [12], the conditional distribution can be presented by

$$f^*|x^*,X,Y \sim N(\frac{1}{\sigma_n^2}\emptyset\left(x^*\right)^T A^{-1}\emptyset(x)\,\Omega^2 Y, \emptyset\left(x^*\right)^T A^{-1}\emptyset\left(x^*\right)) \quad (12)$$

where $A = \sigma_n^{-2}\emptyset(x)\,\Omega^2\emptyset(x)^T + \sum_p^{-1}$.

Nonetheless, according to the equation (12), it can be found that if the size of matrix A is too large, it may cause great calculation complexity when the matrix is being inverted. Therefore, we try to rewrite the equation in the following way:

$$\begin{aligned} f^*|x^*,X,Y \sim N(&\emptyset\left(x^*\right)^T \sum\nolimits_p \left(\Omega^2 K + \sigma^2 I\right)^{-1} y, \emptyset\left(x^*\right)^T \\ &\times \sum\nolimits_p \emptyset\left(x^*\right) - \emptyset\left(x^*\right)^T \\ &\times \sum\nolimits_p \emptyset(x)\left(\Omega^2 K + \sigma^2 I\right)^{-1} \emptyset(x)^T \emptyset\left(x^*\right)) \end{aligned} \quad (13)$$

where by using the kernel method, we can define

$$K = \emptyset(x)^T \sum\nolimits_p \emptyset(x) \quad (14)$$

Consequently, the following formula is obtained:

$$\begin{aligned} \frac{1}{\sigma_n^2}\emptyset(x)(\Omega^2 K + \sigma_n^2 I) &= \frac{1}{\sigma_n^2}\emptyset(x)(\Omega^2\emptyset(x)^T \\ \sum\nolimits_p \emptyset(x) + \sigma_n^2 I) &= A\sum\nolimits_p \emptyset(x) \end{aligned} \quad (15)$$

Now multiplying by A^{-1} from left and $\left(\Omega^2 K + \sigma_n^2 I\right)^{-1}$ from the right meanwhile, it can get:

$$\frac{1}{\sigma_n^2}\emptyset(x)\,A^{-1} = \left(\Omega^2 K + \sigma_n^2 I\right)^{-1} \sum\nolimits_p \emptyset(x) \quad (16)$$

Finally, through the matrix inversion lemma we can simplify the mean and variance of posterior distribution of the predicted point respectively as it shows below. the mean function:

$$\frac{1}{\sigma_n^2}\emptyset\left(x^*\right)^T A^{-1}\emptyset(x)\,\Omega^2 Y = K\left(X^*,X\right)B^{-1}Y \quad (17)$$

the variance function:

$$\begin{aligned} \emptyset\left(x^*\right)^T A^{-1}\emptyset\left(x^*\right) &= K\left(X^*,X^*\right) \\ &-K\left(X^*,X\right)B^{-1}K(X,X^*) \end{aligned} \quad (18)$$

where $B = K(X,X) + \Omega^{-2}\sigma_n^2 I$

Above process is implemented in the view of the weight space, while there exists an alternative and even simpler way of obtaining the identical results by deriving the inference directly in the view of the function space.

The new proposed mathematical form $w_i y_i = w_i x_i + \varepsilon$ can be transformed into another form:

$$y_i = x_i + \frac{\varepsilon}{w_i} \quad (19)$$

Based on this transformed formula, we can obtain the condition distribution and marginal distribution respectively:

$$P(Y|F,X) = N(F, \Omega^{-2}\sigma_n^2 I) \quad (20)$$

$$P(Y|X) = \int P(Y|F,X)\,P(F|X)\,df = N(0, K + \sigma_n^2 I) \tag{21}$$

Consequently, it is inferred that the joint distribution over Y and f^* can be presented as:

$$\begin{bmatrix} Y \\ f^* \end{bmatrix} \sim N(0, \begin{bmatrix} K(X,X) + \Omega^{-2}\sigma_n^2 I & K(X,X^*) \\ K(X^*,X) & K(X^*,X^*) \end{bmatrix} \tag{22}$$

Using the standard rules for typical Gaussians process, the prediction distribution of the function values is:

$$P\left(f^*|X,y,X^*\right) \sim N(m, \sigma_s^2) \tag{23}$$

where $m = K(X^*,X)B^{-1}y$ and $\sigma_s^2 = K(X^*,X^*) - K(X^*,X)B^{-1}K(X,X^*)$.

C. COVARIANCE FUNCTION AND HYPERPARAMETER OPTIMIZATION

Despite GPR-based model is powerful due to its own advantages such as its nonparametric and flexibility, it is the specific kernel function that matters most in patterns and capacities of generalization [33]. Consequently, capturing a proper kernel function is key to achieve better interpretability and further extrapolation [34], [35].

In this article, instead of the traditionally used kernel function, the squared exponential covariance function, we adopt the rational quadratic function with automatic relevance determination distance measure. Compared with traditional covariance functions, it proves that rational quadratic functions are equal to the adding some various kinds of square exponential covariance functions with different length scale l. Thus, according to the theory, rational quadratic covariance function can fit the data smoothly across different length scales [36], which is defined below:

$$k\left(x,x'\right) = \sigma_f^2\left(1 + \frac{1}{2\alpha}\left(x-x'\right)^T M\left(x-x'\right)\right)^{-\alpha} \tag{24}$$

where $M = diag(l)^{-2}, l = [l_1, l_2, \ldots, l_D]^T$. With this explicit selected covariance function, the Weight-GPR model owns the hyperparameters including $\theta = (lnl_1, lnl_2, \ldots, lnl_D, ln\sigma_f, ln\alpha, ln\sigma_n)$. Hyperparameters optimization can be achieved by maximizing the log-marginal likelihood of θ. Given that $y|X \sim N(0, K + \sigma_n^2 I)$, the log-marginal likelihood is expressed as:

$$logP(y|X,\theta) = -\frac{1}{2}y^T B^{-1}y - \frac{1}{2}\log B - \frac{n}{2}log2\pi \tag{25}$$

and conjugate gradients method is applied to obtain the best hyperparameters.

IV. CALCULATE AND ASSIGN THE WEIGHT VALUES

A. OUTLIER DETECTION FOR HIGH-DIMENSIONAL DATA

Outlier detection can be mainly grouped into five categories: distribution-based, depth-based, distance-based, clustering-based and density-based outlier detection [28]. Among them, the outlier detection based on distribution, depth or distance adopts the overall criteria, which is not accurate for some special dataset [26]. Moreover, the deviation between different points is large enough to be considered, so the outlier detection algorithm based on density performs better than others [37]. Therefore, in this article, one of the most prevalent density-based LOF method was adopt, which is proposed by Breunig *et al.* [38] However, when the traditional LOF algorithm is applied in high-dimensional data, the negative impact caused by the problem ''curse of dimensionality'' in Euclidean space need to be considered [29]. And it means that the discrimination between the nearest and farthest neighbours is becoming weaker and weaker with the dimension increasing, which make the outlier detection fail to perform well as expected. To avoid the bad performance caused by the sparse high-dimensional data, the feature-weighting technique is introduced and applied in the Euclidean space of the LOF algorithm. Traditionally, when calculating the Euclidean distance, all data features, also called attributes, get the same treatment. The calculation formula is presented:

$$d\left(x_p, x_q\right) = \sqrt{\sum_i^D \left(x_p^i - x_q^i\right)^2} \tag{26}$$

However, in a real dataset there is always the possibility that different features may have different degrees of relevance, which should be taken into account through the feature-weighting method [37]. In this article, feature weights are obtained based on the Spearman correlation, which is a kind of typical nonlinear correlation measurement methods to measure the monotonic relationship between data. The formula is shown as follow:

$$d\left(x_p, x_q\right) = \sqrt{\sum_i^D \rho_i \left(x_p^i - x_q^i\right)^2} \tag{27}$$

$$\rho_i = 1 - (6\sum \frac{d_i^2}{n\left(n^2-1\right)}) \tag{28}$$

where ρ_i is the feature weights, d_i is the difference between ranks of two data and n stands for the number of data points.

B. IMPROVED LOF ALGORITHM

In this section, after applying the feature-weighting technique in the Euclidean space of LOF algorithm, some necessary steps were summarized in order to explain the principle of the proposed improved LOF algorithm.

Step 1: Measure the similarity of points

The Euclidean distance value improved by the feature weighting technique is used to measure the similarity of the points, and the expression is as follow:

$$d(p,o) = \sqrt{\sum_i^D \rho_i \left(p^i - o^i\right)^2} \tag{29}$$

where p and o stand for two different samples in the training inputs X.

Step 2: Calculate the k-distance of points

In this part, $d_k(p)$ denotes the k-distance of point p, and if it meets such two rules below:

In the set X, there exists at least k samples $o' \in X/\{p\}$, and $d\left(p, o'\right) \leq d(p, o)$

In the set X, no more than (k-1) samples $o' \in X/\{p\}$, and $d\left(p, o'\right) < d(p, o)$ then it holds that

$$d_k\left(p\right) = d(p, o) \tag{30}$$

where $k \in N^+$ is the number of neighbor observations of the observed point p

Step 3: Find the k-distance neighborhood of observed points

$N_k(p)$ denotes the k-distance neighborhood of observation p, which represents the specific points within k-distance. And these points are grouped into a new set Q.

Then the following formula is given:

$$N_k\left(p\right) = \{Q \in X/\{p\} | d\left(p, Q\right) \leq d_k\left(p\right)\} \tag{31}$$

Step 4: Calculate reach-distance of observed points

With respect to point o, the reach-distance of observed point p is denoted as $reach\_\_dist_k(p, o)$, and is calculated by:

$$reach\_\_dist_k(p, o) = max(d\left(p, o\right), d_k\left(p\right)) \tag{32}$$

Step 5: Calculate local reachability density of observed points

The local reachability density of observed point p is denoted as $lrd_k(p)$, given by:

$$lrd_k\left(p\right) = \frac{k}{\sum_{o \in N_k(p)} reach\_\_dist_k(p, o))} \tag{33}$$

Step 6: Calculate LOF values of observed points

The LOF of observed point p is given by:

$$LOF_k\left(p\right) = \frac{1}{k} \sum_{o \in N_k(p)} \frac{lrd_k\left(o\right)}{lrd_k\left(p\right)} \tag{34}$$

It can be obviously observed that the LOF values of the observed points are the ratio of the average local reachability density of neighborhoods to its local reachability density. And it is expected that, for the normal sample, the average local reachability density of neighborhoods is close to the local reachability density of itself, which means the corresponding LOF value approaches to 1. In contrast, if the observed point is an outlier, the LOF value is greater than 1, and that the higher outlier level is, the higher LOF value is, which means the LOF value can represent the outlier level. Moreover, the k value is used to adjust the number of the observed data samples. Moreover, we improved the performance of the LOF algorithm by adjusting the k value properly.

C. THE WEIGHTS OF WEIGHT-GPR MODEL

Theoretically, according to the principle of Weight-GPR model, the higher LOF value of the observed point is, the lower weight value is assigned. As for the normal point, the weight value is equal to 1. Ideally, only when the LOF value is very high, such as $+\infty$ in an extreme case, can the assigned weigh value be set to 0. Therefore, a mathematical transforming formula between LOF values and weight values is constructed after considering the real distribution of LOF values of our experimental dataset. And finally, we experientially establish a relationship as follows:

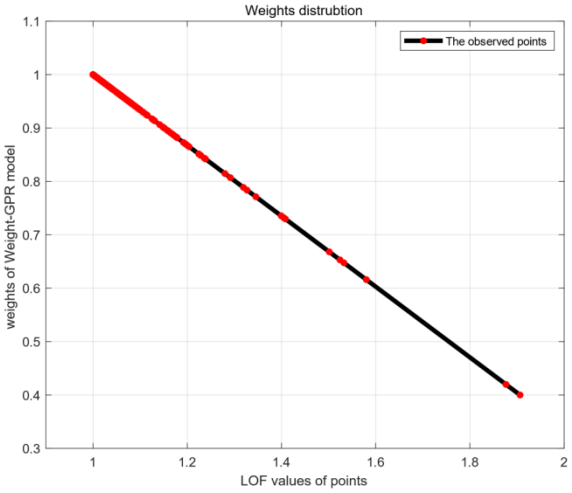

FIGURE 4. Weight values distribution based on LOF values.

TABLE 1. The detail information about the LOF values of the training data.

Outlier level	Low	Intermediate	High
LOF values range	[1.0, 1.1)	[1.1,1.2)	>1.3
Num of samples	272	23	20
Percentage (%)	86.35%	7.30%	6.35%

$$w_i = \frac{0.6}{1 - LOF_{max}} \times LOF_i + \frac{0.4 - LOF_{max}}{1 - LOF_{max}} \tag{35}$$

where the weight value of normal data is set as 1 and that of the particular data with the highest LOF value is set as 0.4. In addition, before training Weight-GPR model, the LOF values of the corresponding training data should be calculated first. The detailed distribution of their LOF values can be intuitively observed in figure 4 and table 1. To sum up, the flow chart of the strip crown prediction process is shown in figure 5.

V. EXPERIMENT RESULTS AND CASE ANALYSIS

A. PREDICTION PERFORMANCE EVALUATION

Four criteria are introduced to evaluate the performance of the prediction model. They are root mean square error (RMSE), mean absolute error (MAE), mean absolute percentage (MAPE) and coefficient of determination (R^2), which are defined respectively as follows:

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{N} \left(y_i - y_i^*\right)^2}{N}} \tag{36}$$

$$\mathrm{MAE} = \frac{1}{N} \sum_{i=1}^{N} |y_i - y_i^*| \tag{37}$$

$$\mathrm{MAPE} = \frac{1}{N} \sum_{i=1}^{N} \left| \frac{y_i - y_i^*}{y_i} \right| \times 100\% \tag{38}$$

$$\rho_i = 1 - (6 \sum \frac{d_i^2}{n\left(n^2 - 1\right)}) \tag{39}$$

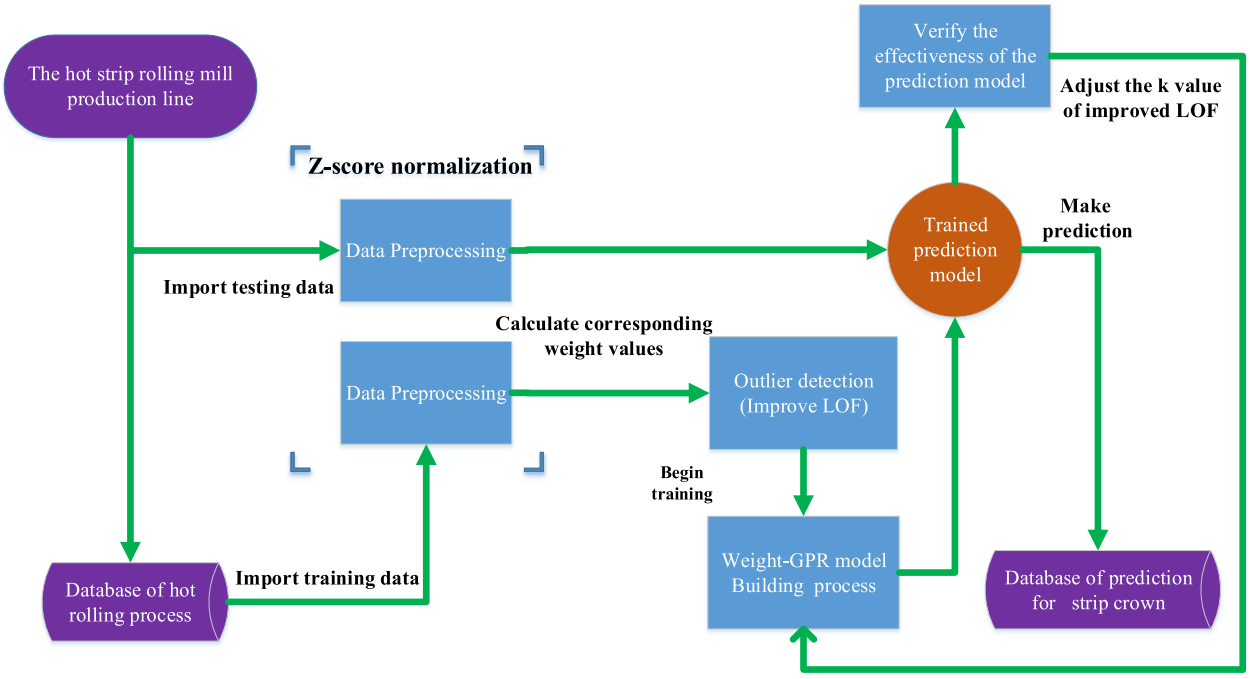

FIGURE 5. The flow chart of the strip crown model prediction process.

where y_i and y_i^* are the measured values and prediction values respectively, N is the total number of predicted data.

The GPR-based models can also provide more information about the possibility and changes in the future due to it can measure the stability of models [44], [45]. The stability of GPR-based models is evaluated by the mean prediction interval width (MPIW), as follows:

$$\mathrm{MPIW} = \frac{1}{n}\sum_{i=1}^{n}\left(U\left(y_i\right) - L\left(y_i\right)\right) \tag{40}$$

where $U(y_i)$ and $L(y_i)$ are the upper bound and lower bound of the predicted value, and the confidence of error interval is set to 95%.

B. EFFECTIVENESS OF THE IMPROVED LOF ALGORITHM

In this section, the effectiveness of the improved LOF algorithm was verified by comparing with the traditional LOF algorithm. After the measurement data being processed with z-score normalization, feature weights of the nine different attributes based on Spearman correlations are measured respectively. The results are shown in detail in figure 6. It can be observed that different attributes have different degrees of relevance of the target output. Take the Entry profile (μm) for example, its calculated feature weighting value is 0.955, which proves the greatest impact on the target output. Then three most important attributes are specifically selected and their space distribution scatter plots are given according to the improved LOF and traditional LOF, respectively, and the outlier degree is represented by the color scale, are shown in figure 7. The samples pointed by the red arrows are part of the identifiable outliers with high outlier level and should be assigned with rather lower weight values. In fact, as it actually shows, for the specific points, the weight values assigned by the improved LOF method are much lower than the weights assigned by the traditional LOF method, which can be found through the color scale. As for the black arrows, they are specifically point out to the circled regions where the samples that are supposed to be outliers in the view of the space distribution. However, they are not identified and detected by the traditional LOF method partly due to "curse of dimensionality" problem in Euclidean space. Therefore, it is obviously that the proposed LOF algorithm is not negatively affected by the "curse of dimensionality" problem in Euclidean space, and gives an effective outlier detection for our dataset.

Further, 20 particular samples are carefully selected in the regions pointed by the six arrows in figure 7. Since the weight values of these samples are much lower than others, they are grouped into a new testing dataset. After using the training dataset mentioned in section 2.2 to establish two different Weight-GPR models, which are based on the improved LOF and traditional LOF methods respectively, we try to use these two models to predict the corresponding target values of this new testing dataset and evaluate the prediction performance. The results are presented in table 2. And considering the small size of this new testing dataset, we only choose RMSE, MAPE, MAE as evaluation standards. The results in table 2 show it is more difficult to predict for the dataset with a large number of outlier values. However, the prediction performance of the Weight-GPR based on improved LOF is better than that of Weight-GPR based on traditional LOF.

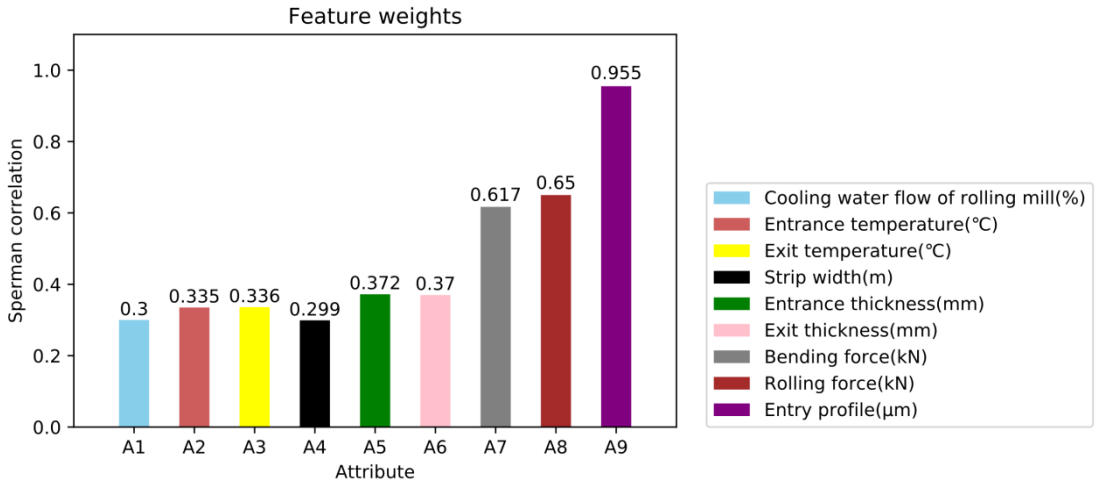

FIGURE 6. Feature weights of the nine different attributes with the Exit profile (μm).

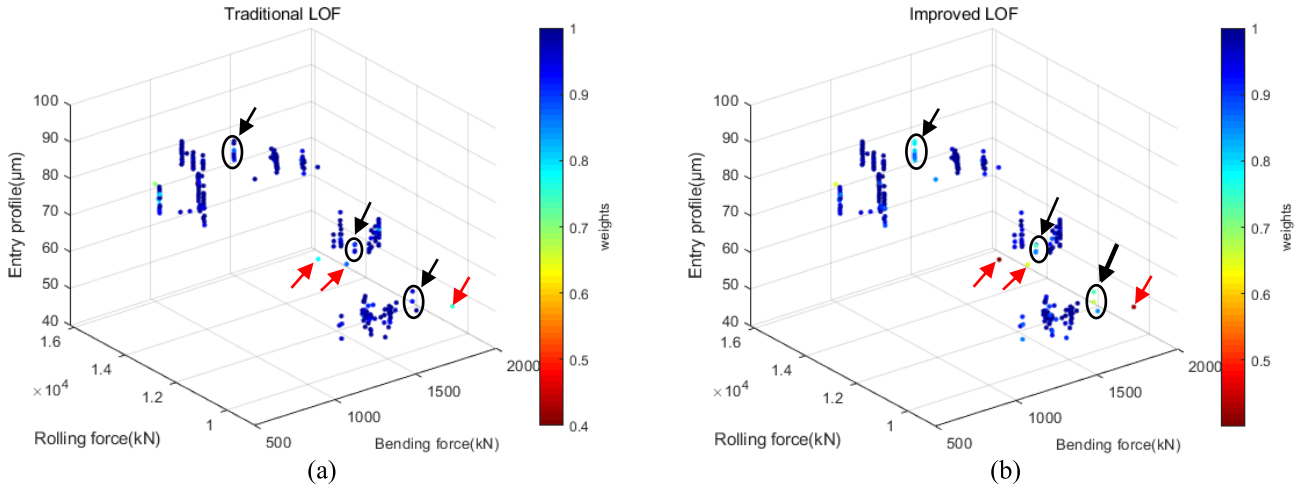

FIGURE 7. Outlier detection for high correlated attributes. (a) Traditional LOF method. (b) Improved LOF method.

To further highlight the effectiveness of the improved LOF algorithm, the two established Weight-GPR models above are applied to predict the original testing dataset (159 data samples). The prediction results are presented in figure 8. It can confirm that although the two models are in good agreement with the true data, the Weight-GPR based on improved LOF performs better in most outlier samples, as indicated by the arrow. And the results of evaluation criteria are shown in table 3. The results show that Weight-GPR based on the improved LOF shows better performance with the lower R^2 and MPIW, and the higher RMSE, MAE and MAPE (%). Comparing table 2 and table 3, it is found that Weight-GPR based on improved LOF has better prediction performance than Weight-GPR based on traditional LOF, both for the dataset with more outlier values and for the dataset with less outlier values.

C. COMPARISON CASE AND ANALYSIS

Using the same training data, the Weight-GPR model based on the improved LOF method, typical GPR (GPML toolbox), ANN (MATLAB neural network toolbox) and SVM (LIBSVM toolbox) are used for crown prediction. Additionally, the optimal parameters of these models are determined by ten-fold cross-validation. Mean Squared Error cost function is applied in ANN and SVM models. Also, the SVM model utilizes the popular radial basis kernel function. And its optimal parameters are obtained by the grid-search. For the ANN model with feed forward-backpropagation structure, the sigmoid activation function is used in a single hidden layer, and the weights and biases are updated according to the Levenberg-Marquardt optimization algorithm. After trying numbers ranging from 5 to 20, the number of hidden nodes is determined to be 15. In the case of the two GPR-based models, their hyperparameters are optimized by the gradient decent optimization algorithm. And it should be mentioned that in this optimization method, the Polack-Ribiere flavour of conjugate gradients is used to compute search directions, and a line search using quadratic and cubic polynomial approximations and the Wolfe-Powell stopping criteria is used together with the slope ratio method for guessing initial

TABLE 2. Prediction performance evaluations of Weight-GPR models based on improved and traditional LOF methods for the 20 particular samples with high outlier level.

	RMSE	MAE	MAPE (%)
Weight-GPR based on improved LOF	2.48	1.36	2.26
Weight-GPR based on traditional LOF	2.57	1.55	2.92

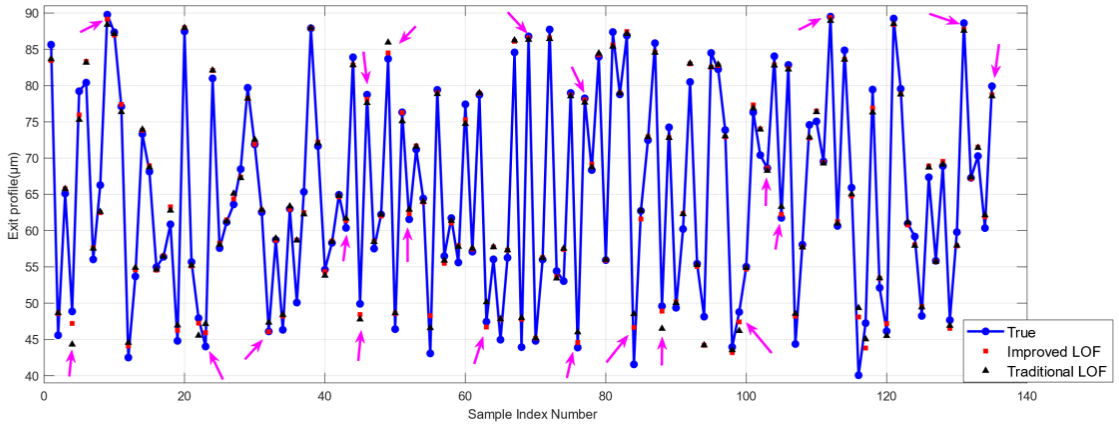

FIGURE 8. Prediction results of Weight-GPR models based on improved and traditional LOF for the testing dataset.

TABLE 3. Prediction performance evaluations of Weight-GPR models based on improved and traditional LOF methods for the testing dataset.

	R^2	RMSE	MAE	MAPE (%)	MPIW
Weight-GPR based on improved LOF	0.982	1.885	1.284	2.260	0.500
Weight-GPR based on traditional LOF	0.978	2.087	1.450	2.572	0.539

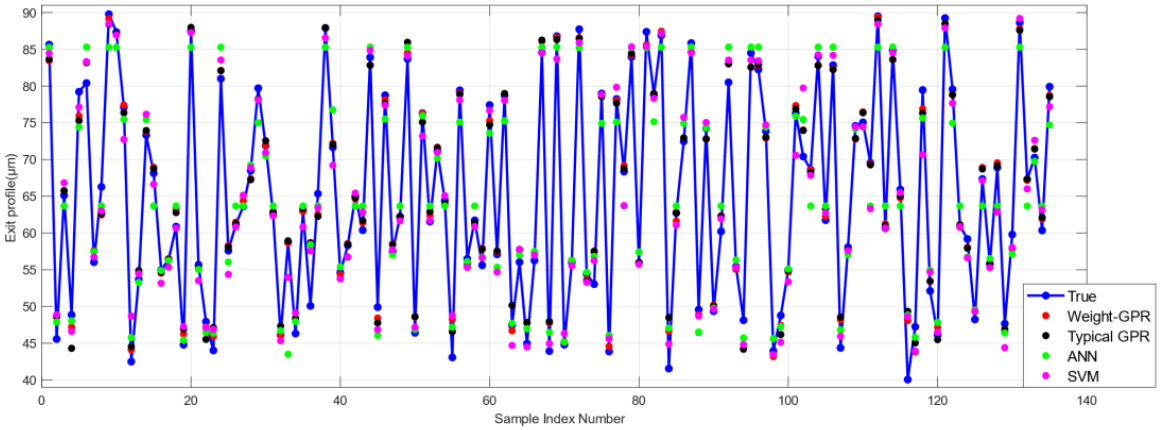

FIGURE 9. Prediction results comparison of different models.

step sizes [41], [42]. Additionally, a bunch of checks are made to make sure that exploration is taking place and that extrapolation will not be unboundedly large, along with applying Cholesky factorization to reduce the calculating complexity. As regards the selection of k value used in improved LOF method, having taken into the quantity of the training data, the k value ranging was adjusted from 25 to 35 continuously according to the performance of Weight-GPR model and chose 30 finally.

Figure 9 compares the strip crown prediction results of the four different models on the testing dataset. It shows that all the models can capture and predict the tendency of the true values, but the prediction values of Weight-GPR model are closer to the true values than others. The detailed prediction performance evaluation results are presented in table 4 and figure 10. Compared with the other three models, the proposed Weight-GPR model has the higher R^2 and the lower MAPE, RMSE and MAE, which means that the proposed

TABLE 4. Prediction performance evaluations of different models for the testing dataset.

	R^2	RMSE	MAE	MAPE (%)	MPIW
ANN	0.952	3.090	2.345	3.744	-
SVR	0.967	2.592	1.867	3.136	-
Typical GPR	0.971	2.397	1.875	3.079	0.596
Weight-GPR	0.982	1.885	1.284	2.260	0.500

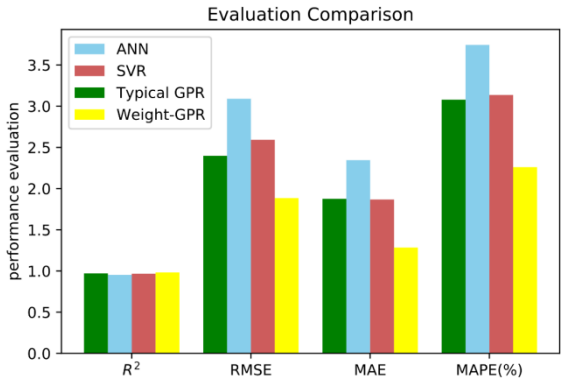

FIGURE 10. Evaluation histogram comparison of different models.

model can perform more accurate in the prediction of the strip crown. Moreover, in the view of MPIW value, it is lower than the typical GPR model, which indicates that the stability of the proposed model is better than that of typical GPR.

VI. CONCLUSION

In this article, GPR model is applied to strip crown prediction in order to quantitatively evaluate the prediction error and stability. Considering the existence of outlier values in the ture data of the factory, this article proposes a novel Weight-GPR based on improved LOF prediction model by combining feature-weighting technology with LOF algorithm. It also can resolve ''curse of dimensionality'' problem in Euclidean space caused by the high-dimensional industrial data. The proposed model not only retains the effective information of outlier values, but also avoids the negative influence brought by outlier values. Compared with Weight-GPR based on the traditional LOF, the proposed model shows the better prediction performance and stability. The prediction experiments based on the real world production line data show that the proposed model can be successfully applied to the prediction of the strip crown in hot rolling process. Also, the performance of the proposed model is compared with typical GPR, ANN and SVM. The results show that the proposed model has the higher R^2 and the lower MAPE, RMSE and MAE, which means that the proposed model can perform more accurate in the prediction of the strip crown. Moreover, in the view of MPIW value, it is lower than the typical GPR model, which indicates that the stability of the proposed model is better than that of typical GPR.

Although the GPR proposed in this article has paid attention to the quantification of prediction errors, it is still based on point prediction. If the training data is sparse, the target is multi-valued or affected by probability events, then the point prediction is not so reliable and accurate. Recent studies have proved that interval prediction is more reliable and informative than point prediction. Therefore, the application of interval prediction in rolling field is a topic ongoing research.

REFERENCES

[1] M. Nioi, S. Celotto, C. Pinna, E. Swart, and H. Ghadbeigi, ''Surface defect evolution in hot rolling of high-Si electrical steels,'' *J. Mater. Process. Technol.*, vol. 249, pp. 302–312, Nov. 2017.
[2] G.-S. Ma, Y.-M. Liu, W. Peng, F.-C. Yin, J.-G. Ding, D.-W. Zhao, H.-S. Di, and D.-H. Zhang, ''A new model for thermo-mechanical coupled analysis of hot rolling,'' *J. Brazilian Soc. Mech. Sci. Eng.*, vol. 39, no. 2, pp. 523–530, Feb. 2017.
[3] S. H. Zhang, B. N. Song, S. W. Gao, M. Guan, J. Zhou, X. D. Chen, and D. W. Zhao, ''Upper bound analysis of a shape-dependent criterion for closing central rectangular defects during hot rolling,'' *Appl. Math. Model.*, vol. 55, pp. 674–684, Mar. 2018.
[4] J. Deng, J. Sun, W. Peng, Y. Hu, and D. Zhang, ''Application of neural networks for predicting hot-rolled strip crown,'' *Appl. Soft Comput.*, vol. 78, pp. 119–131, May 2019.
[5] S.-X. Chen, W.-G. Li, and X.-H. Liu, ''Thermal crown model and shifting effect analysis of work roll in hot strip mills,'' *J. Iron Steel Res. Int.*, vol. 22, no. 9, pp. 777–784, Sep. 2015.
[6] H. S. S. Pour, H. K. Beheshti, Y. Alizadeh, and M. Poursina, ''Calculation of work roll initial crown based on desired strip profile in hot rolling,'' *Neural Comput. Appl.*, vol. 24, no. 5, pp. 1123–1133, Apr. 2014.
[7] Q.-L. Wang, J. Sun, Y.-M. Liu, P.-F. Wang, and D.-H. Zhang, ''Analysis of symmetrical flatness actuator efficiencies for UCM cold rolling mill by 3D elastic–plastic FEM,'' *Int. J. Adv. Manuf. Technol.*, vol. 92, nos. 1–4, pp. 1371–1389, Sep. 2017.
[8] S. John, S. Sikdar, P. K. Swamy, S. Das, and B. Maity, ''Hybrid neural–GA model to predict and minimise flatness value of hot rolled strips,'' *J. Mater. Process. Technol.*, vol. 195, nos. 1–3, pp. 314–320, Jan. 2008.
[9] H. Alaei, M. Salimi, and A. Nourani, ''Online prediction of work roll thermal expansion in a hot rolling process by a neural network,'' *Int. J. Adv. Manuf. Technol.*, vol. 85, nos. 5–8, pp. 1769–1777, Jul. 2016.
[10] Y. A. W. Shardt, S. Mehrkanoon, K. Zhang, X. Yang, J. Suykens, S. X. Ding, and K. Peng, ''Modelling the strip thickness in hot steel rolling mills using least-squares support vector machines,'' *Can. J. Chem. Eng.*, vol. 96, no. 1, pp. 171–178, Jan. 2018.
[11] Z. Wang, G. Ma, D. Gong, J. Sun, and D. Zhang, ''Application of mind evolutionary algorithm and artificial neural networks for prediction of profile and flatness in hot strip rolling process,'' *Neural Process. Lett.*, vol. 50, no. 3, pp. 2455–2479, Dec. 2019.
[12] Z.-H. Wang, Y.-M. Liu, D.-Y. Gong, and D.-H. Zhang, ''A new predictive model for strip crown in hot rolling by using the hybrid AMPSO-SVR-Based approach,'' *Steel Res. Int.*, vol. 89, no. 7, Jul. 2018, Art. no. 1800003.
[13] F. Perez-Cruz, S. Van Vaerenbergh, J. J. Murillo-Fuentes, M. Lazaro-Gredilla, and I. Santamaria, ''Gaussian processes for nonlinear signal processing: An overview of recent advances,'' *IEEE Signal Process. Mag.*, vol. 30, no. 4, pp. 40–50, Jul. 2013.
[14] A. Kamath, R. A. Vargas-Hernández, R. V. Krems, T. Carrington, and S. Manzhos, ''Neural networks vs Gaussian process regression for representing potential energy surfaces: A comparative study of fit quality and vibrational spectrum accuracy,'' *J. Chem. Phys.*, vol. 148, no. 24, Jun. 2018, Art. no. 241702.
[15] M. Seeger, ''Relationships between Gaussian processes, support vector machines and smoothing splines,'' Inst. Adapt. Neural Comput., Univ. Edinburgh, Edinburgh, U.K., Tech. Rep., 2000. [Online]. Available: https://link.springer.com/chapter/10.1007%2F3-540-36187-1_35

[16] J. Vanhatalo, J. Riihimki, J. Hartikainen, P. Jylnki, and A. Vehtari, "GPstuff: Bayesian modeling with Gaussian processes," *J. Mach. Learn. Res.*, vol. 14, no. 1, pp. 1175–1179, Apr. 2013.
[17] C. K. Williams and C. E. Rasmussen, *Gaussian Processes for Machine Learning*, vols. 2–3. Cambridge, MA, USA: MIT Press, 2006.
[18] C. Zhang, H. Wei, X. Zhao, T. Liu, and K. Zhang, "A Gaussian process regression based hybrid approach for short-term wind speed prediction," *Energy Convers. Manage.*, vol. 126, pp. 1084–1092, Oct. 2016.
[19] N. Huang, Y. Wu, G. Lu, W. Wang, and X. Cao, "Combined probability prediction of wind power considering the conflict of evaluation indicators," *IEEE Access*, vol. 7, pp. 174709–174724, 2019, doi: 10.1109/ACCESS.2019.2954699.
[20] Z. Hu, Y. Jin, Q. Hu, S. Sen, T. Zhou, and M. T. Osman, "Prediction of fuel consumption for enroute ship based on machine learning," *IEEE Access*, vol. 7, pp. 119497–119505, 2019, doi: 10.1109/ACCESS.2019.2933630.
[21] W. Kang, J. Xiao, M. Xiao, Y. Hu, H. Zhu, and J. Li, "Research on remaining useful life prognostics based on fuzzy evaluation-Gaussian process regression method," *IEEE Access*, vol. 8, pp. 71965–71973, 2020, doi: 10.1109/ACCESS.2020.2982223.
[22] L. Hao, Z. Jiang, D. Wei, D. Gong, X. Cheng, J. Zhao, S. Luo, and L. Jiang, "Experimental and numerical study on the effect of ZDDP films on sticking during hot rolling of ferritic stainless steel strip," *Metall. Mater. Trans. A*, vol. 47, no. 10, pp. 5195–5202, Oct. 2016.
[23] G. Williams, R. Baxter, H. He, S. Hawkins, and L. Gu, "A comparative study of RNN for outlier detection in data mining," in *Proc. IEEE Int. Conf. Data Mining*, Dec. 2002, pp. 709–712.
[24] H. Liu, S. Shah, and W. Jiang, "On-line outlier detection and data cleaning," *Comput. Chem. Eng.*, vol. 28, no. 9, pp. 1635–1647, Aug. 2004.
[25] I. Ben-Gal, "Outlier detection," in *Data Mining and Knowledge Discovery Handbook*. Boston, MA, USA: Springer, 2005, doi: 10.1007/0-387-25465-X_7.
[26] H. Ding, K. Ding, J. Zhang, Y. Wang, L. Gao, Y. Li, F. Chen, Z. Shao, and W. Lai, "Local outlier factor-based fault detection and evaluation of photovoltaic system," *Sol. Energy*, vol. 164, pp. 139–148, Apr. 2018.
[27] J. Liu and H. Deng, "Outlier detection on uncertain data based on local information," *Knowl.-Based Syst.*, vol. 51, pp. 60–71, Oct. 2013.
[28] C. C. Aggarwal and P. S. Yu, "Outlier detection for high dimensional data," in *Proc. ACM Sigmod Rec.*, vol. 2, 2001, pp. 37–46.
[29] A. Zimek, E. Schubert, and H.-P. Kriegel, "A survey on unsupervised outlier detection in high-dimensional numerical data," *Stat. Anal. Data Mining*, vol. 5, no. 5, pp. 363–387, Oct. 2012.
[30] C. C. Aggarwal, A. Hinneburg, and D. A. Keim, "On the surprising behavior of distance metrics in high dimensional space," in *Database Theory-ICDT* (Lecture Notes in Computer Science), vol. 1973. Berlin, Germany: Springer, 2001, doi: 10.1007/3-540-44503-X_27.
[31] K. Peng, H. Zhong, L. Zhao, K. Xue, and Y. Ji, "Strip shape modeling and its setup strategy in hot strip mill process," *Int. J. Adv. Manuf. Technol.*, vol. 72, nos. 5–8, pp. 589–605, May 2014.
[32] C. E. Rasmussen, "Gaussian processes in machine learning," in *Advanced Lectures on Machine Learning. ML* (Lecture Notes in Computer Science), vol. 3176. Berlin, Germany: Springer, 2003, doi: 10.1007/978-3-540-28650-9_4.
[33] S. Sun, G. Zhang, C. Wang, W. Zeng, J. Li, and R. Grosse, "Differentiable compositional kernel learning for Gaussian processes," 2018, *arXiv:1806.04326*. [Online]. Available: http://arxiv.org/abs/1806.04326
[34] D. Duvenaud, J. Robert Lloyd, R. Grosse, J. B. Tenenbaum, and Z. Ghahramani, "Structure discovery in nonparametric regression through compositional kernel search," 2013, *arXiv:1302.4922*. [Online]. Available: http://arxiv.org/abs/1302.4922
[35] A. Wilson and R. Adams, "Gaussian process kernels for pattern discovery and extrapolation," in *Proc. Int. Conf. Mach. Learn.*, 2013, pp. 1067–1075.
[36] R. K. Pandit and D. Infield, "Comparative analysis of Gaussian process power curve models based on different stationary covariance functions for the purpose of improving model accuracy," *Renew. Energy*, vol. 140, pp. 190–202, Sep. 2019.
[37] D. Pokrajac, A. Lazarevic, and L. J. Latecki, "Incremental local outlier detection for data streams," in *Proc. IEEE Symp. Comput. Intell. Data Mining*, Mar./Apr. 2007, pp. 504–515.
[38] M. M. Breunig, H.-P. Kriegel, R. T. Ng, and J. Sander, "LOF: Identifying density-based local outliers," in *Proc. ACM SIGMOD Int. Conf. Manage. Data (SIGMOD)*, 2000, pp. 93–104.
[39] A. Khosravi, S. Nahavandi, and D. Creighton, "Prediction interval construction and optimization for adaptive neurofuzzy inference systems," *IEEE Trans. Fuzzy Syst.*, vol. 19, no. 5, pp. 983–988, Oct. 2011.
[40] D. L. Shrestha and D. P. Solomatine, "Machine learning approaches for estimation of prediction interval for the model output," *Neural Netw.*, vol. 19, no. 2, pp. 225–235, Mar. 2006.
[41] M. R. Hestenes, *Conjugate Direction Methods in Optimization* (Optimization Techniques Part 1. Lecture Notes in Control and Information Sciences), vol. 6. Berlin, Germany: Springer, 1978.
[42] Y. H. Dai, "Nonlinear conjugate gradient methods," in *Wiley Encyclopedia of Operations Research and Management Science*. Atlanta, GA, USA: American Cancer Society, 2011, doi: 10.1002/9780470400531.eorms0183.

YAN WU received the B.S. degree in materials science and engineering, and the M.S. and Ph.D. degrees in metallurgical physical chemistry from Northeastern University, Shenyang, China, in 2002, 2005, and 2008, respectively. She currently serves as a Lecturer with the School of Metallurgy, Northeastern University. Her research interest includes the application of artificial intelligence in materials and metallurgy.

XU LI received the Ph.D. degree in material processing engineering from Northeastern University, China, in 2009. He is currently a Full Professor with the Sate Key Laboratory of Rolling and Automation, Northeastern University. His research interests include rolling process, fault diagnosis, automation systems, and intelligent control.

FENG LUAN was born in Liaoning, China, in 1979. He received the B.S. and M.S. degrees in biomedical engineering from Northeastern University, Shenyang, China, in 2002 and 2005, respectively, and the Ph.D. degree in electrical engineering and computer science from Seoul National University, Seoul, South Korea, in 2012. He currently serves as an Associate Professor with the School of Computer Science and Engineering, Northeastern University. He is the author/coauthor of 30 peer-reviewed articles on journals and conference proceedings. His research interests include signal processing and machine learning.

YAODONG HE received the B.S. degree from the School of Automation, Shenyang Aerospace University, Shenyang, China, in 2016. He is currently pursuing the master's degree with the School of Materials Science and Engineering, Northeastern University, Shenyang. His current research interests include rolling automation and machine learning.