

Surface Roughness Prediction Method of Titanium Alloy Milling based on CDH Platform

Liu Xianli (✉ xian.li.liu@163.com)

Harbin University of Science and Technology

Sun Yanming

Harbin University of Science and Technology

Yue Caixu

Harbin University of Science and Technology

Wei Xudong

Harbin University of Science and Technology

Sun Qingzhen

Harbin University of Science and Technology

Steven Y.Liang

George W. Woodruff School of Mechanical Engineering, Georgia Institute of Technology, Atlanta

Wang Lihui

KTH Royal Institute of Technology, Stockholm

Qin Yiyuan

Harbin University of Science and Technology

Research Article

Keywords: Titanium alloy milling, surface roughness prediction, CDH data platform, IHHO-SVM algorithm

Posted Date: June 25th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-634199/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Surface roughness prediction method of titanium alloy milling based on CDH platform

Xianli Liu^{1*}· Yanming Sun¹· Caixu Yue¹· Xudong Wei¹· Qingzhen Sun¹· Steven Y. Liang² and Lihui Wang³· Yiyuan Qin¹

¹Key Laboratory of Advanced Manufacturing and Intelligent Technology, Ministry of Education, Harbin University of Science and Technology, Harbin 150080, China

²George W. Woodruff School of Mechanical Engineering, Georgia Institute of Technology, Atlanta, 30332, USA

³KTH Royal Institute of Technology, Stockholm, 25175, Sweden

Corresponding author: Xianli Liu; E-mail: xian.li.liu@163.com

Abstract: Generally, off-line methods are used for surface roughness prediction of titanium alloy milling. However, studies show that these methods have poor prediction accuracy. In order to resolve this shortcoming, a prediction method based on Cloudera's Distribution Including Apache Hadoop (CDH) platform is proposed in the present study. In this regard, data analysis and process platform is designed based on the CDH, which can upload, calculate and store data in real-time. Then this platform is combined with the Harris hawk optimization (HHO) algorithm and pattern search strategy, and an improved hybrid optimization (IHHO) method is proposed accordingly. Then this method is applied to optimize the SVM algorithm and predict the surface roughness in the CDH platform. The obtained results show that the prediction accuracy of IHHO method reaches 95%, which is higher than the conventional methods of SVM, BAT-SVM, GWO-SVM and WOA-SVM.

Keywords: Titanium alloy milling; surface roughness prediction; CDH data platform; IHHO-SVM algorithm

Declarations

Funding

This research is supported by:

(1) National key R&D plan. Network collaborative manufacturing and smart factory special project: “Complex Tool Monitoring and Full Life Cycle Intelligent Management and Control Technology” under Grant No. 2019YFB1704800.

(2) International (regional) cooperation and exchange program of national Natural Science Foundation of China under Grant No. 51720105009.

(3) Outstanding Youth Fund of Heilongjiang Province (Grant Number YQ2019E029).

We have no conflict of interest to declare.

Conflicts of interest/Competing interests

The authors declare that there is no conflict of interests regarding the publication of this article.

Availability of data and material

The datasets used or analyzed during the current study are available from the corresponding author on reasonable request.

Code availability (Not applicable)

Authors' contributions

Xianli Liu, Caixu Yue, Qingzhen Sun and Xudong Wei contributed to the conception of this research. Yanming Sun designed and built CDH platform, and optimized the prediction algorithm. Yiyuan Qin participated in the debugging of the algorithm program; Steven Y. Liang and Lihui Wang gave many constructive suggestions on the revision of the thesis.

Ethics approval

The content studied in this article belongs to the field of metal processing, does not involve humans and animals. This article strictly follows the accepted principles of ethical and professional conduct.

Consent to participate

My co-authors and I would like to opt in to In Review.

Consent for publication

I agree with the Copyright Transfer Statement.

1 Introduction

Studies show that various data including tool type, machine property, workpiece characteristics and sensor information affect the cutting process. Updating these real-time data conforms to the characteristics of big data such as high speed, high diversity, and massive and low-value density. Therefore, it is of significant importance to establish an appropriate real-time data processing platform to integrate input, storage and calculation parts of cutting data. The surface roughness of the metal workpiece directly affects the wear resistance, matching stability, fatigue strength, corrosion resistance, sealing and contact stiffness [1]. Accordingly, accurate and real-time prediction of surface roughness of machined parts can provide a basis to change tool on time.

Currently, extensive investigations have been carried out on the data platform. In terms of composition and architecture of big data platform, Xiayi Gong et al. [2] considered different engineering problems and proposed a big data platform architecture consisting of resource, perception access, service, portal and application layer. To understand the real-time traffic status of the city, Zhang Tao [3] proposed a big data platform architecture that can take the data resource layer as the core and integrate the perception, application, presentation and user layer. Ramech et al. [4] demonstrated that the big data platform architecture should include the data source, data platform, data storage, platform presentation, data analysis and the mining layer. Through distributed file storage and computing, the above platforms greatly improve the efficiency of data storage and mining, data throughput and control of the whole cluster [5][6]. It is worth noting that the distributed storage and computing framework solves two major problems of transmission and process capacity expansion, thereby making big data processing feasible [7][8]. Currently, a close correlation has been established between big data technology and the mechanical field, thereby improving the development of mechanical automation [9][10].

Reviewing the literature indicates that big data architecture has been applied in numerous applications to solve practical problems. Guangming Xian [11] proposed a parallel machine learning algorithm based on the fine-grained pattern spark. The results showed that the parallel EMD-SVM algorithm under the big data spark cloud computing framework can improve accuracy, process speed, and learning convergence speed of intelligent fault classification. Jianglei Dong et al. [12] built a Hadoop-Spark big data framework, stored the collected experimental data in blocks in the form of data sets in HDFS for call and embedded the CSSAE algorithm model with PySpark to evaluate the milling cutter wear state. Furthermore, Min Tao et al. [13] adopted the distributed parallel processing framework of MapReduce to parallelize the signal feature extraction and modeling algorithm and predicted the tool wear driven by big data with high accuracy. Xuning Jiang et al. [14] used MapReduce distributed calculation to process the data and transformed the carbon footprint accounting model in the part processing into the MapReduce function to realize the rapid accounting of carbon footprint in the part processing. Xiongying Duan et al. [15] stored the collected data in the Hadoop distributed file system, processed the data by the Spark computing framework combined with the regression algorithm in the Spark ML machine learning library, trained the established model and finally verified the obtained results. Accordingly, they predicted the service life of the hydraulic actuator for a high-voltage circuit breaker.

Surface roughness refers to the small spacing and tiny peak-trough irregularities on the machined surface. The gradual rise of artificial intelligence provides a new idea to predict the surface roughness of parts. In this regard, machine learning and deep learning have been applied in numerous applications. Esche Aykut et al. [16] applied the artificial neural network (ANN) and response surface method (RSM), and improved the experimental model of the surface roughness. Then they optimized the single hidden layer of the surface roughness by using a 3-8-1 network structure. Longhua Xu et al. [17] proposed a new particle swarm optimization algorithm to train the fast convergence and global optimization ability of the

neural network and support vector machine models. The obtained results showed that compared with conventional intelligent models, this method has higher prediction accuracy and lower mean square error. Nan Xie [18] proposed a surface roughness prediction model based on energy consumption and used the particle swarm optimization support vector machine (PSO-SVM) to predict the turning surface roughness. The results show that the average relative error of this model is low.

It is worth noting that conventional methods mostly adopt off-line processing, which cannot produce timely and effective feedback to the processed parts. Accordingly, it is intended to design and build a data analysis and process platform based on the CDH, which makes full use of the internal structure of the platform to realize the analysis and calculation of real-time data. The present article is expected to provide a reference to simulate the cutting process.

In this article, a CDH platform is designed to integrate data input, storage and computing layer, Spark, Kafka, Flume, Hbase and other architectures, and analyze, calculate and storage data. Then the internal algorithm of the proposed platform is used to model the surface roughness of titanium alloy milling, and a hybrid optimization algorithm (IHHO) is proposed to optimize the SVM modeling. Finally, accuracy of the proposed algorithm and the superiority of the established platform over conventional methods are evaluated.

2 The overall design of the platform

2.1 Data characteristic analysis

Currently, the automation of manufacturing equipment in the field of cutting manufacturing has been preliminarily realized and the process information can be collected in real-time. In this regard, many data have been collected so far in the machining process: (1) Data related to the machine tool, tool accuracy, spindle vibration and so on. (2) Data related to measurement and defect of machining parts such as ablation depth, cutting breadth and so on. (3) The actual monitoring data of the production such as acoustic emission along X-, Y- and Z-directions, vibration variables and acceleration in the cutting process.

The real-time data transmitted by sensors and relevant information collection software have large data scales, diverse data types and fast production speed. In other words, the produced data by the cutting process has the characteristics of big data.

2.2. Requirements of the platform design

CDH platform is a core solution to analyze and process a large number of data generated in the cutting process. Through the deletion of irrelevant data and the correction of wrong data in the collected data, validity of the collected data can be ensured. Then data from different sources are integrated to uniform the data format and ensure the correspondence of data time. Accordingly, reasonable and reliable data resources can be provided for subsequent data analyses. Currently, most of the conventional platforms are built based on Apache's open-source framework Hadoop. Figure 1 presents the infrastructure of this framework.

It is worth noting that the platform should realize the whole process from data input and perform calculations to store data. Therefore, the platform should be designed from three different aspects, including data input, data storage and data analysis modules. Data input module is the platform to meet the high throughputs, which can make the key components withstand the sudden access pressure, and does not collapse due to the sudden overload request. On the other hand, the data storage module consists of a layer with the ability to provide automatic data backup to prevent the platform from normal operation originating from data loss or damage. Finally, the data analysis module is the platform to process the input data in real-time. Accordingly, the data process layer needs the ability of real-time processing.

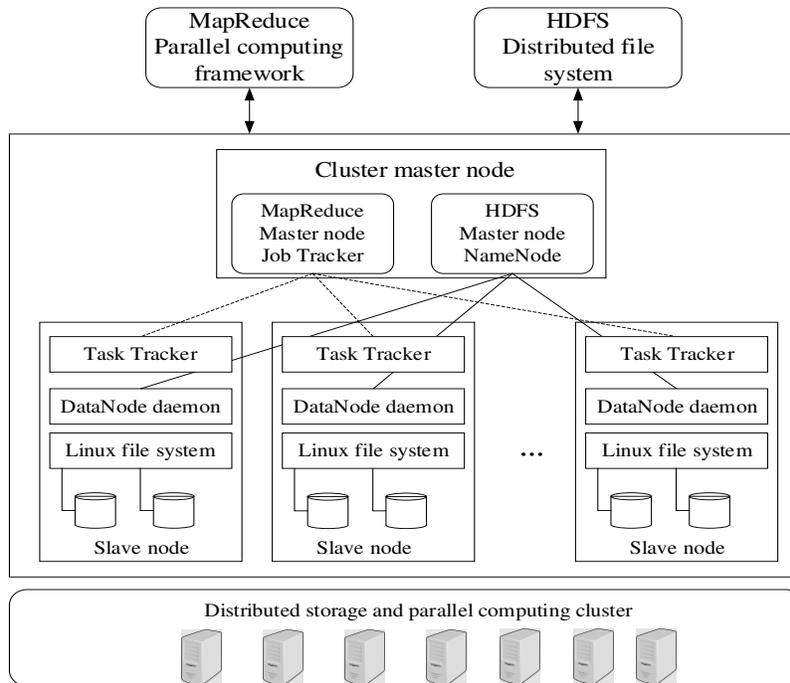


Figure 1: Hadoop infrastructure diagram

2.3. Overall architecture of the platform

According to the overall demand analysis of the platform, the data input, data process and data storage layers are designed respectively, and the appropriate big data architecture is selected.

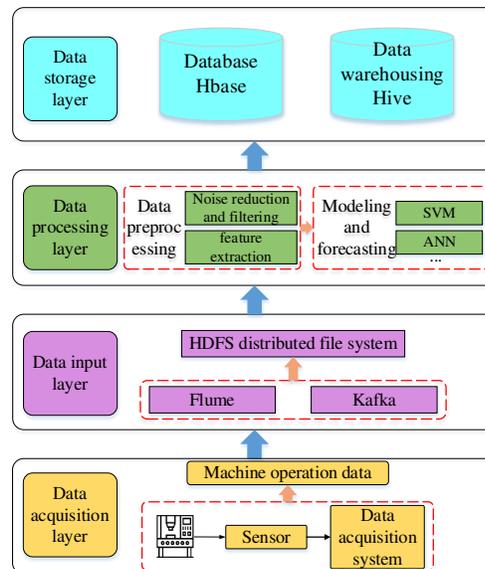


Figure 2: Platform architecture

To this end, Kafka is initially used to upload the collected machine running data to the platform. Then, HDFS and spark is used to store, process or analyze input data. Finally, the Hbase database is used to store and manage the real-time data, and the historical data is stored in the Hive data warehouse. Figure 2 illustrates the platform architecture.

2.4. Data input module

For the original data collection layer, the platform reads the data from either the real-time data file through the Flume and uploads it to Kafka to prepare for subsequent consumptions, or directly through the communication protocol (Object linking and embedding for Process Control , OPC) between Labview and Kafka. In the latter method, data storage in the hard disk is eliminated, thereby saving time. For the upper application system, the platform provides users with various algorithms for data processing and stores the processed results in either HDFS or Hbase.

2.5. Data storage module

The data collected during production and processing are stored in different databases through different acquisition software. It should be indicated that each data type can only reflect a part of the process or reaction so that the real processing state may involve numerous signals. Accordingly, it is of significant importance to improve the database storage structure and have a reasonable storage form for the collected signal data to facilitate the subsequent data call and processing. Hadoop distributed file system (HDFS) is data storage architecture based on Hadoop that applies the distributed principle to facilitate data expansion and fast storage. Meanwhile, Hbase and Hive can be applied as a non-relational database and data warehouse to store real-time data and historical data, respectively.

2.6. Data analysis module

The collected data during the production and processing is a group of large, continuous and time node collection sequences. Therefore, it should be collected in real-time through Labview software and apply the flow processing technology based on Hadoop for the processing. However, abnormal data may appear in the data collection and data transmission processes, thereby disturbing a correct reflection of the actual state of processing. Accordingly, it is necessary to use the screening mechanism of monitoring parameters to monitor possible errors of data in real-time. In this regard, different data processing algorithms have been proposed so far, among them the linked list search algorithm has the best data screening performance, while the moving translation algorithm has the highest data noise reduction.

3. Implementation of the CDH data platform

3.1. Network and software deployment

Generally, the proposed data platform is composed of hardware and software. In the hardware part, the resources of three workstations in the laboratory are combined to complete the deployment of the cluster and realize the integration of limited computing resources.

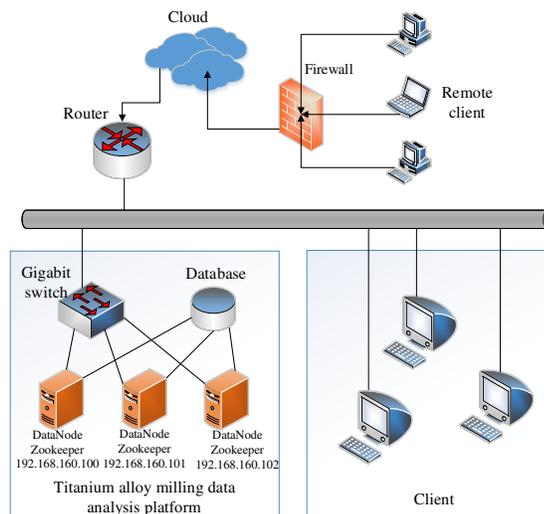


Figure 3: Network configuration of the cutting big data platform

Hadoop, Kafka and Hbase are deployed on the upper part of the cluster. The management and computing nodes are supported by a server (Dell PowerEdge R740XD with two gold medal 5218CPUs, 32 physical cores in total, 64 logical computing cores, 128G memory, 500GB solid-state, and four 10TB SAS hard disks). Moreover, the data computing node is undertaken by two servers (Dell PowerEdge R740 with two gold medals 5218 CPUs, 32 physical cores, 64 logical computing cores, 64G memory, 250GB solid-state). All nodes are connected through Gigabit Ethernet. Figure 3 illustrates the network configuration diagram.

From the software point of view, install virtual machine software (VMware) and Linux operating system are installed in each node. Then a multi-user and the corresponding user permissions are set. Generally, non-root users operate use the Hadoop platform. Accordingly, a SSH password-free login is set for an ordinary user to ensure the safe transmission of data and realize appropriate information sharing between nodes without secret. Finally, JDK is installed, the environment is configured and SecurtCRT is installed to realize remote control of the Linux operating system. Specifications of software are shown in Table 1.

Table 1: Software table of cutting big data platform

Software	Edition
CDH	CDH-6.3.2-1
Cloudera-manager	cloudera-manager-agent-6.3.1
centos	7.4
JDK	Jdk1.8.0_144
VMware-workstation	14.1.8
IDEA	2017

3.2. CM deployment

Cloudera Manager (CM) is the management platform of the CDH that can be deployed easily and operate the complete big data software stack centrally. Through the Cloudera Manager, the installation process can be automated, thereby reducing the required time to deploy the cluster. Studies reveal that the Cloudera Manager can provide a real-time running state view of nodes in the cluster and configure it. Moreover, Cloudera Manager can be applied to optimize the cluster performance and improve utilization by including a series of reporting and diagnostic tools.

When installing CM, it is necessary to deploy the server and agent. In this regard, the Hadoop 100 is set as the server, and Hadoop 101 and Hadoop 102 are set as the agent. It is worth noting that CM provides powerful Hadoop cluster deployment capability, and can automatically deploy nodes in batches. After installing the Cloudera Manager platform, the user can enter the CM management system login interface through the 7180 web port.

3.3. Integration and implementation of the CDH platform architecture

After deploying the CM, the management interface of CM can be accessed through the web port 192.168.160.100:7180. Then installation of big data architecture, including Kafka, HDFS, YARN, Hbase, Hive, Zookeeper and Spark can be completed. The cluster configuration is shown in Table 2. Furthermore, Figure 4 shows the CM management interface of the platform.

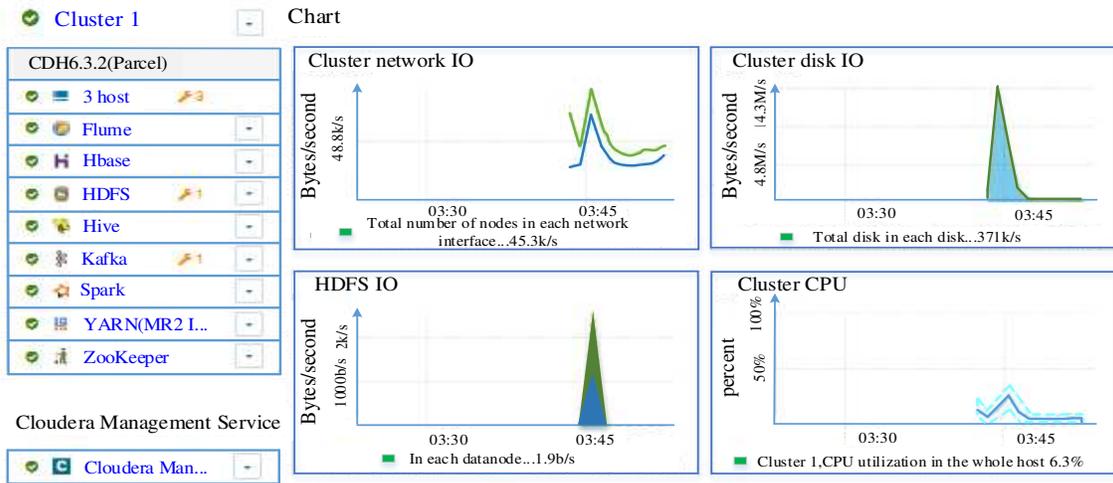


Figure 4: CM management interface of the platform

Table 2: Cluster configuration table

	Hadoop100	Hadoop101	Hadoop102
HDFS	NameNode	DataNode	DataNode
	DataNode		
Yarn	Resource Manager Node Manager	Node Manager	Node Manager
Zookeeper	Zookeeper	Zookeeper	Zookeeper
Kafka	Kafka	Kafka	Kafka
Hbase	Master	Region Server	Region Server
	Region Server		
Flume	Flume	Flume	Flume
Hive	Hive		
Spark	Spark		

4. Surface roughness prediction based on the CDH platform

The CDH platform has three stages, including data uploading, data processing and data storing to predict the surface roughness. In this regard, the Labview data acquisition system is utilized to save the collected data by machine tools to the host computer. Then the Kafka background monitoring program reads data files of the host computer. Once the data is transmitted, the program is immediately called to upload data to the HDFS distributed file storage system of the big data platform. Then the machine learning algorithm inside the platform is called to model, predict and process the uploaded data. Finally, the prediction data is saved to the Hbase database. Figure 5 shows the data processing flow of the established platform.

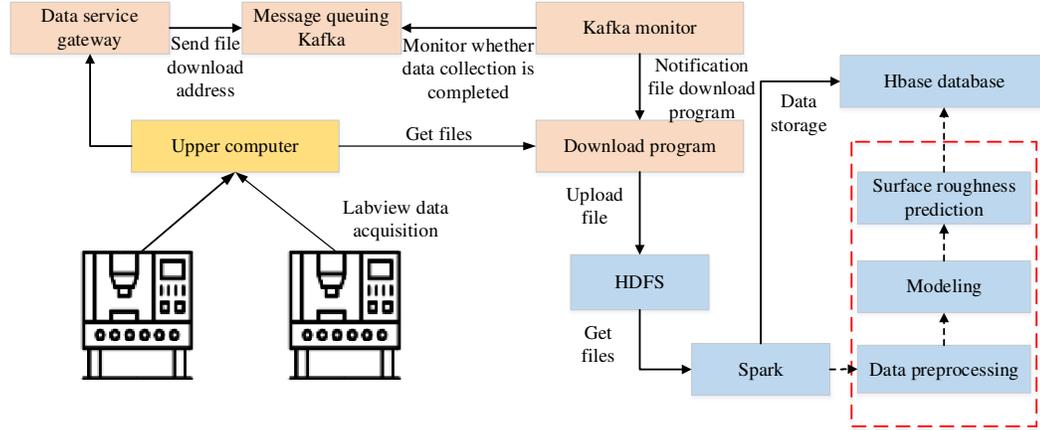


Figure 5: Data processing flow of the Platform

4.1. Data set construction

In order to predict the surface roughness and ensure diversity, authenticity and reliability of the data, the milling experiment of Ti-6Al-4V (TC4) titanium alloy was carried out. The test workpiece consists of a rectangular titanium alloy (TC4) block with a dimension of 100 mm * 50 mm * 50 mm. Then an alumina-coated finishing tool (ST210-R4-20030, Xiamen Jinlu Company) with an integral carbide vertical milling cutter was applied to achieve the desired surface finishing. The milling process was conducted in accordance with single tooth and down milling methods with a dry cutting mode. The experiment was carried out on a three-coordinate vertical CNC machining center (VDL-1000E, Dalian machine tool Company). The surface roughness R_a was measured by a surface roughness measuring instrument (TR-200). Figure 6 shows instruments and test equipment.



Figure 6: Ti-6Al-4V (TC4) milling test tool

During the experiment, surface roughness values were measured by adjusting four cutting parameters (f_z , v_c , a_e , a_p). In this regard, a total of 144 experiments were conducted. Each group of parameters was measured three times, and the average value was taken as the final measurement value. For the data collected in the experiment, the first 110 groups of data were set as the training set to train the model, and the last 34 groups of data were utilized to verify the model.

Table 3: Data parameters of the training set

NO	f_z (mm/z)	v_c (m/min)	a_e (mm)	a_p (mm)	R_a
1	0.06	60	0.5	12	0.337
2	0.08	60	0.5	12	0.47
3	0.1	60	0.5	12	0.628
4	0.12	60	0.5	12	0.696
5	0.06	60	1	12	0.377
...

107	0.1	60	1.5	8	0.711
108	0.12	60	1.5	8	0.791
109	0.06	80	0.5	8	0.283
110	0.08	80	0.5	8	0.381

Table 4: Validation set data parameter table

NO	$f_z(\text{mm/z})$	$v_c(\text{m/min})$	$a_e(\text{mm})$	$a_p(\text{mm})$	Ra
111	0.1	80	0.5	8	0.464
112	0.12	80	0.5	8	0.571
113	0.06	80	1	8	0.291
...
142	0.08	120	1.5	8	0.471
143	0.1	120	1.5	8	0.532
144	0.12	120	1.5	8	0.674

4.2. Data transmission module

The Kafka message queue is introduced into the cluster internal structure. The background monitor presents the message queue all the time. Once the message is detected and stored, the relevant information is immediately collected and the program is called to upload the signal file to the platform to facilitate the subsequent operation. When the data information is successfully uploaded to the platform, a message displays on the console, indicating that the data uploading is done successfully.

4.3. Data processing module

In the previous stage, the milling test data is uploaded to the platform, where the interior support vector machine (SVM) algorithm is used to model and predict the surface roughness. In order to get a more accurate prediction value of the surface roughness, the pattern search algorithm is combined with the Harris Hawk optimization algorithm (HHO) and a hybrid optimization algorithm (IHHO) is proposed. Then the proposed IHHO algorithm is applied to model the parameters of SVM and predict the surface roughness. Compared with BAT, GWO and WOA optimized SVM, and introduced indexes such as mean square error, mean absolute error and curve fitting degree R^2 , which proved the superiority of IHHO algorithm.

4.3.1. Harris hawk optimization algorithm

Harris Hawks Optimization (HHO) is a group intelligence optimization algorithm proposed by Heidari et al. [19] to simulate the predatory behavior of Harris hawks. Wenlong Fu et al. [20] combined the sine-cosine algorithm (SCA) and cyclic variation strategy and proposed an improved Harris-hawks optimization algorithm (HHO). Then, the MSCAHHO algorithm was applied to optimize parameters of SVM and finally, the optimized SVM model was used for fault classification. Reasonable diagnosis results have been achieved in this regard. Tayab et al. [21] applied the Harris-Hawks optimization algorithm to the feedforward neural network. The results showed that compared with the conventional artificial neural network based on the particle swarm optimization (PSO), the least square support vector machine based on the PSO and the neural network based on BP, the mean absolute error of this method is reduced by 33.30%, 49.54% and 60.76%, respectively.

The implementation process of the HHO algorithm can be mainly divided into the exploration stage, exploration and development transformation stage and the development stage [23].

In the exploration stage, Harris hawk randomly inhabits in some places, waits and detects its prey according to the following two strategies:

$$X(t+1) = \begin{cases} X_{rand}(t) - r_1 |X_{rand}(t) - 2r_2 X(t)| & q \geq 0.5 \\ (X_{rabbit}(t) - X_m(t)) - r_3 (LB + r_4 (UB - LB)) & q < 0.5 \end{cases} \quad (1)$$

Where $X(t+1)$ is the position of individuals in the next iteration, $X_{rand}(t)$ is a randomly selected position, $X_{rabbit}(t)$ is the position of individuals with the optimal fitness, and r_1, r_2, r_3, r_4, q are random numbers between $[0,1]$. Moreover, $X_m(t)$ is the average position as the following:

$$X_m(t) = \frac{1}{n} \sum_{i=1}^N X_i(t) \quad (2)$$

The transition stage from exploration to development: HHO algorithm converts different exploitative behaviors according to the escape energy of prey. The escape energy of prey can be expressed in the form below:

$$E = 2E_0 \left(1 - \frac{t}{T}\right) \quad (3)$$

Where E_0 is the initial energy and T is the maximum number of iterations. When the escape energy $|E| \geq 1$, the search is carried out. When $|E| < 1$, the development is carried out.

Development phase: In the HHO algorithm, four possible strategies are proposed to simulate the attack phase. When the parameter $|E| \geq 0.5$, the soft siege was carried out, while a hard siege was carried out for $|E| < 0.5$.

1. When $r \geq 0.5$ and $|E| \geq 0.5$, the soft siege strategy is as follows:

$$\begin{aligned} X(t+1) &= \Delta X(t) - E |J X_{rabbit}(t) - X(t)| \\ \Delta X(t) &= X_{rabbit}(t) - X(t) \end{aligned} \quad (4)$$

Where $\Delta X(t)$ is the difference between optimal individuals and the current individual, J is a random number between $[0,2]$ and r is a random number between $[0,1]$.

2. When $r \geq 0.5$ and $|E| < 0.5$, the strategy of hard siege can be expressed as follows:

$$X(t+1) = X_{rabbit}(t) - E |\Delta X(t)| \quad (5)$$

3. When $r < 0.5$ and $|E| > 0.5$, the siege strategy can be rewritten in the form below:

$$X(t+1) = \begin{cases} Y & F(Y) < F(X(t)) \\ Z & F(Z) < F(X(t)) \end{cases} \quad (6)$$

$$\begin{aligned} Y &= X_{rabbit}(t) - E |J X_{rabbit}(t) - X(t)| \\ Z &= Y + S \times \text{LF}(D) \end{aligned}$$

Where S is a D -dimensional random vector and LF is a Levy flight function.

4. When $r < 0.5$ and $|E| < 0.5$, the siege strategy is as follows:

$$X(t+1) = \begin{cases} Y & F(Y) < F(X(t)) \\ Z & F(Z) < F(X(t)) \end{cases} \quad (7)$$

$$\begin{aligned} Y &= X_{rabbit}(t) - E |J X_{rabbit}(t) - X_m(t)| \\ Z &= Y + S \times \text{LF}(D) \end{aligned}$$

4.3.2. Improved Harris hawk optimization algorithm (IHHO)

Pattern search method [22] (also known as the Hooke Jeeves algorithm) is a direct search algorithm that does not rely on derivatives. Studies show that this algorithm has a good performance at solving multi-dimensional optimization problems. Moreover, it has a simple structure, high accuracy, and strong local searching ability. The main purpose of this algorithm is to find the best direction of function value

decline to search and get the optimal solution.

The pattern search method is to search and move the current search point according to the fixed pattern and step size to find the feasible descent direction. The algorithm consists of two alternative searches, including axial and pattern search. The axial search is carried out along n-coordinate axes to determine a new base point and direction, in which the function value decreases. On the other hand, the pattern search is carried out along the continuous direction of two adjacent base points, trying to make the function value decline faster [23]. The optimization steps of the improved Harris hawk optimization algorithm (IHHO) are as follows:

(1) Population initialization. Each individual is initialized according to the upper and lower bounds of each dimension of the search space.

(2) Calculating the initial fitness. The position of the individual with the best fitness is set as the current prey position.

(3) Updating the location. The escape energy of prey is updated by equation (3-3), and then the corresponding position update strategy in search or development behavior is performed according to the escape energy and generated random number.

(4) Calculating the fitness. The individual fitness after position updating is calculated and compared with that of the prey. If the updated position is better than that of the prey, it is considered as the new prey position.

(5) Repeat step 3 and step 4, unless the number of iterations exceeds the maximum number of iterations. Then the current prey position is considered as the estimated position of the target.

(6) A new solution scheme is obtained by searching the pattern of the current optimal individual. Compared with the optimal solution of the current population, the better one is retained.

4.3.3. Construction of the surface roughness prediction model based on IHHO-SVM

Support Vector Machine (SVM) is an extended machine learning technology of the statistical learning theory. It has superior characteristics, including simple structure, good generalization ability and global optimality. Accordingly, SVM has been widely applied to identify the data classifier, where the result will output an optimal classification hyperplane. In this method, the given data should not be misclassified, that is, the nearest data distributed on both sides of the hyperplane should have the largest distance from the plane. The hyperplane can be mathematically expressed in the form below:

$$f(x)=\omega^T\varphi(x_i)+b \quad (8)$$

Where ω , $\varphi(x_i)$ and b denote the weight, mapping function and offset top, respectively.

When data is linearly divisible, the classification accuracy increases as the classified distance increase. In this regard, the following optimization problems should be solved to make the distance of classification larger:

$$\begin{aligned} \min & \frac{1}{2}\omega^T\omega \\ \text{s.t. } & y_i(\omega^T \cdot x_i+b)\geq 1, i=1,2,\dots,n \end{aligned} \quad (9)$$

Therefore, the optimal hyperplane problem is transformed into a mathematical optimal value problem. In the present study, the Lagrange dual transformation is applied for the optimization and the following expression is achieved in this regard [24]:

$$L(\omega, b, \alpha) = \frac{1}{2}\omega^T \cdot \omega - \sum_{i=1}^n \alpha_i (y_i (\omega^T x_i + b) - 1) \quad (10)$$

Where $\alpha_i \geq 0$, $i = 1,2,\dots, N$ denotes Lagrange coefficient, take the partial derivatives of ω and b , respectively, and set them equal to 0. The following is the constraint condition:

$$\left\{ \begin{array}{l} \text{s. t. } \alpha_i \geq 0, i = 1, 2, \dots, n \\ \sum_{i=1}^n \alpha_i y_i = 0 \end{array} \right. \quad (11)$$

Finally, the classification function can be expressed in the form below [24]:

$$f(x) = \text{Sgn}\left(\sum_{i=1}^n y_i \alpha_i (x_i \cdot x) + b_0\right) \quad (12)$$

Then Support Vector Machine (SVM) can be applied to complete the linear and nonlinear analysis of the data by adopting an unreasonable kernel function. Conventional kernel functions are as follows:

- (1) Linear kernel function: $k(x, y) = x^T y + c$
- (2) Polynomial kernel function: $k(x, y) = (ax^T y + c)^d$
- (3) Radial basis function kernel function: $k(x, y) = \exp\left(-\frac{\|x-y\|^2}{2\delta^2}\right)$.

Where δ is the bandwidth of the radial basis function kernel.

It should be indicated that the radial basis function (RBF) has only one parameter δ , which makes it easier to select the optimal value. Accordingly, the radial basis kernel function is selected in the present study. However, there are often some variables in the training of the SVM model that affect the training accuracy. Among these variables, the most important parameters are penalty variable factor C and kernel function variable δ . In order to ensure the accuracy of model training, it is necessary to use an appropriate intelligent optimization algorithm to optimize the support vector machine and seek the optimal combination of solutions.

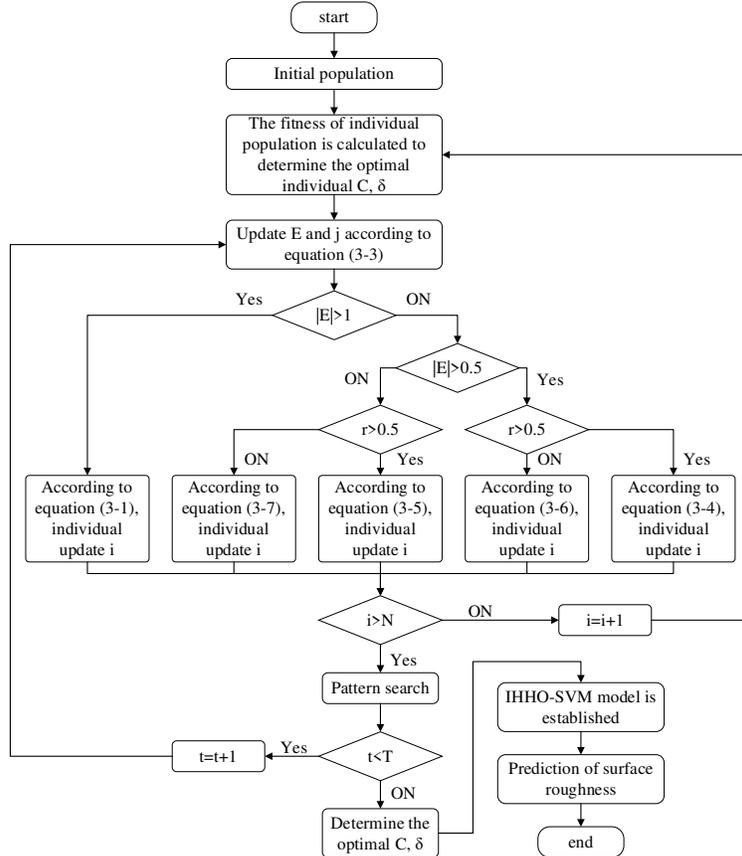


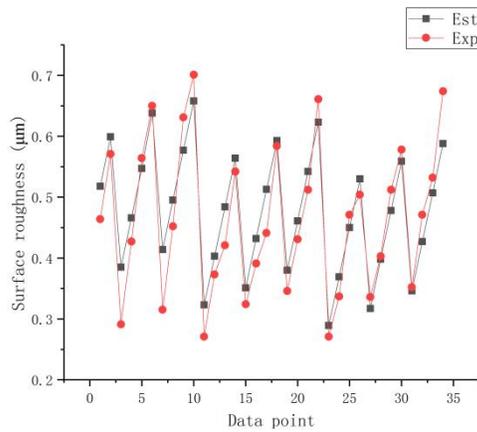
Figure 7: Flowchart of the IHHO-SVM

The performed investigations demonstrate that the improved Harris hawk optimization algorithm (IHHO) has a strong global searching ability. Consequently, applying the IHHO algorithm to search C and δ can avoid falling into the local optimal solution so that the global optimal solution can be obtained. Figure 7 shows the processing flowchart of the IHHO-SVM algorithm.

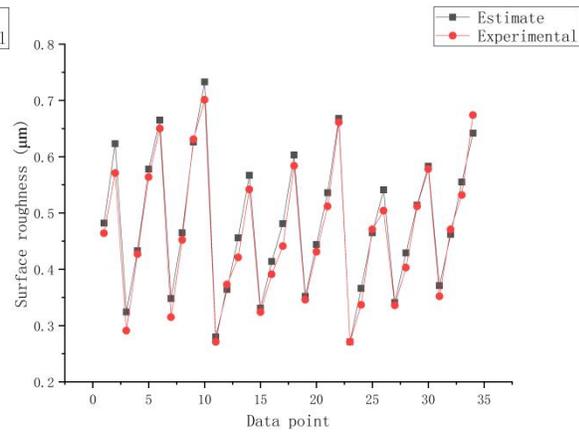
It is observed that the original search and optimization method of HHO is used to get the optimal solution, then the pattern search strategy is implemented and the obtained result is compared with the current optimal solution. The best solution is retained as the output, which is used to build an improved IHHO-SVM model to predict the surface roughness.

4.3.4. Forecast results

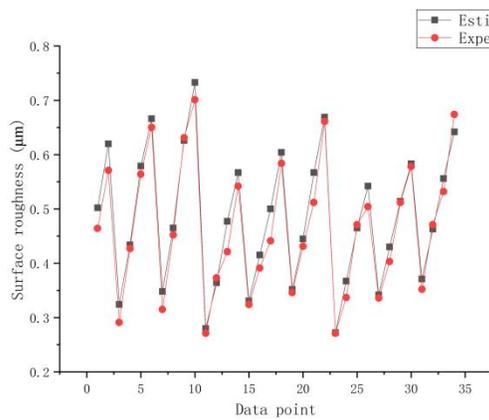
Figure 8 illustrates the point and line diagrams of the predicted and real values obtained from different schemes. It is observed that compared with other algorithms, the predicted value of the IHHO-SVM algorithm is closer to the experimental value.



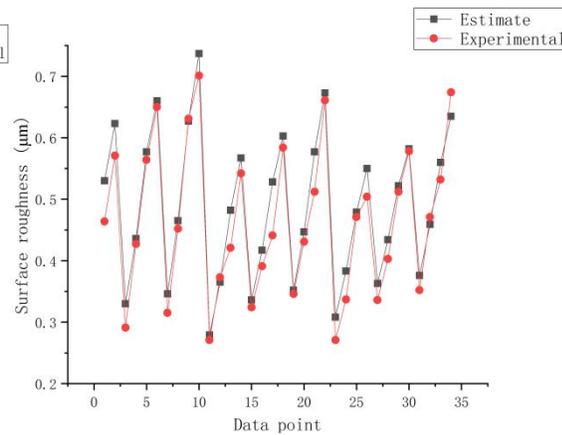
(a) SVM algorithm



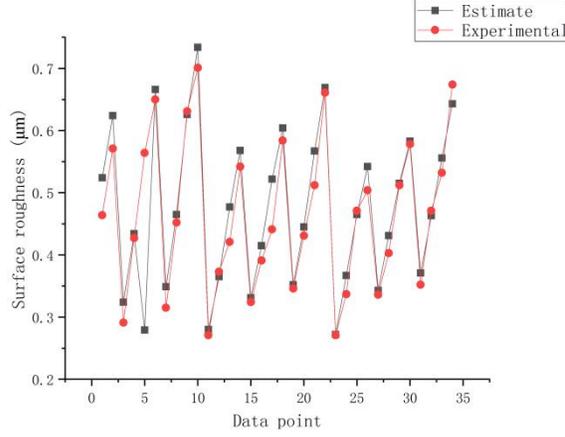
(b) IHHO-SVM algorithm



(c) GWO-SVM algorithm



(d) WOA-SVM algorithm



(e) BAT-SVM algorithm

Figure 8: Predicted and experimental values of different algorithms

In order to characterize the model reliability, root mean squared error (RMSE) and mean absolute error (MAE) are introduced in this article as statistical indicators. These indicators are defined as the following:

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |\hat{y}_i - y_i| \quad (13)$$

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2} \quad (14)$$

Where n is the number of forecast data. Moreover, y_i and \hat{y}_i denote the actual value and the predicted value, respectively.

Table 5 presents root mean squared error (RMSE), mean absolute error (MAE) and curve fitting degree R^2 of different algorithms. It is observed that RMS, MAE and R^2 of the optimized SVM are better than those of the standard SVM. More specifically, the RMSE of IHHO-SVM is 0.0225, MAE is 0.0186, R^2 is 0.94. Compared with other optimization algorithms, the prediction error and fitting accuracy are significantly improved. The results show that applying the improved HHO algorithm to optimize the SVM model is an effective scheme to improve the prediction accuracy of surface roughness.

Table 5: Comparison table of RMSE, MAE and R^2 of different algorithms

	SVM	IHHO-SVM	GWO-SVM	BAT-SVM	WOA-SVM
RMSE	0.0414	0.0225	0.0270	0.0573	0.0342
MAE	0.0351	0.0186	0.0215	0.0310	0.0274
R^2	0.85	0.94	0.93	0.927	0.918

4.4. Data storage module

In the data processing module, the predicted value of surface roughness is obtained, and the predicted value is stored in the Hbase database using the platform integrated database architecture. Meanwhile, the stored data in the Hbase can be queried automatically to facilitate subsequent use.

5. Conclusion

In the present study, the CDH platform is used to predict the surface roughness of titanium alloy milling and resolve the shortcoming of conventional methods, including low prediction accuracy and real-

time feedback. In this paper, the milling test of titanium alloy is completed, and the rectangular block with the size of 100 mm * 50 mm * 50 mm is cut by down milling. The main work is as follows:

(1) A data processing platform based on the CDH is designed and implemented, which uses the advantages and characteristics of various architectures integrated on the platform to meet the requirements of data transmission, data processing and data storage. In the studied cases, it was observed that the platform can accurately upload, process and store data, and has certain real-time performance.

(2) In order to predict the surface roughness more accurately and improve the prediction accuracy, the pattern search strategy is combined with the Harris hawk optimization algorithm (IHHO) and an improved optimization method is proposed to improve the global search ability and avoid falling into local extremum. Then the proposed algorithm is applied to optimize the support vector machine (SVM), and the global optimization of parameters C and δ is carried out to find the optimal solution. The obtained results show that the improved IHHO-SVM algorithm outperforms other optimization algorithms and prediction accuracy of up to 95% can be achieved.

References

1. Wang XM, Han J (2019) Study on surface roughness of TC4 titanium alloy in high speed milling. *Mechanical design and manufacturing* (5): 232-236
2. Gong XY, Li BH, Chai X, Gu M (2014) Overview of big data platform technology. *Journal of system simulation* 26(03): 489-496
3. Zhang T, Ling P (2014) Design, development and application of smart transportation big data platform. *Proceedings of the 9th China Intelligent Transportation annual conference*
4. Ramesh B (2015) *Big Data Architecture*. Big Data, Springer India
5. Zhong Y, Huang X, Liu D, Huang Y, Tian W, Wang J (2013) NoSQL storage scheme for large scale equipment monitoring data. *Computer integrated manufacturing system* 19(12): 3008-3016
6. Xu H, Yang Y, Zhang L (2013) Extension of LaUD-KV aggregation operation supporting large-scale monitoring data analysis. *Computer integrated manufacturing system* 19(012): 3035-3042
7. Dou M, Wen L, Wang J, Yan Z (2013) Parallel conversion algorithm of massive event logs based on MapReduce. *Computer integrated manufacturing system* 19(08): 1784-1793
8. Wu H, Ni Z, Wang H (2012) Ant colony algorithm based on MapReduce. *Computer integrated manufacturing system* (07): 1503-1509
9. Weng L (2020) Application of big data technology in mechanical automation. *Information recording materials*, 21(10): 162-163
10. Chen X (2021) Development trend of mechanical automation under the background of big data. *Technological innovation and productivity* (01): 73-75 + 78
11. Xian G (2020) Parallel Machine Learning Algorithm Using Fine-Grained-Mode Spark on a Mesos Big Data Cloud Computing Software Framework for Mobile Robotic Intelligent Fault Recognition. *IEEE Access* (99): 1-1
12. Dong J (2020) Research on evaluation method of milling cutter wear state based on CSSAE. *Dissertation, Dalian University of technology*
13. Min T (2019) Research on Online Tool Wear Warning Technology Driven by Big Data. *Dissertation, Nanjing University of Aeronautics and Astronautics*
14. Jiang X (2018) Research on carbon footprint accounting in parts processing based on Hadoop. *Dissertation, Zhejiang University of technology*
15. Duan X, Zhao Y, Chen Y, Ma C, Liao M (2020) Mechanical life prediction and evaluation of high voltage circuit breaker based on spark. *High voltage apparatus* 56(09): 80-86
16. Eser A, Ayyldz EA, Ayyldz M, Ghaleb M, Dabwan A (2021) Artificial Intelligence-Based Surface Roughness Estimation Modelling for Milling of AA6061 Alloy. *Advances in Materials Science and Engineering* 6: 1-10
17. Xu L, Huang C, Niu J, Li C, Wang X (2021) An improved case-based reasoning method and its application to predict machining performance. *Soft Computing* 25: 5683–5697

18. Nan X, Zhou J, Zheng B (2018) An energy-based modeling and prediction approach for surface roughness in turning. *The International Journal of Advanced Manufacturing Technology* 96:5-8
19. Heidari AA, Mirjalili S, Faris H (2019) Harris hawks optimization: algorithm and applications. *Future Generation Computer Systems* 97: 849-872
20. Fu W, Shao K, Tan J, Wang K (2020) Fault Diagnosis for Rolling Bearings Based on Composite Multiscale Fine-Sorted Dispersion Entropy and SVM with Hybrid Mutation SCA-HHO Algorithm Optimization. *IEEE Access* (99): 1-1
21. Usman BT, Ali Z, Yang F, Lu J, Muhammad K (2020) Short-term load forecasting for microgrid energy management system using hybrid HHO-FNN model with best-basis stationary wavelet packet transform. *Energy* 203
22. Zhang M, Zhang D, He R (2018) Chicken colony optimization algorithm based on pattern search method. *Microelectronics and computer* 35(04): 46-52
23. Gong C, Wang Z (2012) Proficient in MATLAB optimization calculation. Beijing: Electronic Industry Press
24. Zhu, Q (2018) Research and application of support vector machine based on improved ant colony algorithm. Dissertation , Anhui University of science and technology