EAI Endorsed Transactions
on Scalable Information Systems

Research Article EAI.EU

Research on Fault Diagnosis Method of CNC Machine Tools Based on Integrated MPA Optimised Random Forests

Xiaoyan Wang[1,*] and Donggang Xu[1]

[1] College of Advanced Materials Engineering Zhengzhou Technical College, Zhengzhou 450121, Henan, China
[2] College of Rail Transit Engineering Zhengzhou Technical College, Zhengzhou 450121, Henan, China

Abstract

INTRODUCTION: Intelligent diagnosis of CNC machine tool faults can not only early detection and troubleshooting to improve the reliability of machine tool operation and work efficiency, but also in advance of the station short maintenance to extend the life of the machine tool to ensure that the production line of normal production.
OBJECTIVES: For the current research on CNC machine tool fault diagnosis, there are problems such as poorly considered feature selection and insufficiently precise methods.
METHODS: This paper proposes a CNC machine tool fault diagnosis method based on improving random forest by intelligent optimisation algorithm with integrated learning as the framework. Firstly, the CNC machine tool fault diagnosis process is analysed to extract the CNC machine tool fault features and construct the time domain, frequency domain and time-frequency domain feature system; then, the random forest is improved by the marine predator optimization algorithm with integrated learning as the framework to construct the CNC machine tool fault diagnosis model; finally, the validity and superiority of the proposed method is verified by simulation experiment analysis.
RESULTS: The results show that the proposed method meets the real-time requirements while improving the diagnosis accuracy.
CONCLUSION: Solve the problem of poor accuracy of fault diagnosis of CNC machine tools and unsound feature system.

Keywords: cnc machine tools, fault diagnosis, integrated techniques, marine predator algorithm, random forests

Received on 23 December 2023, accepted on 25 April 2024, published on 2 May 2024

doi: 10.4108/eetsis.5785

[*]Corresponding author. Email: 29625664@qq.com

1. Introduction

CNC machine tools have the advantages of high processing accuracy, high productivity, high degree of flexibility, high degree of automation, and are widely used in various fields of industry [1]. With the manufacturing industry towards high-end, information technology, service, intelligent direction continues to develop, Shukang machine tool complexity and the degree of information technology is also increasing, long-time operation and use of the machine tool failure rate continues to rise [2]. Once the failure of CNC machine tools, light will lead to parts scrapped, heavy will lead to enterprise production stagnation, resulting in irreparable economic losses [3]. CNC machine tool troubleshooting generally refers to the operation of the equipment or basically do not dismantle all the equipment, to grasp the operating status of the equipment and determine the fault parts and causes [4]. Currently, CNC machine tool fault diagnosis still uses the traditional fault diagnosis service mode, the complexity and variety of the system leads to an increase in the difficulty of diagnosis and repair, and most of the maintenance and diagnosis work is dependent on the energy of the manufacturer's maintenance personnel [5]. With the development of big data technology and artificial intelligence technology, intelligent CNC machine tool fault diagnosis technology has become an inevitable trend [6]. Intelligent CNC machine tool fault diagnosis method can not only early detection and troubleshooting to improve the reliability of machine tool operation and work efficiency, but also in advance of the station short maintenance to extend the applicable life of the machine tool to ensure the normal production of the production line [7]. Therefore, the study of intelligent CNC machine tool fault diagnosis method is the current precise production line intelligent demand [8].

Intelligent fault diagnosis of CNC machine tools refers to the technology of automatic diagnosis and prediction of CNC machine tool faults by using artificial intelligence and machine learning and other technologies [9]. The study of intelligent fault diagnosis of CNC machine tools should not only analyse the triggering factors of CNC machine tool faults, but also accurately determine and find the fault location and give the fault type [10]. Currently, the research methods of CNC machine tool fault diagnosis technology mainly include expert system, fault tree, neural network, fuzzy theory, rough set, support vector machine and so on [11]. Literature [12] proposed an expert system based on PLC information fault diagnosis, and studied the diagnostic reasoning strategy based on fault tree according to the fault manifestation and diagnostic process of FMS system; Literature [13] researched a machine tool fault diagnostic process based on fault tree, modular decomposition of the fault tree and the introduction of the minimum cut-set relative composite degree of sorting, which accelerated the diagnostic speed and shortened the diagnostic scale; Literature [14] uses fuzzy neural network to carry out detection and fault diagnosis of water pumps, and verifies the good recognition ability of fuzzy neural network; Literature [15] carries out fault diagnosis by classifying vibration signals of electric pumps through the extreme learning machine, and proves the effectiveness of fault diagnosis method based on the ELM network; Literature [16] researches the fault diagnosis model of aero-engine sensors based on the extreme learning machine, and proves that the ELM performance is better than the neural network method; Literature [17] investigates a fault tree-based fault diagnosis process of machine tools. better than the neural network method; literature [17] through the construction of fuzzy theory-based fault diagnosis model, according to the hydraulic system failure phenomenon, quickly and accurately find the cause of the fault; literature [18] proposed based on rough set theory of diesel engine valve lash fault diagnostic identification method, to overcome the traditional fault diagnostic methods only provide a fault category defects. In view of the above literature analysis, the existing CNC machine tool fault diagnosis methods have the following defects [19]:

1) Inadequate consideration of failure triggers;

2) Fault type construction is not comprehensive enough;

3) CNC machine tool fault diagnosis methods are not accurate enough and have poor real-time performance.

Ensemble Learning [20] accomplishes learning tasks by constructing and combining multiple weak learners. Population Heuristics [21] refers to the construction of optimisation algorithms inspired by natural phenomena and human activities. Random forests are constructed by integrating a large number of decision trees using the idea of Bagging as the basic unit [22]. Integrated learning and heuristic optimisation algorithms to improve the random forest method makes the fault diagnosis accuracy increase, and its application to the problem of CNC machine tool fault diagnosis has become a research hotspot for experts and scholars in the field.

Aiming at the problems existing in the current CNC machine tool fault diagnosis method, this paper proposes a CNC machine tool fault diagnosis method based on the integrated heuristic optimisation algorithm to improve the random forest. The main contributions of this paper are: (1) by describing the fault diagnosis process of CNC machine tools and analysing and extracting the fault characteristics of CNC machine tools; (2) combining the integrated learning technology with the marine predator algorithm to optimize the random forest method and propose the fault diagnosis model of CNC machine tools; and (3) verifying the method of this paper through simulation to have a higher evaluation accuracy and real-time performance.

2. CNC machine tool troubleshooting problem analysis

2.1. Fault diagnosis process of CNC machine tools

The fault diagnosis process of CNC machine tools includes raw fault signal online acquisition, noise reduction processing, fault feature extraction, fault pattern recognition and fault prediction etc., the specific process is shown in Figure 1 [23].

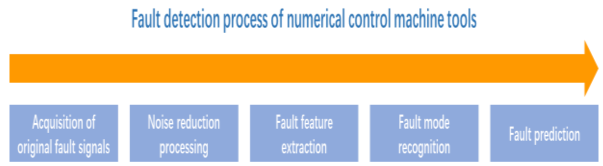


Figure 1. Fault diagnosis flowchart of CNC machine tools

As can be seen from Figure 1, CNC machine tool fault data mainly comes from the CNC machine tool host and each control system installed vibration, temperature, frequency and other different types of sensors; the original signal contains a large number of system noise and environmental noise, the original signal needs to be noise reduction and amplification; fault diagnosis feature extraction as a CNC machine tool state parameter determination and fault location of the key technology, the need to extract the key features, accurate prediction of Fault type and fault severity; Fault diagnosis through classification algorithms to build a mapping model between fault features and fault categories; Fault prediction is to use the fault diagnosis model to predict and diagnose the faults of CNC machine tools.

2.2. Fault diagnosis feature extraction

Fault diagnosis features include time domain features, frequency domain features [24].

Signal time domain characteristics

As the most direct and simple feature extraction method, time domain feature extraction mainly includes two categories: quantitative and dimensionless. The quantitative indexes include mean value, mean square value, RMS value, peak value, etc., and the dimensionless indexes include skewness, steepness, waveform factor, impulse factor, margin factor, etc. [25].

Signal frequency domain characteristics

The signal frequency domain features include mean square frequency, centre of gravity frequency and frequency variance, where the mean square frequency X_{msf} reflects the variation of the signal's main band position, the centre of gravity frequency X_{fc} reflects the signal's centre of gravity position, and the frequency variance X_{vf} reflects the signal's spectral energy distribution [26].

Decomposition of signal features based on wavelet packet noise reduction and EMD

The signal time-domain features and frequency-domain features mainly reflect the overall situation of the signal and cannot reflect the signal locally. In order to obtain accurate fault information, new signal processing methods are needed, for this reason, this paper proposes a joint time-frequency domain analysis method based on wavelet packet and EMD [27].

Wavelet transform has the ability of multi-resolution analysis and excellent local feature characterisation in both time and frequency domains. During wavelet packet decomposition, the main components of the original signal are retained and high frequency noise is filtered out.

The EMD decomposition method decomposes the signal into a number of IMF components, which satisfy that the number of signal extreme points and over-zero points is equal or the difference is one; at any point, the envelope of the local maxima and minima has an average value of zero. The EMD decomposition of the original signal and the wavelet packet noise cancellation processed signal are given in Figures 2 and 3, respectively. As seen in Figure 2 and 3, the quality of the EMD decomposition after wavelet packet noise reduction processing is significantly improved, and the high-frequency noise is effectively filtered out.

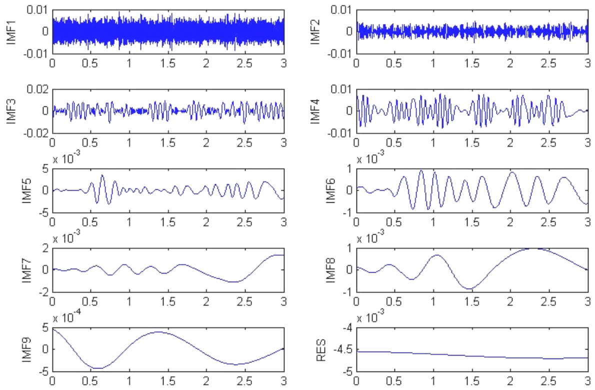


Figure 2. EMD decomposition of the original signal

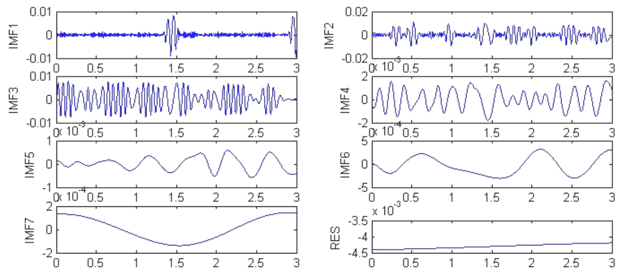


Figure 3. EMD decomposition of wavelet packet noise cancellation processed signal

For the EMD decomposed signal, its energy value is used as a feature:

$$E_i = \int \left| c_i(t) \right|^2 dt \tag{1}$$

In Eq. (1), c_i denotes the IMF signal component.

2.3. Fault diagnosis feature selection

Due to the complex structure of CNC machine tools, in order to comprehensively and accurately obtain the operating data of CNC machine tools, a large amount of sensor data is required, which results in the formation of high-dimensional features [28]. In order to reduce the time of classifier diagnosis and increase the diagnostic accuracy, this paper adopts the fuzzy clustering method to carry out the dimensionality reduction analysis of the mixed domain features, and the specific feature selection and dimensionality reduction process is shown in Figure 4, and the main steps are as follows:

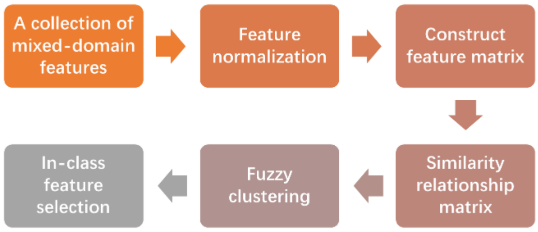


Figure 4. Flowchart of feature selection

Step 1: Establish similarity relationships. Firstly, the features within the mixed domain feature set are normalised to obtain normalised data:

$$x_i = \frac{y_i - y_{\min}}{y_{\max} - y_{\min}} \tag{2}$$

In Eq. (2), x_i denotes the normalised data, y_i denotes the original data to be normalised, and $y_{\min}$ and $y_{\max}$ denote the minimum and maximum values of the meta-three-year hi data, respectively.

Assume that the set of features after normalisation is $X_i = \{X_1, X_2, \cdots, X_n\}$, and n is the number of features. The feature matrix with number of samples m is equation (3):

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{m1} & x_{m1} & \cdots & x_{mn} \end{bmatrix} \tag{3}$$

Step 2: Calculate the similarity relationship matrix. The correlation coefficient between features X_i and X_j is shown in equation (4):

$$r_{ij} = \frac{\sum_{k=1}^{m}(x_{ki} - \bar{x}_i)(x_{kj} - \bar{x}_j)}{\sqrt{\sum_{k=1}^{m}(x_{ki} - \bar{x}_i)^2}\sqrt{\sum_{k=1}^{m}(x_{kj} - \bar{x}_j)^2}} \tag{4}$$

In Eq. (4), x_{ki} denotes the ith feature of the kth sample, x_{ki} denotes the jth feature of the kth sample, and $\bar{x}_i$ and $\bar{x}_j$ denote the mean values of the features X_i and X_j, respectively. The correlation coefficient is a measure of the degree of linear correlation between variables. r_{ij} reflects the degree of correlation between two features X_i and X_j. The correlation coefficient can be positive or negative, if r_{ij} is positive, it means they are positively correlated; if r_{ij} is negative, it means they are negatively correlated.

Step 3: Intra-class feature selection based on fuzzy clustering algorithm. Assuming that the number of sample categories is C, the number of samples under each category is the same, all n, and the number of features is J, the intra-class distance of the jth feature of the cth category is:

$$D^{c,j} = \frac{1}{n(n)-1}\sum_{l,m=1}^{n}\left|x_k^{c,j} - x_l^{c,j}\right| \quad (l \neq m) \tag{5}$$

In Eq. (5), $x_k^{c,j}$ and $x_l^{c,j}$ denote the kth sample and lth sample, respectively.

The average intra-class distance for the jth feature C class is:

$$D_n^j = \frac{\sum_{c=1}^{C} D^{c,j}}{C} \tag{6}$$

The mean value of the jth feature class c is:

$$\bar{x}^{c,j} = \frac{1}{n}\sum_{m=1}^{n} x_m^{c,j} \tag{7}$$

The average interclass distance for the jth feature C categories is:

$$D_w^j = \frac{1}{C(C-1)}\sum_{ci,cj=1}^{C}\left|\bar{x}^{ci,j} - \bar{x}^{cj,j}\right| \tag{8}$$

The evaluation function is constructed according to the principle that the smaller the spacing within classes the better, and the larger the spacing between classes the better:

$$\alpha_j = D_w^j - D_n^j \tag{9}$$

Within-class feature selection was carried out by sorting based on α_j size.

2.4. Stratified diagnosis

A single machine learning algorithm when can not solve the large-scale recognition space problem. Combined with the characteristics of CNC machine tool fault diagnosis, this paper adopts a hierarchical diagnosis model [29], schematically shown in Figure 5.

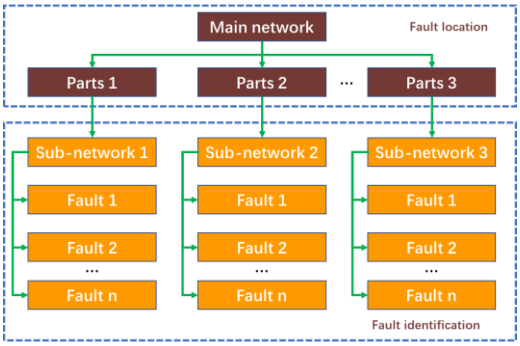


Figure 5. Schematic diagram of hierarchical diagnosis

As can be seen from Fig. 5, the first layer is mainly fault localisation, which is divided into different modules according to the components to which the fault belongs. The main network activates the corresponding sub-modules according to the diagnosis results. The second layer is responsible for diagnosing the specific categories of faults and the degree of faults in each component. Through the two layers of diagnostic output, the final diagnostic decision is obtained.

3. Ocean Predator Algorithm Optimising Random Forest Methods

Neural network iterative optimisation algorithm generally uses gradient descent method, which is easy to fall into the local optimum, in order to overcome the above defects, this paper adopts the intelligent optimisation algorithm - internal search optimisation algorithm to improve the neural network.

3.1. Marine Predator Algorithm

Marine Predators Algorithm (MPA) [30] is a novel meta-heuristic optimisation algorithm inspired by the theory of survival of the fittest in the ocean, where a marine predator chooses the optimal foraging strategy by choosing between Lévy wandering or Brownian wandering, as shown in Figure 6. The algorithm finds the optimal solution by simulating the foraging behaviour of marine predators. Its main features include strong ability to find the optimal solution, fast convergence speed, and easy implementation. The algorithm has been applied in several fields, such as wireless sensor network coverage optimisation.

Optimisation strategy

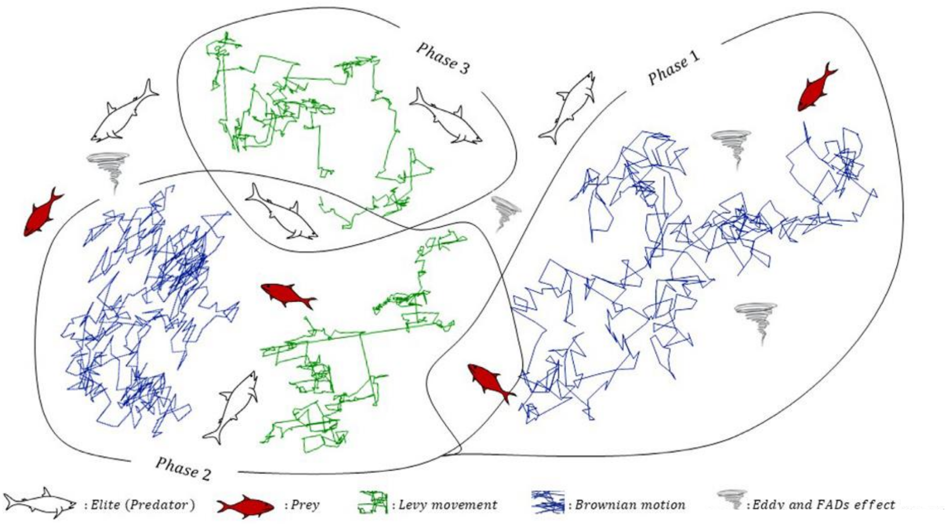


Figure 6. Schematic diagram of the marine predator algorithm decision

The Ocean Predator algorithm is an intelligent population-based optimisation algorithm where the initial solutions are distributed as uniformly as possible in the search space in order to ensure the quality of the search.

$$\boldsymbol{X}_1 = \boldsymbol{X}_{\min} + rand(1, \dim) \times (\boldsymbol{X}_{\max} - \boldsymbol{X}_{\min}) \quad (10)$$

In Eq. (10), $X_{\min}$ and $X_{\max}$ denote the upper and lower boundaries of the search space, respectively, and *dim* denotes the dimension of the population.

The MPA is divided into three main phases: the exploration phase, the balancing phase, and the development phase, which take place before, during, and after the merit seeking process, respectively.

Exploration phase: this phase usually occurs at the beginning of the iteration, in order to search for more space, the predator's action mainly obeys the Brownian motion, which has a large step size in favour of the algorithm's exploratory ability. The mathematical model of the exploration phase is shown below:

$$While\ iter < \frac{1}{3} iter_{\max}$$

$$\boldsymbol{stepsize}_i = \boldsymbol{R}_B \times (\boldsymbol{Elite} - \boldsymbol{R}_B \times \boldsymbol{Prey}_i) \quad i=1,...,n \quad (11)$$

$$\boldsymbol{Prey}_i = \boldsymbol{Prey}_i + P.\boldsymbol{R} \times \boldsymbol{stepsize}_i$$

In Eq. (12), $iter$ denotes the current iteration number, and $iter_{\max}$ denotes the maximum number of iterations. $\boldsymbol{R}_{\mathrm{B}}$ is a random vector with normal distribution based on Brownian motion, $\boldsymbol{Prey}_i$ denotes the position information of the first i individual, $\boldsymbol{Elite}$ is the global optimal individual position information, P is a constant with the value of 0.5, and $\boldsymbol{R} \in (0,1)$ is a uniformly distributed random vector.

Equilibrium phase: in this phase, the predator balances the exploration and exploitation of the search space, and therefore divides the population into two parts, one relying on the large step size of the Brownian motion for a wide range of searches, and the other utilising the smaller step size of the Levy distribution for deeper searches. The mathematical model for this phase is described below:

$$While\ \frac{1}{3} iter_{\max} \leq iter < \frac{2}{3} iter_{\max}$$

For the first part of the population, exploitation behaviour is mainly carried out:

$$\boldsymbol{stepsize}_i = \boldsymbol{R}_{\mathrm{L}} \times (\boldsymbol{Elite}\text{-}\boldsymbol{R}_{\mathrm{L}} \times \boldsymbol{Prey}_i) \quad i\text{=}1,\ldots,n \tag{12}$$

$$\boldsymbol{Prey}_i = \boldsymbol{Prey}_i + P.\boldsymbol{R} \otimes \boldsymbol{stepsize}_i \tag{13}$$

where $\boldsymbol{R}_{\mathrm{L}}$ is a random vector obeying a Lévy distribution.

For the second part of the population, exploratory behaviour was predominant:

$$\boldsymbol{stepsize}_i = \boldsymbol{R}_{\mathrm{B}} \times (\boldsymbol{R}_{\mathrm{B}} \times \boldsymbol{Elite}\text{-}\boldsymbol{Prey}_i) \quad i\text{=}1,\ldots,n \tag{14}$$

$$\boldsymbol{Prey}_i = \boldsymbol{Elite} + P.CF \times \boldsymbol{stepsize}_i \tag{15}$$

$$CF = (1\text{-}\frac{\mathrm{iter}}{iter_{\max}})^{2\frac{\mathrm{iter}}{iter_{\max}}} \tag{16}$$

where CF is an adaptive parameter that controls the predator step size.

Exploitation phase: in the final phase of the search, the predator locally exploits the search space, which is mathematically modelled as follows:

$$While\ iter > \frac{2}{3} iter_{\max}$$

$$\boldsymbol{stepsize}_i = \boldsymbol{R}_{\mathrm{L}} \times (\boldsymbol{R}_{\mathrm{L}} \times \boldsymbol{Elite}\text{-}\boldsymbol{Prey}_i) \tag{17}$$

$$\boldsymbol{Prey}_i = \boldsymbol{Elite} + P.CF \times \boldsymbol{stepsize}_i \tag{18}$$

In addition, Fish Aggregating Devices (FADs) are susceptible to being used as food by predators, thus losing the real prey. Therefore to avoid FADs, a larger step size is used for movement. The mathematical model of this behaviour is described below:

$$\boldsymbol{Prey}_i = \begin{cases} \boldsymbol{Prey}_i + CF[\boldsymbol{X}_{\min} + \\ \vec{\boldsymbol{R}} \times (\boldsymbol{X}_{max}\text{-}\boldsymbol{X}_{\min})] \times \boldsymbol{U}\ , if\ rand \leq FADs \\ \boldsymbol{Prey}_i + [FADs(1\text{-}r) + \\ r](\boldsymbol{Prey}_{r1}\text{-}\boldsymbol{Prey}_{r2}) \quad , if\ rand > FADs \end{cases} \tag{19}$$

where $FADs = 0.2$, denotes the probability of being affected by FADs, and $\boldsymbol{U}$ is a binary vector including either 0 or 1. When a random vector of 0 to 1 is generated and is less than 0.2, all the vector elements are changed to 0, and vice versa. $\boldsymbol{Prey}_{r1}$ and $\boldsymbol{Prey}_{r2}$ are two randomly selected individuals.

Process steps

According to the analysis of the principle and mechanism of MPA algorithm, the flowchart of MPA algorithm is shown in Figure 7, and the specific steps are as follows:

Step 1: Initialise the number of MPA populations with the number of iterations;

Step 2: Initialise the MPA population. Initialise the MPA population using the random uniform distribution strategy, calculate the fitness value, or obtain the current optimal value and optimal solution;

Step 3: Based on the number of iterations, the search phase is selected.The MPA algorithm selects the exploratory phase, equilibrium phase and exploitation phase search strategies to update the population location before, during and after the iterations, respectively;

Step 4: Calculate the fitness value and use the selection strategy to select and retain the better solution;

Step 5: Determine whether the number of iterations reaches the maximum number of iterations. If it reaches, output the optimal solution and optimal value; otherwise, return to step 3.

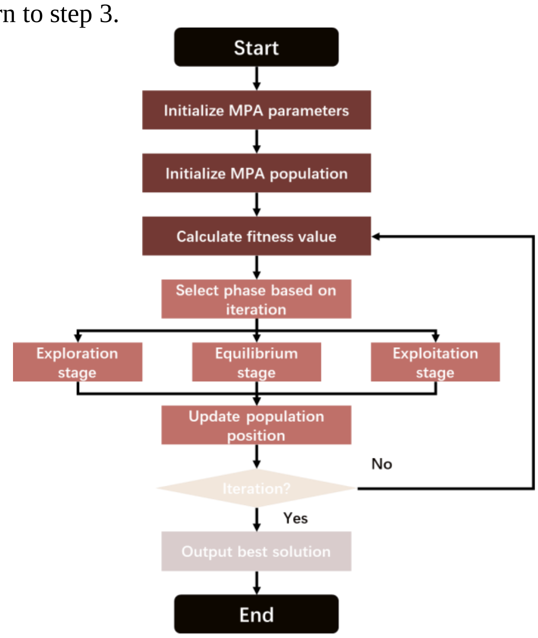


Figure 7. Flowchart of the marine predator algorithm

3.2. Random Forest Algorithm

Random Forest (RF) is an integrated learning algorithm based on decision trees, and its algorithmic principle is based on the Bagging integration algorithm and the random subspace method [31].The RF algorithm uses the samples to be classified, trains to produce decision trees, and determines the predicted classification results by aggregating all the results of all the decision trees by aggregating the voting, and its specific steps are as follows:

Step 1: Randomly draw samples, construct different sub-datasets, and train decision trees;

Step 2: Randomly select m attributes and use the information gain strategy to select the optimal attributes for decision tree node splitting and train to form a decision tree;

Step 3: Build different decision trees according to Steps 1~2 and integrate to construct a random forest model.

3.3. Random forest parameter optimisation based on MPA algorithm

Coding method and fitness function design

In order to improve the accuracy of RF, the MPA algorithm is used to optimise the RF hyperparameters, i.e., to optimise the number of decision trees and the minimum number of leaves of the RF algorithm, and in this paper, we use the real number encoding method to encode the number of decision trees and the minimum number of leaves. In order to accurately reflect the training RF advantages and disadvantages, this paper adopts the accuracy as the fitness function.

MPA-RF algorithm steps

According to the coding method and fitness function, the steps of the random forest classification and identification method based on MPA algorithm are as follows:

Step 1: The MPA algorithm encodes the initial parameters of the random forest, and also initialises the algorithm parameters such as the population parameters and the number of iterations; and calculates the value of the fitness function;

Step 2: Select exploratory, equilibrium, and exploitation phase search strategies to update the location information of MPA populations before, during, and after the iteration;

Step 3: Calculate the fitness function value and update the global optimal solution;

Step 4: Judge whether the termination condition is satisfied. If it is satisfied, exit the iteration, output the optimal RF hyperparameters and execute step 5, otherwise continue to execute step 3;

Step 5: Decode the MPA based optimised RF hyperparameters to obtain the number of RF decision trees and the minimum number of leaves;

Step 6: Construct the MPA-RF classifier and use the sample set to train to get the classification model.

4. Integrated Learning Technologies

In order to make the training process less prone to overfitting phenomenon and improve the MPA-RF accuracy, the integrated MPA-RF-Ada algorithm is constructed based on the MPA optimisation to improve the RF algorithm by combining with AdaBoost technology [32]. Assume the training sample set $\boldsymbol{X}=\left[x_{ij}\right]_{n\times d}$, where n is the number of samples and d is the dimension of the feature set of the mixed domain of fault signals; the base classifier MPA-RF, denoted as $f_{MPA-RF}(\cdot)$; the number of base evaluators is T . The specific algorithm is described as follows:

Step 1: Initialise the weight distribution of the samples $D_1=\left(\omega_{1,1},\omega_{1,2},\cdots,\omega_{1,n}\right)$, the weights of each sample are calculated as follows:

$$\omega_{1,i}=\frac{1}{n},i=1,2,\cdots,n \tag{20}$$

Step 2: For the iteration round $t=1,2,\cdots,T$, use the training sample base evaluator $h_t=f_{ISA-BP}\left(X,D_t\right)$ with the current distribution D_t ;

Step 3: Calculate the error of the base evaluator h_t on the set of training samples:

$$\varepsilon_t=\sum_{i=1}^{m_s}\omega_{t,i}e_{t,i} \tag{21}$$

$$e_{t,i}=\begin{cases}1 & y_i\neq h_t\left(x_i\right)\\ 0 & y_i=h_t\left(x_i\right)\end{cases} \tag{22}$$

In Eq. (22), $e_{t,i}$ is the error of the ith sample on the t base evaluator, $e_{t,i}=1$ denotes that the error is 1 at the supervised signal $y_i\neq h_t\left(x_i\right)$, and $e_{t,i}=0$ denotes that the error is 0 at the supervised signal $y_i=h_t\left(x_i\right)$.

Step 4: Calculate the weight coefficients for the base evaluator h_t a_t :

$$a_t=0.5\lg\frac{\varepsilon_t}{1-\varepsilon_t} \tag{23}$$

Step 5: Update the sample distribution D_{t+1} of the training sample set until the maximum number of iteration rounds is reached.

$$\omega_{t+1,i}=\omega_{t,i}e^{-a_t y_i^2} \tag{24}$$

Step 6: Linearly combine the T base evaluators to end up with the strong evaluator, the integrated MPA-RF (MPA-RF with AdaBoost, MPA-RF-Ada):

$$f_{MPA-RF-ada}(X) = round\left(\sum_{t=1}^{T}\left(\ln\frac{1}{a_t}\right)G(X)\right) \quad (25)$$

where $G(X)$ is the median of all $a_t h_t(X)$ and $round(\cdot)$ represents rounding.

5. Ideas of CNC machine fault diagnosis method based on MPA-RF-Ada algorithm

Combining AdaBoost and MPA-RF, this section proposes a method of CNC machine fault diagnosis model based on MPA-RF-Ada algorithm. The diagnostic model mainly takes the signal mixing domain features as input and the diagnostic identification type as output, and constructs the mapping relationship between the signal mixing domain features and the diagnostic identification type. The flowchart of the prediction of the talent team construction based on the MPA-RF-Ada algorithm is shown in Figure 8. The specific steps are as follows:

Step 1: According to the sensor status of the CNC machine, the CNC machine faults are collected, and the CNC machine fault diagnosis time-domain, frequency-domain, and time-frequency-domain features are extracted; the fuzzy clustering method is used to reduce the dimensionality of the features; and the dataset is divided into a training set, a validation set, and a test set;

Step 2: Initialise AdaBoost parameters. Randomly initialise the hyperparameters of MPA-RF; set the number of weak classifiers T' ; initialise the distribution weights of the training samples D'_1 ;

Step 3: Train the weak classifier MPA-RF.

(1) Use MPA algorithm to encode the initial parameters of RF, as well as initialise the algorithm parameters such as population parameters and iteration number; initialise the population and calculate the objective function value;

(2) Select exploration phase, equilibrium phase, and development phase search strategies to update the location information of MPA populations before, during, and after the iteration;

(3) Compare the value of the objective function of the population with the value of the objective function of the current global optimal solution and update the global optimal solution;

(4) Determine whether the termination condition is satisfied, if so, exit the iteration, output the optimal random forest hyperparameters, and execute step (2), otherwise continue to execute step (5);

(5) Decode the MPA-based optimised RF hyperparameters to obtain the optimal random forest hyperparameters and construct the weak classifier MPA-RF model;

Step 4: Calculate the weight coefficients a'_t as well as update the sample distribution D'_{t+1} ; train the weak classifier until the end of the iteration rounds and output the strong classifier MPA-RF-Ada;

Step 5: Identify the current test set using the trained strong classifier and output the corresponding fault types.

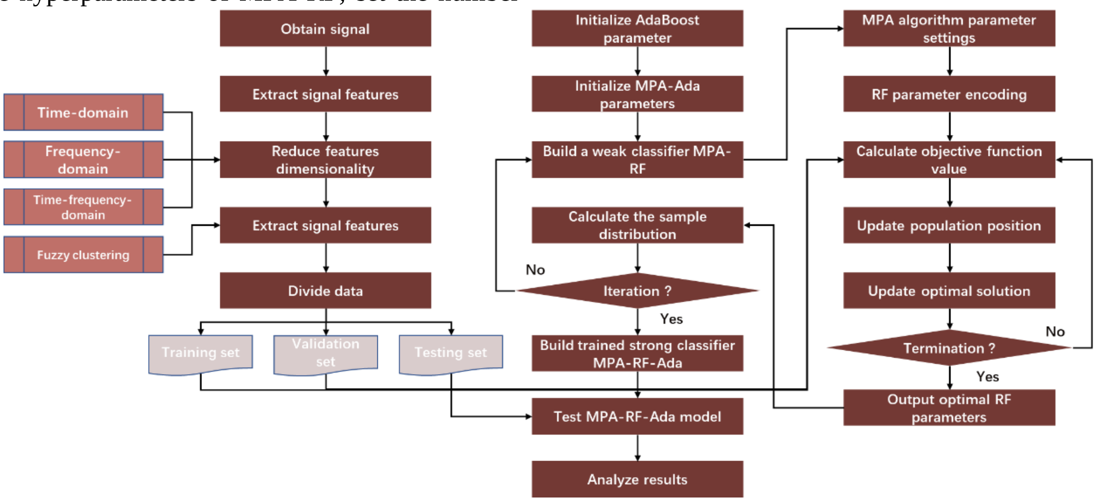


Figure 8. Flowchart of CNC machine tool fault diagnosis method based on MPA-RF-Ada algorithm

6. Experiments and results

In order to verify the accuracy and timeliness of the CNC machine tool fault diagnosis method proposed in this paper, five algorithms are selected for comparison, and the specific parameters of each algorithm are set as in Table 1.The data are collected from a W enterprise specialising in piston rod machining, and are divided into a training set, a validation set, and a test set, in which the training set is mainly used to train the model, the validation set is mainly used to compute the fitness value in the optimisation process, and the test set is mainly used to test the fault diagnosis model. The experimental simulation environment is Windows 10, CPU 2.80GHz, 8GB RAM, programming language Matlab2016a.

Table 1 CNC machine tool fault diagnosis method parameter setting

arithmetic	parameterisation
RF	N_tree=500, m_try=floor(80.5)
PSO-RF	Optimisation of RF hyperparameters using PSO with a maximum number of iterations of 200 and population size set as in MPA-RF-Ada, Vmax=30, Vmin=-30, r=0.5
GWO-RF	Optimisation of RF hyperparameters using GWO with a maximum number of iterations of 200 and population size set as in MPA-RF-Ada
MPA-RF	Optimisation of RF hyperparameters using MPA with a maximum number of iterations of 200, population size set as in MPA-RF-Ada, P = 0.5, FADs = 0.2
GWO-RF-Ada	Optimisation of RF hyperparameters using GWO with a maximum number of iterations of 200 and the same population size and number of classifiers settings as MPA-RF-Ada
MPA-RF-Ada	The RF hyperparameters were optimised using MPA with a maximum number of iterations of 200, the number of populations and the number of classifiers were set as described in Section 5.1, P = 0.5, FADs = 0.2

6.1. Parameter setting analysis

In order to analyse the impact of MPA algorithm population size and RF algorithm classifier number on the CNC machine tool fault diagnosis method, this paper compares and analyses the performance of CNC machine tool fault diagnosis method under the conditions of different population sizes and different number of classifiers, respectively. Figure 9 gives the graph of the impact of different population sizes and different numbers of classifiers on diagnostic accuracy, and Figure 10 gives the graph of the impact of different population sizes and different numbers of classifiers on diagnostic time.

As can be seen from Figure 9, as the population size of the MPA algorithm increases, the accuracy of CNC machine tool fault diagnosis has an increasing trend; as the number of classifiers increases, the accuracy of CNC machine tool fault diagnosis increases. As can be seen from Figure 10, as the population size of the MPA algorithm increases, the CNC machine tool fault diagnosis time is also increasing; as the number of classifiers increases, the CNC machine tool fault diagnosis time is increasing. In summary, the intelligent optimisation algorithm selected in this paper has a population size of 70 and a number of classifiers of 100.

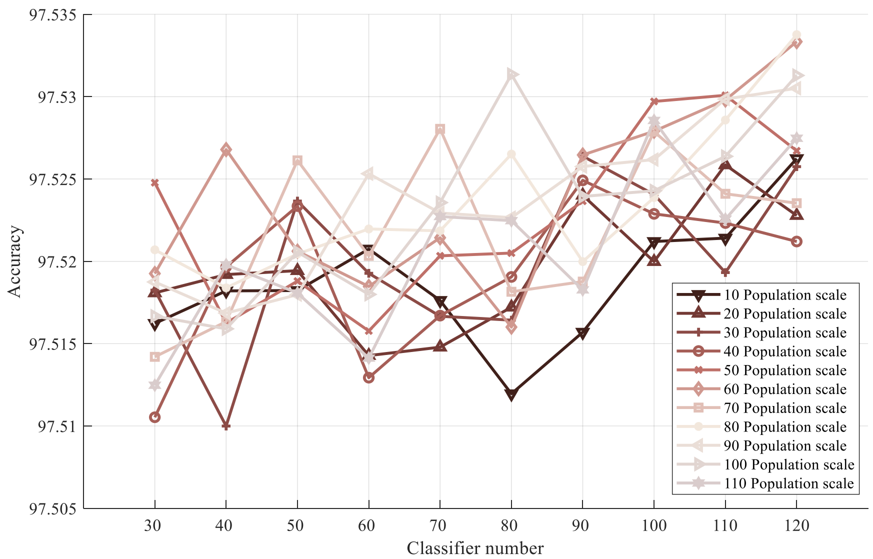


Figure 9. Effect of different population sizes and number of classifiers on diagnostic accuracy

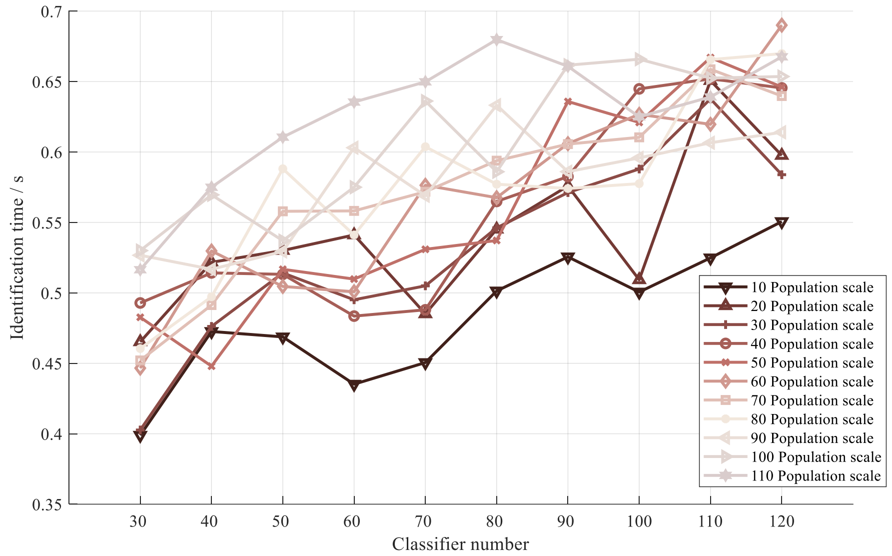
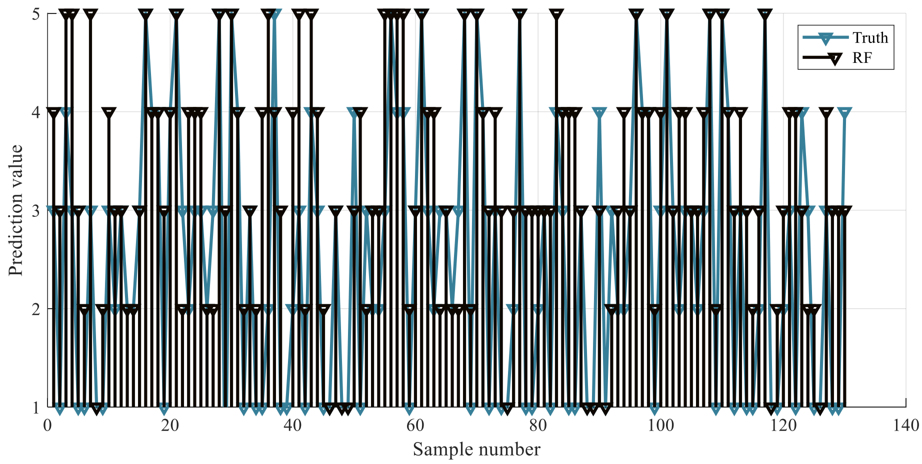


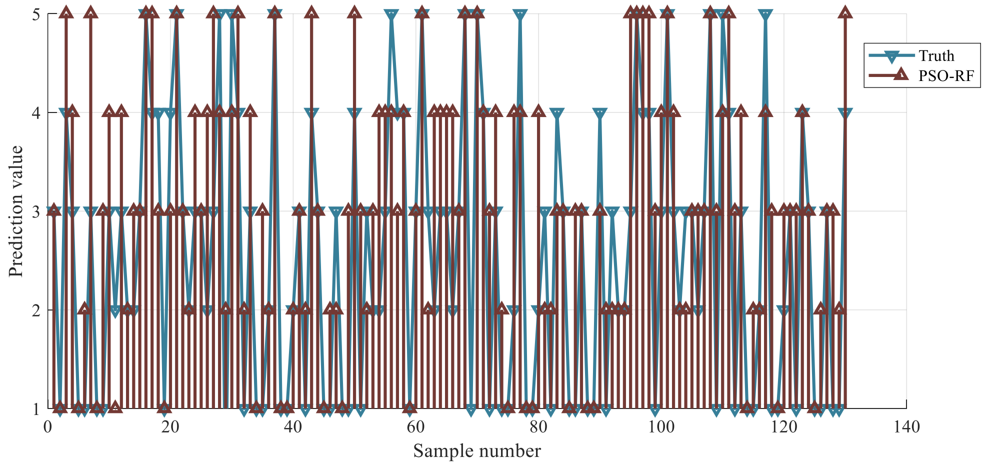
Figure 10. Effect of different population sizes and number of classifiers on diagnosis time

6.2. Experimental prediction performance analysis

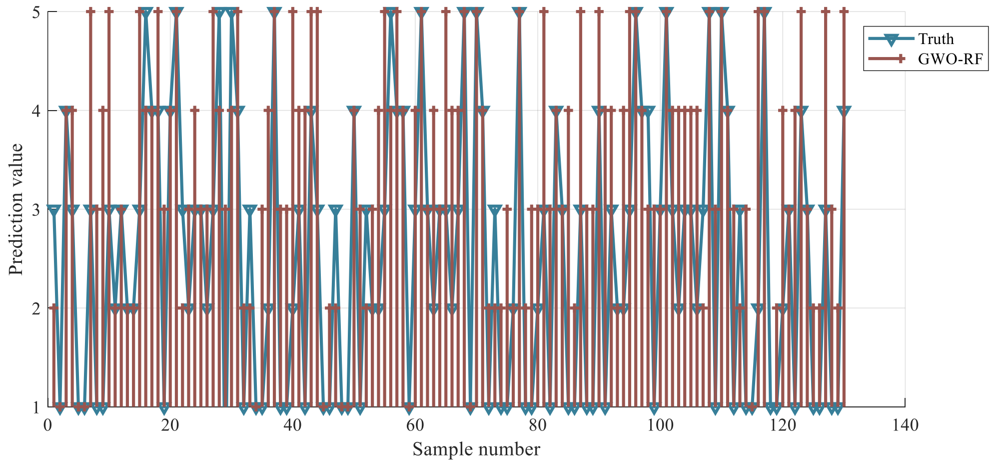
In order to verify the effectiveness and superiority of the CNC machine tool fault diagnosis method based on the MPA-RF-Ada algorithm, MPA-RF-Ada is compared with five other models such as RF, PSO-RF, GWO-RF, MPA-RF, and GWO-RF-Ada, and the evaluation results of each model are shown in Figure 11, Figure 12, and Figure 13. Figure 11 gives the results of fault diagnosis of CNC machine tools based on each algorithm. From Figure 11, it can be seen that comparing the MPA-RF-Ada and MPA-RF algorithms shows that the integration technique improves the diagnostic accuracy of the MPA-RF algorithm; comparing the MPA-RF with the RF, PSO-RF, and GWO-RF algorithms shows that the MPA algorithm optimises RF hyper-parameters, which improves RF diagnostic accuracy; and comparing the MPA-RF-Ada with the RF, PSO-RF, GWO-RF, MPA-RF, GWO-RF-Ada shows that the diagnostic results of MPA-RF-Ada are better than other models. At the same time, the accuracy of CNC machine tool fault diagnosis identification based on MPA-RF-Ada is the highest.



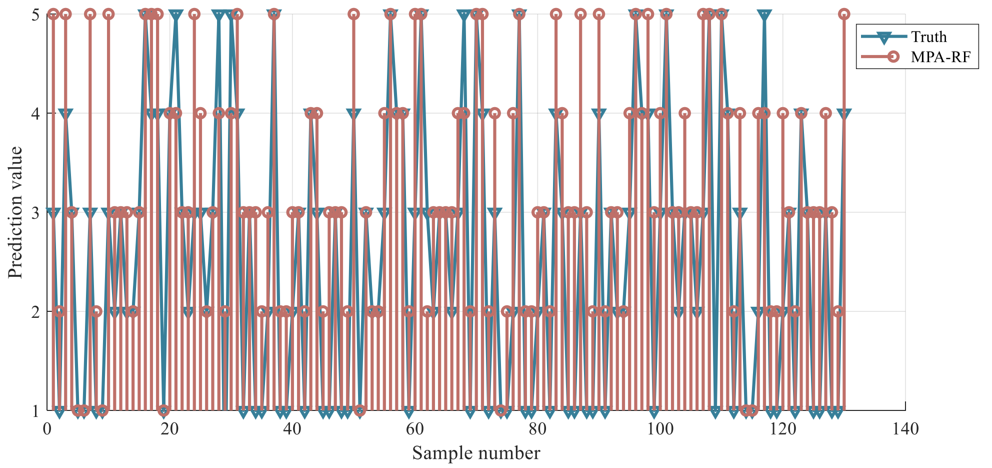
(a) RF



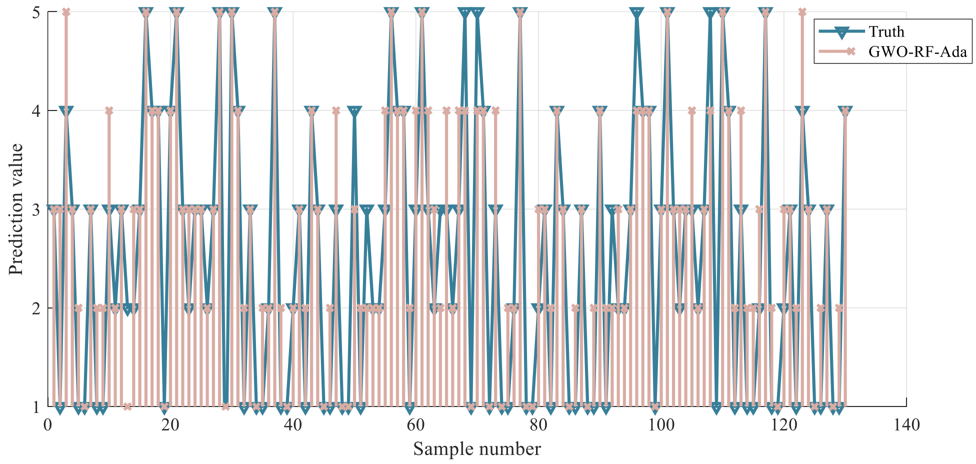
(b) PSO-RF



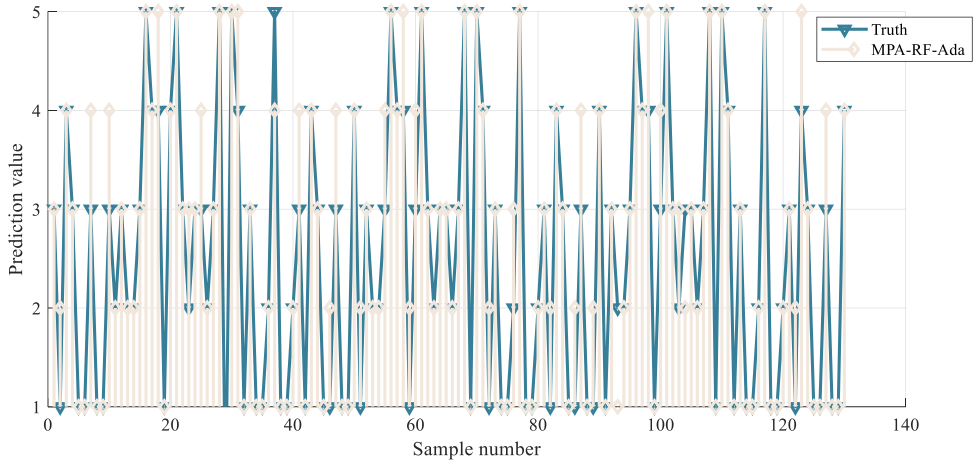
(c) GWO-RF



(d) MPA-RF



(e) GWO-RF-Ada



(f) MPA-RF-Ada

Figure 11. Fault diagnosis results of CNC machine tools based on each algorithm

In order to further verify the superiority of the CNC machine tool fault diagnosis method based on the MPA-RF-Ada algorithm, the diagnostic performance results of each algorithm are statistically given in this section, as shown in Figure 12 and Figure 13. As can be seen from Figure 12, the accuracy of CNC machine tool fault diagnosis based on MPA-RF-Ada algorithm is greater than other algorithms, and the diagnostic effect is better than other algorithms. From Figure 13, it can be seen that the fault diagnosis time of CNC machine tools based on MPA-RF-Ada algorithm is smaller than other algorithms, and the time performance is better than other algorithms. In conclusion, the CNC machine tool fault diagnosis method based on MPA-RF-Ada algorithm is better than other algorithms and meets the real-time requirements.

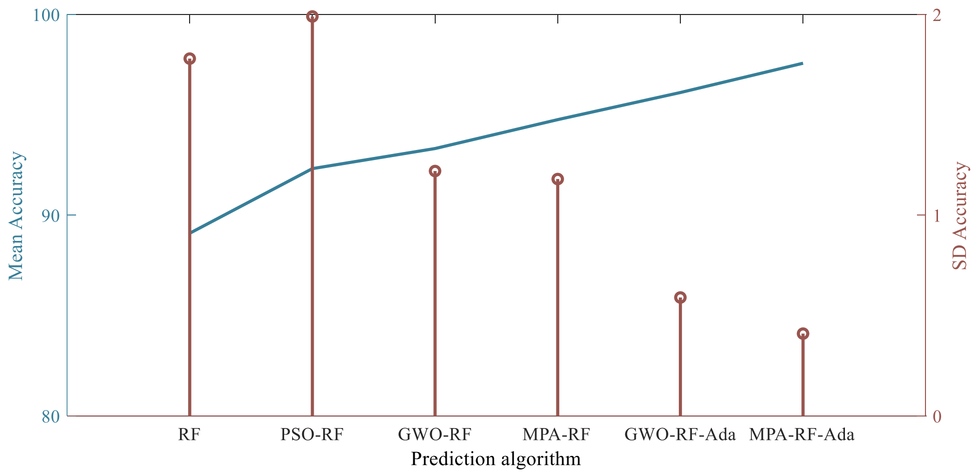


Figure 12. Comparison of fault diagnosis accuracy analysis of CNC machine tools based on each algorithm

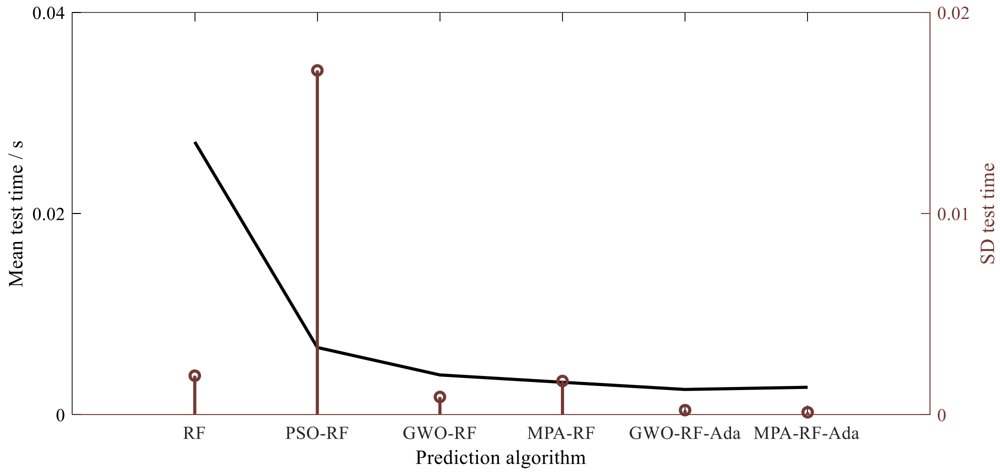


Figure 13. Comparison of fault diagnosis time analysis of CNC machine tools based on each algorithm

7. Conclusion

Aiming at the defects of the CNC machine tool fault diagnosis method, which has incomplete features , low accuracy and low real-time performance, this paper adopts the integrated learning technology, marine predator optimisation algorithm, and random forest to construct the CNC machine tool fault diagnosis method. The method extracts CNC machine tool fault diagnosis signal features by analysing the CNC machine tool fault diagnosis process. Combined with the integrated learning technology, the MPA algorithm is used to improve the random forest and construct the CNC machine tool fault diagnosis model. Simulation experiments are carried out using CNC machine tool fault diagnosis data, and the following conclusions are drawn:

(1) By comparing the diagnostic performance of MPA-RF with RF, PSO-RF, and GWO-RF algorithms, the MPA algorithm can improve the diagnostic accuracy of RF;

(2) Integrated learning further improves the diagnostic model accuracy by comparing the diagnostic performance of the MPA-RF-Ada and MPA-RF algorithms;

(3) MPA-RF-Ada prediction time meets real-time requirements.

The MPA algorithm optimisation used in this paper can easily fall into a local optimum, making the diagnostic model subject to certain limitations. In future work, the introduction of multi-strategy optimisation of the MPA algorithm will be considered to improve the efficiency of the algorithm.

Acknowledgements.

This work is supported by Key Scientific Research Project of Henan Higher Education Institutions in 2024，" Research on the Key Technologies for Predicting and Controlling CNC Cutting Flutter Based on Deep Neural Network Learning Algorithms" (24B460034)

References

[1] Liu Y , Guo L , Gao H , You Z, Zhang B. Machine vision based condition monitoring and fault diagnosis of machine tools using information from machined surface texture: a review[J].Mechanical Systems and Signal Processing, 2022, 164(11):108068.

[2] Shafiullah M , Alshumayri K A , Alam M S .Machine learning tools for active distribution grid fault diagnosis[J].Advances in engineering software,. 2022.

[3] Liu J .Fault Detection Method of CNC Machine Tool Based on Wavelet Transform[J].Parallel Processing Letters, 2022.

[4] Albayati M G , Faraj J , Thompson A , Patil P, Gorthala R, Rajasekaran S. Semi-Supervised Machine Learning for Fault Detection and Diagnosis of a Rooftop Unit[J].Big Data Mining and Analytics, 2023, 6(2):170-184.

[5] Akshay E , Sugumaran V , Elangovan M .Single Point Cutting Tool Fault Diagnosis in Turning Operation Using Reduced Error Pruning Tree Classifier[J]. Structural Durability and Health Monitoring (English), 2022(003):016.

[6] Jiao J , Lin J , Zhao M , Liang K, Ding C. Cycle-consistent Adversarial Adaptation Network and its application to machine fault diagnosis.[J].Neural networks : the official journal of the International Neural Network Society, 2022, 145:331-341.

[7] Jia P , Li P , Zheng F ,Feng Q, Zhang B.Simultaneous measurement of 5DOF spindle error motions in CNC machine tools[J].Applied optics, 2022, 61(19). 5704-5713.

[8] Yuan J , Li L , Shao H , Han M, Huang H. Material recognition for fault diagnosis in machine tools using improved Mel Frequency Cepstral Coefficients[J]. Journal of Manufacturing Processes, 2023.

[9] Iqbal M , Madan A K .CNC Machine-Bearing Fault Detection Based on Convolutional Neural Network Using Vibration and Acoustic Signal[J]. Vibration Engineering & Technologies, 2022.

[10] He Y , Xiang H , Zhou H , Chen J. In-situ fault detection for the spindle motor of CNC machines via multi-stage residual fusion convolution neural networks[J].Computers in Industry, 2023.

[11] Wu Y , Liu L , Qian S .A small sample bearing fault diagnosis method based on variational mode decomposition, autocorrelation function, and convolutional neural

network[J].The International Journal of Advanced Manufacturing Technology, 2023.
[12] FAN Liping, LANG Lang, XIAO Jingjing, et al. Research on intelligent diagnosis expert system for multi-parameter monitor faults based on fault tree[J]. Journal of Biomedical Engineering, 2022, 39(3):10.
[13] Su Yishun,Shen Bin. Research on reasoning machine for fault diagnosis of CNC machine tools based on fault tree[J]. Mechatronics, 2014(9):5.
[14] WU Shiaocen,HAN Jingxian. Research on electric vehicle transmission fault detection based on fuzzy neural network[J]. Journal of Jiamusi University: Natural Science Edition, 2023, 41(5):112-115.
[15] YANG Jingzong, SHI Chunzhao, YANG Tianqing, WU Limei. Fault diagnosis method based on LMD and ABC optimised KELM[J]. Industrial Engineering, 2022(3):025.
[16] ZHAO Wanli, GUO Yingqing, XU Kejie, YANG Qingcai, WANG Kun, GUO Pengfei. Multi-model-based hybrid fault diagnosis method for aero-engine sensors[J]. Propulsion Technology, 2023, 44(3):224-232.
[17] Liu Lu,Shi Xianjun. Adaptive fuzzy rule-based fault diagnosis method based on model and data-driven[J]. Journal of Naval Aviation University, 2023, 38(4):368-374.
[18] LI Guoyou, YANG Mengqi, HANG Bingpeng, LI Chenguang, WANG Weijiang. Chemical process fault diagnosis based on fuzzy rough set and whale optimisation support vector machine[J]. Vibration and Shock, 2022, 41(2):8.
[19] Qin Q , Li L , Zhao G , Li Z. Uncertainty quantification for spindle axial thermal error of CNC machine tools considering hysteresis effect[J]. Precision Engineering, 2023.
[20] Yang X , Wang S , Meng L , Peng Y, Chen J. Short-term photovoltaic power prediction with similar-day integration by BP-AdaBoost based on the Grey-Markov model[J].Electric Power Systems Research, 2023.
[21] Zhang Mengjian, Wang Deguang, Wang Min, Yang Jing. Novel intelligent optimisation algorithms for solving engineering constrained problems and prospects[J]. Computer Applications, 2022, 42(2):8.
[22] Chang Jucai, Qi Pengfei, Chen Xiao. Recognition of rock hardness of multiple working conditions cutting in roadheader based on feature preference and random forest[J]. Journal of Coal, 2023, 48(2):1070-1084.
[23] Losi E , Venturini M , Manservigi L , Ceschini G F, Bechini G, Cota G.Prediction of Gas Turbine Trip: A Novel Methodology Based on Random Forest Models[J].Journal of engineering for gas turbines and power: Transactions of the ASME, 2022(3):144.
[24] Chan L , Cheng C .A novel intelligent fault diagnosis method based on variant sparse filtering and back-propagation[J].Noise & vibration worldwide, 2022(3):53.
[25] Wang Z , Xuan J , Shi T .Alternative multi-label imitation learning framework monitoring tool wear and bearing fault under different working conditions[J].Adv. Eng. Informatics, 2022, 54:101749.
[26] Moussaoui L , Aouaouda S , Rouaibia R .Fault tolerant control of a permanent magnet synchronous machine using multiple constraints Takagi-Sugeno approach[J].Electrical Engineering & Electromechanics, 2022.
[27] Feng W , Zhang G , Ouyang Y ,et al. Fault Diagnosis of Oil-Immersed Transformer based on TSNE and IBASA-SVM[J].Recent patents on mechanical engineering, 2022.
[28] Pan H , Xu H , Zheng J .A novel symplectic relevance matrix machine method for intelligent fault diagnosis of roller bearing[J].Expert Systems with Applications, 2022, 192:116400.
[29] Li W , Li C , Wang N ,Li J, Zhang J. Energy Saving Design Optimization of CNC Machine Tool Feed System: A Data-Model Hybrid Driven Approach[J].IEEE transactions on automation science and engineering: a publication of the IEEE Robotics and Automation Society, 2022.
[30] Yang J , Zheng M , Chen S .Illumination correction with optimised kernel extreme learning machine based on improved marine predators algorithm[J]. Colour Research And Application, 2022, 47(3):630-643.
[31] Yao Dengju, Yang Jing, Zhan Xiaojuan. Feature selection algorithm based on random forest[J]. Journal of Jilin University: engineering edition, 2014(1):5.
[32] WANG Mei, TIAN Ye, YANG Erlong, DANG Qingjun, KAN Chunling. A prediction method of reverse compression drive effect in oil wells based on AdaBoost-SVM algorithm[J]. Practice and understanding of mathematics, 2023, 53(2):8.