# A machine learning-based approach for product maintenance prediction with reliability information conversion

Hua Zhang[1,3], Xue He[1,2], Wei Yan[4,5*] , Zhigang Jiang[2,3] and Shuo Zhu[1,3]

## Abstract

Predictive maintenance (PdM) cannot only avoid economic losses caused by improper maintenance but also maximize the operation reliability of product. It has become the core of operation management. As an important issue in PdM, the time between failures (TBF) prediction can realize early detection and maintenance of products. The reliability information is the main basis for TBF prediction. Therefore, the main purpose of this paper is to establish an intelligent TBF prediction model for complex mechanical products. The reliability information conversion method is used to solve the problems of reliability information collection difficulty, high collection cost and small data samples in the process of TBF prediction based on reliability information for complex mechanical products. The product reliability information is fully mined and enriched to obtain more reliable and accurate TBF prediction results. Firstly, the Fisher algorithm is employed to convert the reliability information to expand the sample, and the compatibility test is also discussed. Secondly, BP neural network is used to realize the final prediction of TBF, and PSO algorithm is used to optimize the initial weight and threshold of BP neural network to avoid falling into local extreme value and improve the convergence speed. Thirdly, the mean-absolute-percentage-error and the Coefficient of determination are selected to evaluate the performance of the proposed model and method. Finally, a case study of TBF prediction for a remanufactured CNC milling machine tool (XK6032-01) is studied in this paper, and the results show that the feasibility and superiority of the proposed TBF prediction method.

**Keywords:** Predictive maintenance, Time between failures, Reliability Information conversion, Machine learning

## 1 Introduction

With the rapid development of Internet of Things, information technology and Artificial Intelligence, Predictive maintenance (PdM) has gradually attracted great attention of scholars [1–4]. As a core of operation management in the industrial system, PdM plays a vital role in reducing the frequency of failures, improving operating efficiency, and ensuring product quality [5, 6], and has become a research hotspot of Prognostics and Health Management (PHM) in aerospace and manufacturing fields [7–9]. At present, there are many methods to solve the problem of PdM of equipment. These methods can be roughly divided into two categories [10]: PdM methods based on fault mechanism [11–16] and data-driven. The data-driven methods are divided into mathematical statistics [17–21] and artificial intelligence [22–26]. During the implementation of PdM, through the accurate prediction of product failure time, the most cost-effective maintenance window can be determined before its failure. Furthermore, it can prevent the operating deterioration of equipment and minimize the shutdown time and maintenance costs through appropriate maintenance activities [27]. Therefore, improving

*Correspondence: yanwei81@wust.edu.cn
[4]School of Automobile and Traffic Engineering, Wuhan University of Science and Technology, Wuhan, 430081, China
[5]Department of Mechanical Engineering, School of Engineering, Cardiff University, Cardiff, CF24 3AA, UK
Full list of author information is available at the end of the article

the prediction accuracy of TBF is of great significance and can bring tangible benefits to the industry.

As an important reliability index, fault time can show the dynamic evolution process of faults, and have been predicted by a variety of data-driven methods, such as autoregressive moving average (ARMA) [28], singular spectrum analysis (SSA) [29], support vector regression (SVR) [30], artificial neural network (ANN) [31], etc. Due to the difficulty of the reliability test for complex system, the reliability information (such as TBF) is usually difficult to obtain [32]. Most of the above methods are for component, and are not applicable to the fault time prediction of complex mechanical products.

To fill this gap, the paper proposed a machine learning-based TBF prediction model with reliability information conversion. Mainly through the information conversion method based on Fisher algorithm to deeply mine the multi-source reliability information, convert the multi-source reliability information (such as fault data of similar products, unit, subsystem, and reliability simulation test data, etc. [33]) into the reliability information of the target product, expand the sample size of the small sample target product. For data samples with certain errors after capacity expansion, BP neural network with good prediction effect on time series data is selected to realize the final prediction of TBF, and PSO algorithm is used to optimize the initial weight and threshold of BP neural network to avoid falling into local extreme value and improve the convergence speed. Thus, more accurate TBF prediction results can be obtained. The rest of this paper is organized as follows: Sect. 2 provides an overview of existing methods for PdM. Section 3 introduces the methodology used in this paper, including reliability information conversion, compatibility test and TBF prediction based on PSO-BP. Section 4 is a case study, and the prediction results of TBF are also analyzed in this section, followed by a brief summary in Sect. 5.

## 2  Literature review

TBF prediction of complex mechanical products is a typical small sample problem. There are two methods to solve the problem: one is mathematical methods suitable for data analysis of small samples, such as Bayesian and poor information method; The former can improve the evaluation accuracy of the field test with small sample by fusing the simulation prior information [34–37], but there is problem of how to scientifically determine the prior distribution before information fusion. The evaluation accuracy will even drop when the prior information is distorted. In recent years, scholars have carried out relevant research on fault prediction, process evaluation and event prediction in poor information situation, and achieved good results [38–42]. But the core prediction model of this method is the use of Grey Model alone or in combination with the

Bootstrap. This type of prediction method does not essentially avoid the problem of averaging the weights after superposition calculation. The mean weight will cause the instability and large prediction errors of the prediction model. The second is expansion of small samples, such as Bootstrap [43–45]. The Bootstrap relies too much on the subsample, which is not conducive to the robust of parameter evaluation. While expanding the capacity of small sample, the information conversion method can overcome the problems of the above two methods. This method was first proposed by Qian Xuesen in his research on two bombs and one satellite, that is, in order to evaluate reliability, it is often necessary to convert and synthesize experimental data in different environments [46]. At present, there are mainly two information conversion methods for converting reliability by environmental factors and regression conversion of reliability information. Because the latter has certain requirements on the distribution of reliability information in use, that is, it is required to obey a specific distribution. Therefore, it is necessary to find a method that is not affected by the distribution of reliability information to solve the problem of insufficient fault data in the TBF prediction problem of mechanical products with small samples. As the Fisher algorithm is a linear discriminant method based on the idea of variance analysis, which can better distinguish each population, and have no requirements for the distribution of the population [47]. Therefore, this paper uses the information conversion method based on Fisher algorithm to expand the capacity of small and medium-sized sample in TBF prediction of complex mechanical products.

TBF belongs to time series data. Due to the excellent performance in solving the problem of time series prediction, machine learning is very consistent with the needs of practical engineering problems [48]. To solve time series analysis problems, researchers have applied linear regression model, SVM [49], decision tree and other machine learning models. The time series prediction based on machine learning is a typical supervised learning task, in which the input attribute is the time series data of historical data, and the output label is the future data (on the training set). Machine learning model is a data-driven way to learn sequence features and establish the mapping relationship between historical data and future data. In practical application, the time series prediction method based on machine learning performs better than the method based on statistics [50]. In order to improve the accuracy of time series prediction, researchers try to make a breakthrough in model complexity and sequence feature expression [51]. As a classical feedforward neural network, BP neural network has the advantages of good self-learning, self-organization and adaptability, and strong fault tolerance for its structural characteristics, which allows certain errors of input samples. Good simulation effect can

be achieved for the expanded data samples [52], which is conducive to reduce the error of the method proposed in this paper. However, the traditional BP neural network has the defects that it is easy to fall into local extremum and the convergence speed is slow or even non convergent [53]. In view of this defect, scholars use genetic algorithm (GA) to optimize BP neural network [54, 55] and improve the convergence speed. Compared with GA algorithm, particle swarm optimization (PSO) algorithm is simple in calculation and requires less parameters adjustment. Using PSO algorithm to optimize the initial weight and threshold of BP neural network cannot only avoid falling into local extreme value but also improve the convergence speed [56]. Therefore, BP neural network with good prediction effect on time series data is selected to realize the final prediction of TBF in this paper, and PSO algorithm is also adapted to optimize BP neural network to verify the feasibility of information conversion method based on Fisher algorithm.

## 3 Methodology

### 3.1 Reliability information conversion

The reliability information of a mechanical product (such as: fault data of similar product, unit, subsystems, and data of reliability simulation, etc. [52]) is various and have multidimensional characteristics. Specifically, according to the performance reliability analysis of the products, the reliability information can be divided into three dimensions: time, hierarchy, and information source. In terms of time dimension, it includes design phase information, manufacturing phase information, simulation operation phase information, etc. In the hierarchy dimension, it includes part level information, unit level information, subsystem level information and system level information. In this dimension, the amount of information presents an inverted pyramid; In the dimension of information source, it includes expert judgment information, performance degradation information, similar product information, etc. In order to use the information effectively and fully reflect its value, it needs to be processed deeply in the process of using reliability information, that is, the reliability information needs to be effectively converted before they can be used. Since the Fisher algorithm is a linear discriminant method established based on the idea of variance analysis, it can better distinguish each population, and this discriminant method does not make any requirements on the distribution of the population [53]. Therefore, the paper proposed an information conversion method based on Fisher algorithm to convert the reliability information. The following paragraph describes the detail.

The central idea of the conversion method is to apply the Fisher algorithm to divide the information samples, pair the sample points in two different environments, and then use the geometric mean of the ratio to calculate the environmental factor. Assume that the reliability information samples in environments A and B are represented by $X$ and $Y$, respectively, $X = \{x_1, x_2, \ldots, x_n\}$, $Y = \{y_1, y_2, \ldots, y_m\}$ ($n \geq m$). As shown in Fig. 1, to convert the sample information in environment A into environment B, the specific process of information conversion based on Fisher algorithm is as follows:

1) Rearrange the order of sample information. Rearrange the sample information in the samples $X$ and $Y$ in ascending order:

$$X : x_{(1)} < x_{(2)} < \cdots < x_{(n)},$$
$$Y : y_{(1)} < y_{(2)} < \cdots < y_{(m)}.$$

2) Find center points of sample data. Calculate the center points of sample $X$ and $Y$ respectively according to Eq. (1).

$$|x_{(r)} - \overline{X}| = \min_{1 \leq i \leq n} |x_{(i)} - \overline{X}|,$$
$$|y_{(s)} - \overline{Y}| = \min_{1 \leq j \leq m} |y_{(s)} - \overline{Y}|. \tag{1}$$

$x_{(r)}$ divides the sample $X$ into three parts: $\{x_{(1)}, x_{(2)}, \ldots, x_{(r-1)}\}, \{x_{(r)}\}, \{x_{(r+1)}, x_{(r+2)}, \ldots, x_{(n)}\}$;

$y_{(s)}$ divides the sample $Y$ into three parts: $\{y_{(1)}, y_{(2)}, \ldots, y_{(s-1)}\}, \{y_{(s)}\}, \{y_{(s+1)}, y_{(s+2)}, \ldots, y_{(m)}\}$.

3) Construct sample pairings. If: $n = m$, and $r = s$, construct sample pairing: $(x_{(i)}, y_{(j)}), i = j = 1, 2, \ldots, n$; otherwise, keep sample pairing: $(x_{(r)}, y_{(s)})$, then, the samples are paired in two cases:

(i) For the left sample points of $x_{(r)}$ and $y_{(s)}$, if $r - 1 \geq s - 1$, cluster the sample $X$ based on Fisher, the cluster class is $s - 1$, and the mean value of each class is $y_j^*$, construct sample pairings: $\{y_j^*, y_j\}, 1 \leq j \leq s - 1$; Otherwise, cluster the sample $Y$ based on Fisher, the cluster class is $r - 1$, and the mean value of each class is $y_j^*$, construct sample pairings: $\{y_j^*, y_j\}, 1 \leq i \leq r - 1$.

(ii) For the right sample points of $x_{(r)}$ and $y_{(s)}$, if $n - r \geq m - s$, cluster the sample $X$ based on Fisher, the cluster class is $m - s$, and the mean value of each class is $y_j^*$, construct sample pairings: $\{y_j^*, y_j\}, s + 1 \leq j \leq m$; Otherwise, cluster the sample $Y$ based on Fisher, the cluster class is $n - r$, and the mean value of each class is $y_j^*$, construct sample pairings: $\{y_j^*, y_j\}, r + 1 \leq i \leq n$.

4) Calculate the conversion factor. After pairing the samples with the method in step 3), calculate the conversion factor according to Eq. (2).

$$K = \sqrt[m]{\frac{y_1^*}{y_1} \cdot \frac{y_2^*}{y_2} \cdots \frac{y_m^*}{y_m}}. \tag{2}$$

5) Get the conversion function. The following Eq. (3) can convert the information of the sample $X$ under environment A into the sample $Y$ under environment B, so as to
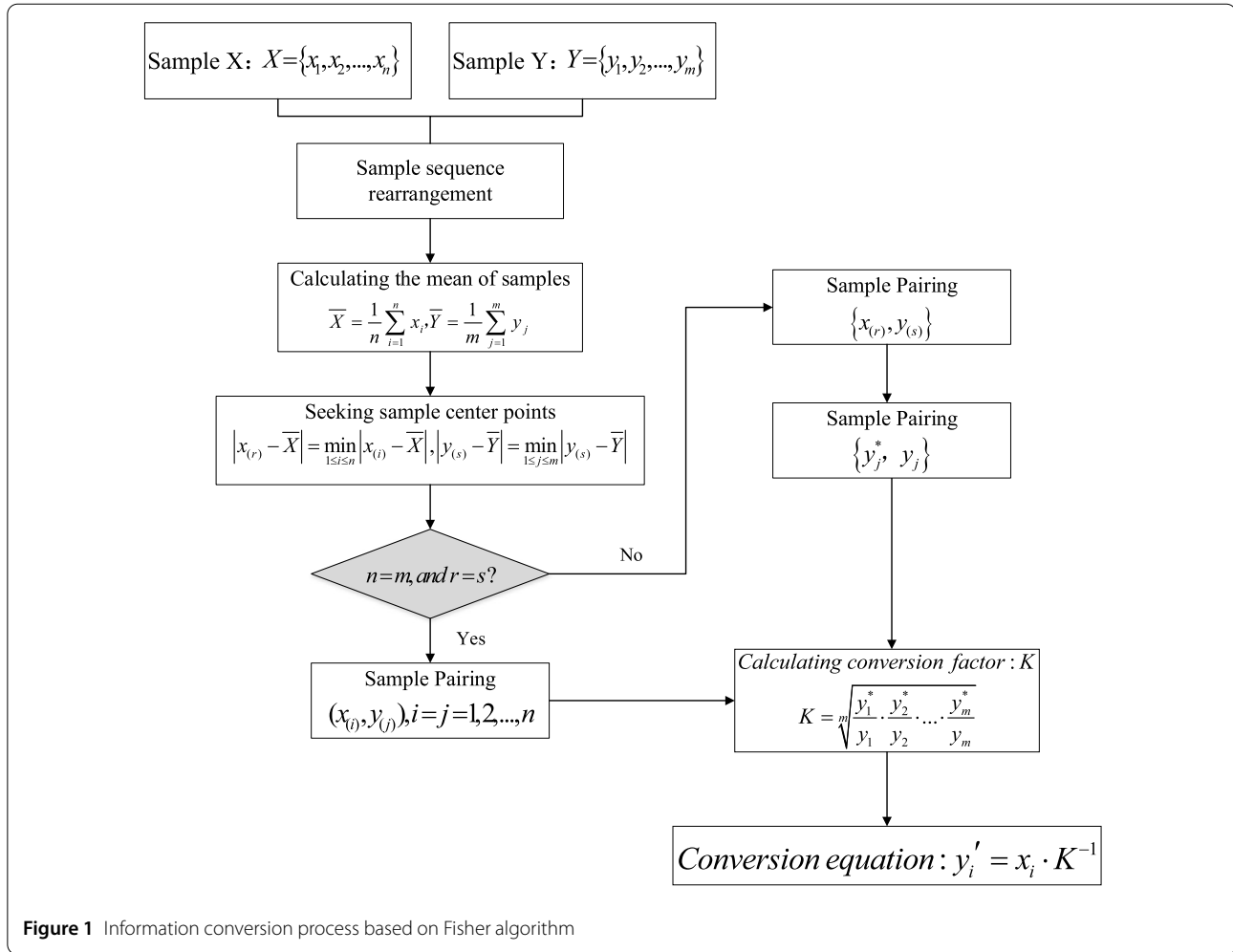
**Figure 1** Information conversion process based on Fisher algorithm

achieve the purpose of expanding the capacity of the sample $Y$.

$$y_i' = x_i \cdot K^{-1}. \tag{3}$$

### 3.2 Compatibility test

After the information conversion through the method in Sect. 3.1, it is necessary to use a certain method to judge whether the conversion result is valid, that is, the reliability information can truly reflect the characteristics of statistical parameters, which requires that the reliability information and the small sample field information of the target mechanical product approximately obey the same distribution, and this goal is achieved through the compatibility test. Wilcoxon rank sum test, Smirno test, Mood test, etc. can all be used to test the compatibility directly using sample data [57]. The rank sum test is adopted in this paper, which can effectively test whether the data from different samples obey the same overall distribution when the overall distribution is not clear. Assume that the reliability information sample is: $X = \{x_1, x_2, \ldots, x_n\}$. The on-site relia-

bility information sample of the target mechanical product is: $Y = \{y_1, y_2, \ldots, y_m\}$, consider a trade-off hypothesis:

Null hypothesis $H_0$: $X$ and $Y$ belong to the same population;

Alternative hypothesis $H_1$: $X$ and $Y$ do not belong to the same population.

Combine the sample information of the two samples into a new sample population, and arrange the new sample information in order from small to large to obtain a new sample: $Z$, $Z = \{z_1, z_2, \ldots, z_{n+m}\}$, $z_1 \leq z_2 \leq \cdots z_{n+m}$. The subscript $i$ is called the rank of the new sample information, $i = 1, 2, \ldots, n + m$. If $x_k = z_j$, $x_k \in \{x_1, x_2, \ldots, x_n\}$, then it is considered that the rank of elements $x_k$ in sample $X$ in the new sample is $j$, and note that: $r(x_k) = j$. The rank sum of sample $X$ is $T = \sum_{k=1}^{n} r(x_k)$.

Based on the rank sum test table, under significance level $\alpha$, if $T$ meets: $T_1(\alpha) < T < T_2(\alpha)$, the null hypothesis should be accepted, otherwise, the null hypothesis should be rejected and the alternative hypothesis need to be accepted. When the values of $n$ and $m$ are both large ($n, m > 10$), the rank sum $T$ approximately obeys normal distribution, that
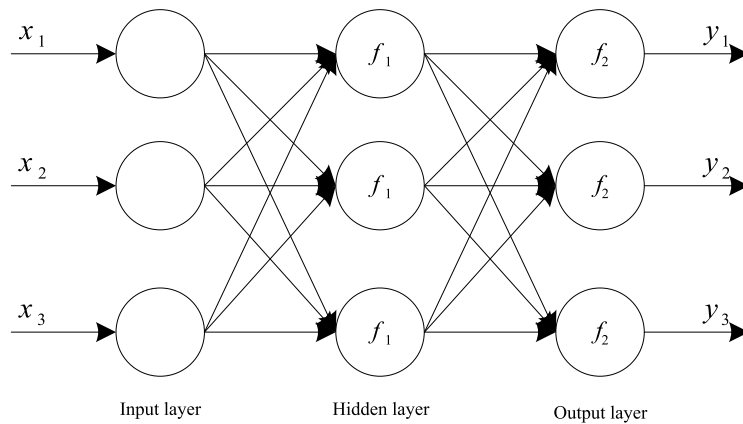
**Figure 2** BP neural network prediction structure

is:

$$T = \sum_{k=1}^{n} r(x_k) \sim N\left( \frac{n(n+m+1)}{2}, \frac{nm(n+m+1)}{2} \right). \quad (4)$$

At this time, the $Z$ test method of normal distribution is used to test the compatibility. Under the significance level $\alpha$, the test rule is as shown in Eq. (2):

$$P\left\{ \left| \frac{T - \frac{n(n+m+1)}{2}}{\sqrt{\frac{nm(n+m+1)}{12}}} \right| < \mu_\alpha | \mathrm{H}_0 \right\} = 1 - \alpha. \quad (5)$$

Here $\mu_\alpha$ represents the $\alpha$ quantile of the standard normal distribution. When Eq. (5) is established, under the confidence: $1 - \alpha$, the null hypothesis will be accepted; otherwise, the two samples are considered not to meet the requirements of the compatibility test.

### 3.3 TBF prediction based on PSO-BP

#### 3.3.1 Prediction model

After the reliability information is converted to the target mechanical product through the above steps, an enlarged data sample is obtained, which is composed of reliability information of different dimensions and source. Since the reliability information of each different source (such as the TBF to be discussed in this paper) generally has the characteristics of time series, the BP Neural Network with better prediction effect on time series data is selected to realize the final prediction of TBF in this paper, to verify the feasibility of the information conversion method based on Fisher algorithm. The prediction model is shown in Eq. (6).

$$y(t) = f(y(t - n), y(t - n + 1), \ldots, y(t - 1)). \quad (6)$$

Here $y(t)$ represents the TBF value at time $t$, $n$ is the number of input neuron nodes of the BP neural network,

that is, the TBF at time $t$ is affected by the previous $n$th times, the specific value of $n$ is determined according to the number of faults. The predicted network structure is shown in Fig. 2.

Here $f$ is the hidden layer, and the number of nodes in the hidden layer of the prediction model is determined by the empirical equation and the trial-and-error method. The empirical equation is defined as Eq. (7).

$$m = \sqrt{n + l} + \alpha. \quad (7)$$

Here $m$ is the number of hidden layer nodes, $n$ is the number of input nodes, $l$ is the number of output nodes, and $\alpha$ is a constant between 1 and 10. It can be seen from the above analysis that the value of $m$ is between 3 and 12. Determining the optimal number of nodes by trial and error method.

BP Neural Network is a multi-layer feedforward neural network. Its weights are adjusted by Back-Propagation learning algorithm, which has good nonlinear mapping ability, generalization ability and fault tolerance ability, and is easy to implement in engineering. But the initial value is randomly assigned, and it is easy to fall into a local minimum [58]. PSO is an algorithm with good global searching ability. It only needs optimized function and does not need other auxiliary information. PSO algorithm is adapted in this paper to find the global optimal particle by updating particle speed and position, so as to optimize the weight and threshold of BP neural network, and then realize the establishment of TBF prediction model based on PSO-BP. The process of PSO-BP prediction model is shown in Fig. 3.

#### 3.3.2 Performance evaluation

The ultimate goal of this paper is to use the above method to predict the TBF of the target mechanical product.
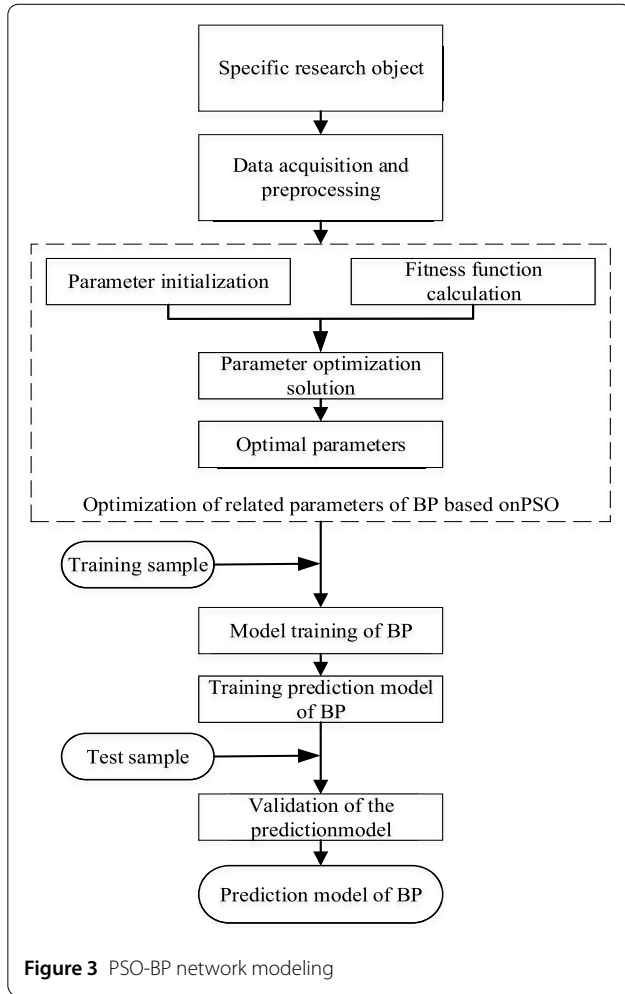
**Figure 3** PSO-BP network modeling

Firstly, the converted reliability information is used as the training set to train the prediction model, and then the TBF in the actual operation stage of the target mechanical product is selected as the test set to test the prediction effect of the model, and two evaluation metrics as shown in Eq. (8): the mean-absolute-percentage-error (MAPE) and the coefficient of determination ($R^2$) are selected to evaluate the performance of the model and verify the feasibility of information conversion method based on Fisher algorithm.

$$\begin{cases} MAPE = \frac{\sum_i |\frac{y_i - \hat{y}_i}{y_i}|}{N}, \\ R^2 = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y}_i)^2}. \end{cases} \tag{8}$$

Here $y_i$ is the true value, $\hat{y}_i$ is the predicted value, $\bar{y}_i$ is the mean value of the true value, and $N$ is the sample size.

## 4 Case study
### 4.1 Case description
Taking the remanufactured CNC milling machine tool (XK6032-01) as an example, due to its single piece produc-

**Table 1** Fault data during actual operation stage

| Target product | TBF(h) | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| XK6032-01 | 421 | 457 | 106 | 575 | 178 | 2390 | 889 | 246 |

tion mode and part blank value, it is impossible for remanufacturing product entities to carry out huge destructive on-site tests. In addition, it is extremely difficult to build the system level simulation test model and platform of personalized remanufacturing products, which is costly and difficult to accurately simulate all working conditions in the process of re-service. Therefore, the target product of this case belongs to a typical small sample. The TBF data (as shown in Table 1) and its related reliability information (TBF) were collected respectively. Perform TBF prediction for the remanufactured CNC milling machine according to the process shown in Fig. 4.

### 4.2 Results
#### 4.2.1 Conversion results of reliability information
The machine tool is used as the target environment, and the reliability information of each source is converted to the working environment of the remanufactured machine tool according to the information conversion method in Sect. 3.1. The compatibility test is carried out respectively with the fault data of the actual operation stage, and finally the expanded fault information table of the CNC machine tool as shown in Table 2 is obtained.

According to the number of faults of Table 2, the number of input neurons in the TBF prediction model based on PSO-BP neural network is determined to be $n = 3$, that is, the TBF at the fourth moment is affected by the TBF at the first three moments. In the process of training the model, the fault data of the first three moments is used as input, and the fault data of the fourth moment is used as output. For information sources with more than 4 failures, the sliding time window is used to obtain the value, the window size is set to 4, and the step size is set to 1. Based on this, according to the order of Table 2, the data from different sources of each dimension is input into the neural network training model in turn. A total of 34 groups of data can be used for model training. Take the data in Table 1 as the test set (5 groups in total) to test the trained model, and the prediction performance of this model is measured by the metrics in Eq. (8).

#### 4.2.2 Prediction results
In this prediction, the neural network toolbox in Matlab is used for network training. The prediction model reaches convergence after 378 iterations. The prediction performance of the training set is shown in Fig. 5, and the prediction performance of the test set is shown in Fig. 6.

As can be seen from Fig. 5, the predicted value of each point in the training set almost completely coincides with
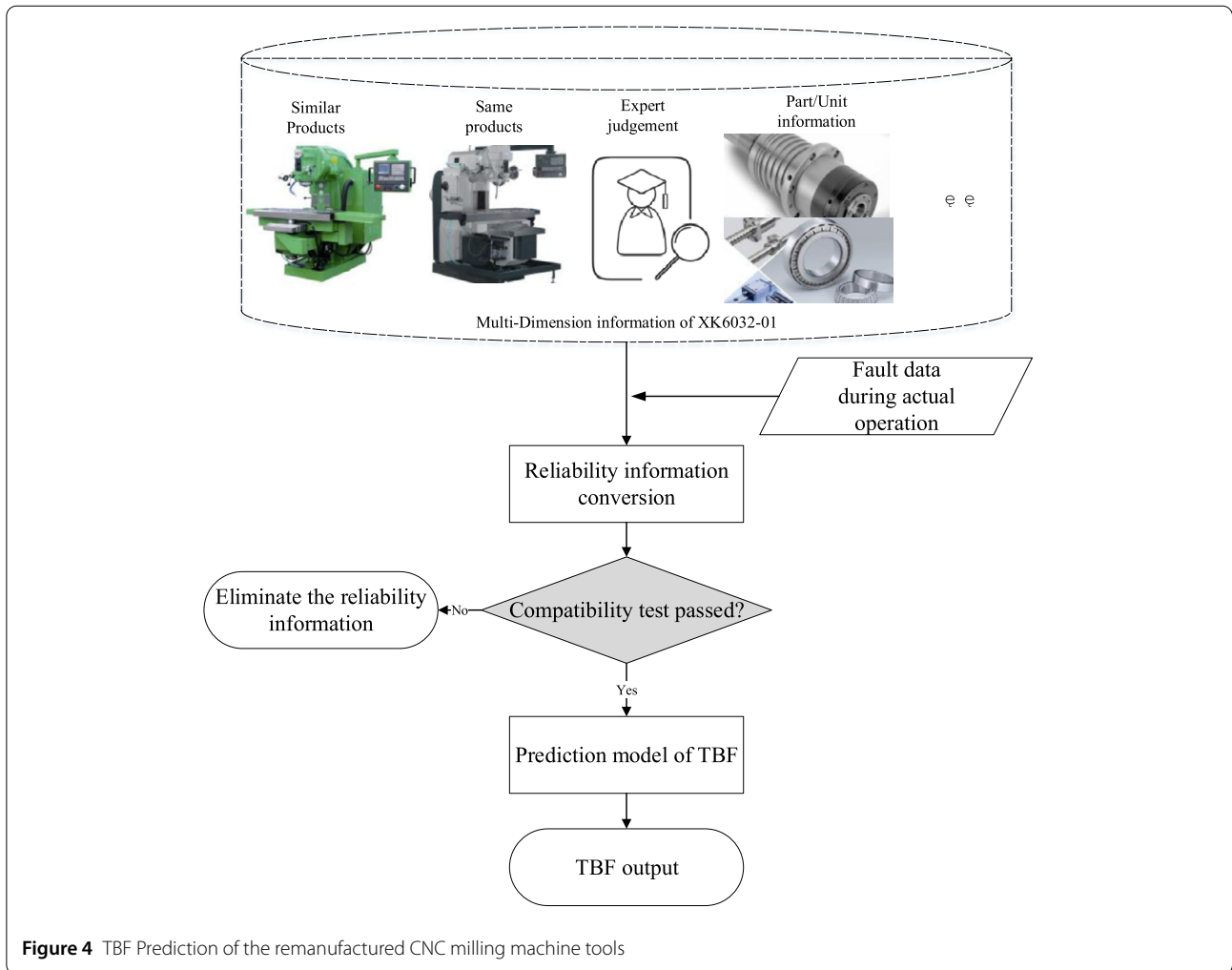
**Figure 4** TBF Prediction of the remanufactured CNC milling machine tools

**Table 2** Fault data after converted (TBF)

| Serial number | Information source | Conversion results(h) | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 1 | Workbench | 35 | 252 | 347 | 467 | | | | |
| 2 | Servo unit | 23 | 786 | 236 | 15 | | | | |
| 3 | Spindle | 124 | 535 | 516 | 2725 | | | | |
| 4 | Control system | 13 | 263 | 134 | 125 | 463 | 571 | | |
| 5 | Similar MT 1 | 478 | 486 | 85 | 542 | 186 | 2486 | 934 | 235 |
| 6 | Similar MT 2 | 524 | 876 | 35 | 1896 | 376 | 1288 | 1045 | |
| 7 | Similar MT 3 | 496 | 480 | 64 | 530 | 192 | 2576 | 928 | 240 |
| 8 | Same MT 1 | 698 | 976 | 495 | 1056 | 964 | 1275 | | |
| 9 | Same MT 2 | 153 | 1384 | 356 | 89 | 1768 | 2757 | | |
| 10 | Same MT 3 | 31 | 458 | 356 | 976 | 275 | 1897 | 243 | 1125 |
| 11 | Trial phase | 528 | 94 | 56 | 15 | 2451 | 1763 | | |

*Note*: Similar MT stands for similar remanufacturing CNC machine tools, and Same MT stands for same remanufacturing CNC machine tools.

the true value, and the error is almost zero, indicating that the TBF prediction model based on PSO-BP has a very good prediction effect on the training set.

It can be seen from Fig. 6 that the change trend of the true value of the test set and the predicted value are basically the same, indicating that the prediction model has a good prediction effect on the change trend of TBF. The specific prediction error is shown in Table 3.

It can be seen from Table 3 that, except for test point 3, the absolute value of the relative error of the other points
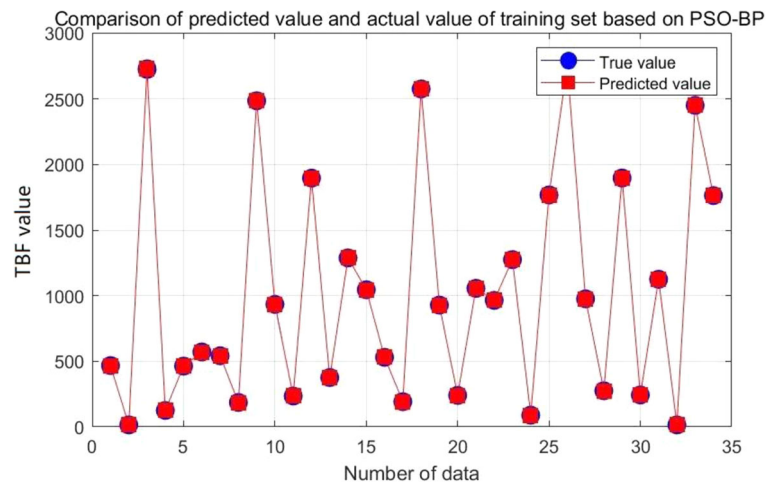
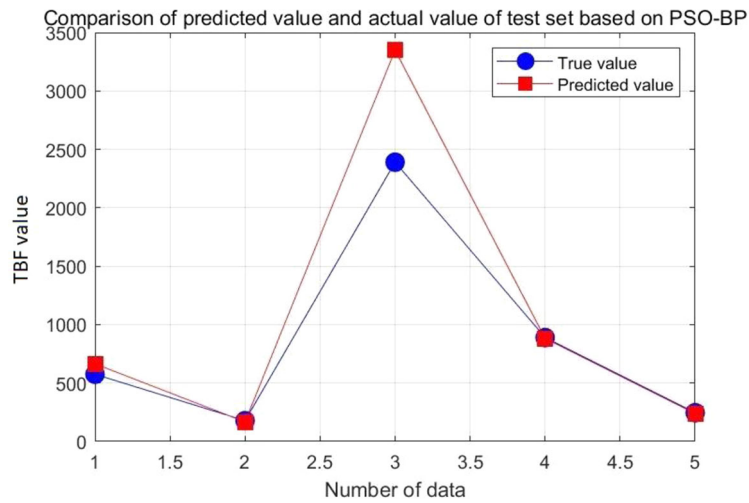**Figure 5** Comparison of the predicted value of the training sample with the true value



**Figure 6** Comparison of the predicted value of the test sample with the true value

**Table 3** The relative error of each point in the test set

|  | Test point 1 | Test point 2 | Test point 3 | Test point 4 | Test point 5 |
|---|---|---|---|---|---|
| True value | 575 | 178 | 2390 | 889 | 246 |
| Predicted value | 604.20 | 167.29 | 3301.87 | 881.51 | 238.69 |
| Relative error | −0.0508 | 0.0602 | −0.3815 | 0.0084 | 0.0297 |

is less than 10%, which is within the allowable error range. At this time, the mean-absolute-percentage-error of the model is 0.1061, in general, if the MAPE is less than 10%, the model prediction accuracy is high; the coefficient of determination of the model is 0.9902, in general, $R^2$ is within $[0, 1]$, and the value is closer to 1, the higher the prediction accuracy of the model will be. It can be seen that the pre-

diction accuracy of the model is high, which proves that the proposed method has a good prediction effect on TBF.

### 4.3 Discussion

In order to verify whether PSO-BP neural network model is superior to other intelligent models in predicting the time series data after sample expansion, standard BP neural network and GA-BP neural network are selected for comparative analysis. The experimental results are shown in Table 4.

To express the prediction accuracy and convergence speed of each model more intuitively, the above results are displayed in the form of bar graph, as shown in Fig. 7. Because the iteration times are dimensional parameters,
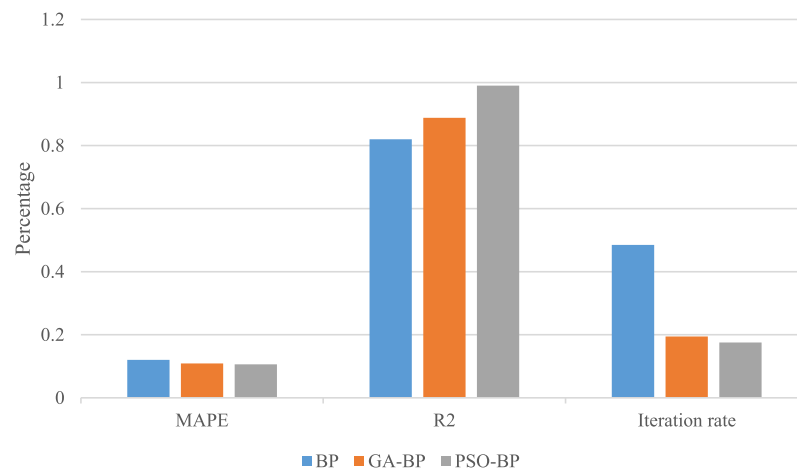
**Figure 7** Performance comparison of prediction models

**Table 4** Performance comparison of prediction models

| Metrics | BP | GA-BP | PSO-BP |
|---------|--------|---------|--------|
| MAPE | 0.12038 | 0.10893 | 0.1061 |
| $R^2$ | 0.81995 | 0.88784 | 0.9902 |
| Epoch | 2425 | 973 | 878 |

they are processed to calculate the iteration rate, that is, the ratio of iteration times to total times.

Generally, the smaller the MAPE, the closer to 1 of $R^2$, the smaller the iteration rate, the higher the prediction accuracy of the model, and the faster the convergence speed. It can be seen from Fig. 7 that the performance of PSO-BP is better than the other two intelligent models, indicating that this model is more suitable for the prediction of time series data with small sample after the expansion of sample size.

Through the above analysis, some advantages of the proposed approach can be summarized as follows. 1) The information conversion method based on Fisher algorithm converts the multi-source reliability information into the target product itself, overcomes the influence of the specific distribution of the reliability information, expands the sample size of complex mechanical products. 2) The TBF prediction model based on PSO-BP neural network avoids the problem of data samples affected by certain errors after capacity expansion, has high accuracy, and provides method support for PdM of complex mechanical products.

However, the proposed reliability information conversion method based on the Fisher algorithm has some problems and deficiencies in the process of converting experimental information. The basic idea of the Fisher algorithm is similar to dichotomy, that is, it needs to divide all data into two categories firstly, and then divide one of them into two categories too, recursive in turn. Therefore, the clas-

sification result obtained by this method is a local optimal solution, which has a certain impact on the accuracy of environmental factors.

## 5 Conclusions

Aiming at the problems of reliability information collection difficulty, high collection cost and small data samples in the process of TBF prediction of complex mechanical products, a machine learning-based TBF prediction model with reliability information conversion is proposed in this paper. The information conversion method based on the Fisher algorithm is used to deeply mine the reliability information of multiple-source reliability information, convert the multi-source reliability information into the target product itself, and expand the sample size of small sample target products. For data samples with certain errors after capacity expansion, BP neural network with good prediction effect on time series data is selected to realize the final prediction of TBF, and PSO algorithm is used to optimize the initial weight and threshold of BP neural network to avoid local extreme value and improve the convergence speed. Taking a remanufactured machine tools as an example to verify the proposed approach. The results show that the TBF prediction model based on PSO-BP proposed in this paper has a good prediction effect on the sample data after the reliability information conversion and expansion using the Fisher algorithm, and provides a strong basis for the PdM and reliability growth of remanufactured machine tools. In the future work, we should consider an optimization method to get the global optimal solution of data segmentation and pairing, improve the sample pairing effect of data, and improve the accuracy of conversion.

## Declarations

**Conflicts of interest**
The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Competing interests**
The authors declare that they have no competing interests.

**Author contribution**
All authors contributed to the study conception and design. Conceptualization, methodology, data collection and analysis were performed by HZ, XH and WY. Review & Editing and supervision were performed by ZJ and SZ. The first draft of the manuscript was written by HZ, XH and WY and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

**Author details**
[1]Hubei Key Laboratory of Mechanical Transmission and Manufacturing Engineering, Wuhan University of Science and Technology, Wuhan, 430081, China. [2]Key Laboratory of Metallurgical Equipment and Control Technology, Ministry of Education, Wuhan University of Science and Technology, Wuhan, 430081, China. [3]Academy of Green Manufacturing Engineering, Wuhan University of Science and Technology, Wuhan, 430081, China. [4]School of Automobile and Traffic Engineering, Wuhan University of Science and Technology, Wuhan, 430081, China. [5]Department of Mechanical Engineering, School of Engineering, Cardiff University, Cardiff, CF24 3AA, UK.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

### References

1. J. Yan, Y. Meng, L. Lu et al., Industrial Big Data in an Industry 4.0 Environment: Challenges, Schemes and Applications for Predictive Maintenance (IEEE Access, 2017), pp. 1–1
2. A.J. Guillén, A. Crespo, M. Macchi et al., On the role of prognostics and health management in advanced maintenance systems. Prod. Plan. Control 27(12), 991–1004 (2018)
3. S. Sule, Predictive maintenance, its implementation and latest trends. J. Eng. Manuf. 231(9), 1670–1679 (2017)
4. Q. Liu, D. Ming, W. Lv et al., Manufacturing system maintenance based on dynamic programming model with prognostics information. J. Intell. Manuf. 30(3), 1155–1173 (2017)
5. C. Chen, Y. Liu, S.X. Wang et al., Predictive maintenance using Cox proportional hazard deep learning. Adv. Eng. Inform. 44, 101054 (2022)
6. P.H. Cui, J.Q. Wang, W.P. Zhang et al., Predictive maintenance decision-making for serial production lines based on deep reinforcement learning. Comput.-Integr. Manuf. Syst. 27(12), 3416–3428 (2021)
7. J. Padgett, M. Sanchez-Silva et al., Maintenance and operation of infrastructure systems: review. J. Struct. Eng. 142(9), F4016004 (2016)
8. K. Wang, W. Yi, *How AI Affects the Future Predictive Maintenance: A Primer of Deep Learning* (2018)
9. T.P. Carvalho, A.A. Fabrízzio, M.N. de Soares, R. Vita et al., A systematic literature review of machine learning methods applied to predictive maintenance. Comput. Ind. Eng. 137, 106024 (2019)
10. Z. Tiago, C.A. Da Costa, R. Da Rosa Righi et al., Predictive maintenance in the industry 4.0: a systematic literature review. Comput. Ind. Eng. 150, 106889 (2020)
11. C. Kong, Review on advanced health monitoring methods for aero gas turbines using model based methods and artificial intelligent methods. Int. J. Aeronaut. Space Sci. 15(2), 123–137 (2014)
12. K.P. Boshnakov, V.I. Petkov, L.A. Doukovska et al., Predictive maintenance model based approach for objects exposed to extremely high temperatures, in *Signal Processing Symposium* (2013)
13. J. Ortiz, R.A. Carrasco, Model-based fault detection and diagnosis in ALMA subsystems: observatory operations: strategies, in *Processes & Systems VI* (2016)
14. B. Lung, M. Monnin, A. Voisin et al., Degradation state model-based prognosis for proactively maintaining product performance. CIRP Ann. 57(1), 49–52 (2008)
15. Y. Lei, N. Li, S. Gontarz et al., A model-based method for remaining useful life prediction of machinery. IEEE Trans. Reliab. 65(3), 1314–1326 (2016)
16. R. Magargle, L. Johnson, P. Mandloi et al. A simulation-based digital twin for model-driven health monitoring and predictive maintenance of an automotive braking system. Proceedings of the 12th International Modelica Conference, Prague, Czech Republic, May 15–17, 2017
17. Basic T, Distribution W.9. The Weibull Distribution (2009)
18. J.W. Hu, P. Chen, Predictive maintenance of systems subject to hard failure based on proportional hazards model. Reliab. Eng. Syst. Saf. 196, 106707 (2020)
19. X.J. Zhou, F.X. Li, L. Jay, A Reliability-Based Sequential Preventive Maintenance Model. J. Shanghai Jiaotong Univ. (2005)
20. C. Li, Y. Zhang, M. Xu, Reliability-based maintenance optimization under imperfect predictive maintenance. Chin. J. Mech. Eng. 25(1), 160–165 (2012)
21. J. Shimada, S. Sakajo, A statistical approach to reduce failure facilities based on predictive maintenance, in (International Joint Conference on Neural Networks) (2016)
22. J. Daily, J. Peterson, *Predictive Maintenance: How Big Data Analysis Can Improve Maintenance. Supply Chain Integration Challenges in Commercial Aerospace* (Springer, Berlin, 2017), pp. 267–278
23. W.J.C. Verhagen, R. Curran, L.W.M. De Boer, Component-Based Data-Driven Predictive Maintenance to Reduce Unscheduled Maintenance Events (2017)
24. M. Baptista, S. Sankararaman, I.P.D. Medeiros et al., Forecasting fault events for predictive maintenance using data-driven techniques and ARMA modeling. Comput. Ind. Eng. 115, 41–53 (2018)
25. J. Zenisek, F. Holzinger, M. Affenzeller, Machine learning based concept drift detection for predictive maintenance. Comput. Ind. Eng. 137, 106031 (2019)
26. J.J. Zhang, P. Wang, R.Q. Yan et al., Long short-term memory for machine remaining life prediction. J. Manuf. Syst. 48, 78–86 (2018)
27. A.V. Horenbeek, L. Pintelon, A dynamic predictive maintenance policy for complex multi-component systems. Reliab. Eng. Syst. Saf. 120, 39–50 (2013)
28. M. Baptista, S. Sankararaman, I.P. de Medeiros et al., Forecasting fault events for predictive maintenance using data-driven techniques and ARMA modeling. Comput. Ind. Eng. 15, 41–53 (2018)
29. S. Rocco, M. Claudio, Singular spectrum analysis and forecasting of failure time series. Reliab. Eng. Syst. Saf. 114, 126–136 (2013)
30. M.C. Moura, E. Zio, I.D. Lins et al., Failure and reliability prediction by support vector machine. Reliab. Eng. Syst. Saf. 96(11), 1527–1534 (2011)
31. I. Lazakis, Y. Raptodimos, T. Varelas, Predicting ship machinery system condition through analytical reliability tools and artificial neural networks. Ocean Eng. 152, 404–415 (2018)
32. J. Jana, S. Radka, Finite-sample density and its small sample asymptotic approximation. Stat. Probab. Lett. 81, 1311–1318 (2011)
33. B.S. Reliability, Evaluation of the servo turret with accurate failure data and interval censored data based on EM algorithm. J. Mech. Sci. Technol. 34, 1503–1513 (2020)
34. T. Li, W.H. Zhang, S.J. Zhang et al., Research-on gray-scale prediction model of product failure interval time based on genetic algorithm weight. Modem Manuf. Eng. 9, 101–106 (2021)
35. K. Li, X.M. Shi, J. Li et al., Bayesian estimation of ammunition demand based on multinomial distribution. Discrete Dyn. Nat. Soc. 2021, 5575335 (2021)
36. J. Lenhard, A transformation of Bayesian statistics: computation, prediction, and rationality. Stud. Hist. Philos. Sci. 92, 144–151 (2022)

37. A.K. Mandal, S. Rakshit, C.S. Stalin et al., Estimation of the size and structure of the broad line region using Bayesian approach. Mon. Not. R. Astron. Soc. **502**(1), 2140–2157 (2021)

38. E. Ml, C. Sjkab, A. Rl et al., Long-term forecasting of El Nio events via dynamic factor simulations. J. Econom. **214**(1), 46–66 (2020)

39. B. Zeng, H. Duan, Y. Bai et al., Forecasting the output of shale gas in China using an unbiased grey model and weakening buffer operator. Energy **151**, 238–249 (2018)

40. Y.C. Hu, Constructing grey prediction models using grey relational analysis and neural networks for magnesium material demand forecasting. Appl. Soft Comput. **93**, 106398 (2020)

41. C. Wang, G.G. Yen, M. Jiang, A grey prediction-based evolutionary algorithm for dynamic multiobjective optimization. Swarm Evol. Comput. **56**(1), 100695 (2020)

42. L. Ye, X.T. Xia, C. Zhen et al., Evaluation of dynamic uncertainty of rolling bearing vibration performance. Math. Probl. Eng. **2**, 1–17 (2019)

43. H.L. Suo, J.M. Gao, Z.Y. Gao et al., Bootstrap and maximum entropy based small-sample product lifetime probability distribution. IFAC-PapersOnLine **48**(3), 219–224 (2015)

44. S. Marelli, B. Sudret, An active-learning algorithm that combines sparse polynomial chaos expansions and bootstrap for structural reliability analysis. Struct. Saf. **75**, 67–74 (2018)

45. A.A. Mikshowsky, D. Gianola, K.A. Weigel, Improving reliability of genomic predictions for Jersey sires using bootstrap aggregation sampling. J. Dairy Sci. **99**(5), 3632–3645 (2016)

46. H.Y. Li, *The Decision Making on Scheduled Maintenance of Civil Aircraft Based on the Reliability* (Nanjing University of Aeronauticsand Astronautics, Nanjing, 2016)

47. H.M. Zhong, Y.N. Liang, S.L. He et al., Construction of a knowledge base for log interpretation using KNN-Fisher. Geophys. Prospect. Pet. **60**(3), 395–402 (2021)

48. N.K. Ahmed, A.F. Atiya, N.E. Gayar et al., An empirical comparison of machine learning models for time series forecasting. Econom. Rev. **29**(5–6), 594–621 (2010)

49. J.F. Yang, Y.J. Zhai, D.F. Wang et al., Time series prediction based on support vector regression. Proc. CSEE **25**(017), 110–114 (2005) (in Chinese)

50. G. Bontempi, S.B. Taieb, Y.A. Le Borgne, *Machine Learning Strategies for Time Series Forecasting//European Business Intelligence Summer School* (Springer, Brussels, 2012), pp. 62–77

51. C. Wan, W.Z. Li, W.X. Ding et al., A multivariate time series forecasting algorithm based on self-evolution and pre-training. Chinese J. Comput. **45**(3), 513–525 (2022)

52. Z. Dai, Z. Li, Y. Jiao et al., Reliability assessment based on BP neural network for relay protection system with a few failure data samples. Electr. Power Autom. Equip. **34**(11), 129–134 (2014)

53. J.W. Wang, J. Wang, S. Liu et al., Construction period prediction of open-cut subway station based on PSO-BP. J. Civ. Eng. Manag. **39**(2), 7–18 (2022)

54. Y. Shang, L.Y. Kang, M.X. Zhang et al., Prediction method of electricity stealing behavior based on multi-dimensional features and BP neural network. Energy Rep. **8**(4), 523–531 (2022)

55. H.J. Cao, L. Liu, B. Wu et al., Process optimization of high-speed dry milling UD-CF/PEEK laminates using GA-BP neural network. Composites, Part B, Eng. **221**, 1359–8368 (2021)

56. B. Bai, J.Y. Zhang, X. Wu et al., Reliability prediction-based improved dynamic weight particle swarm optimization and back propagation neural network in engineering systems. Expert Syst. Appl. **177**, 114952 (2021)

57. B. Wang, P. Jiang, B. Guo, Reliability evaluation of aerospace valves based on multi-source information fusion. Acta Armament. **43**(1), 199–206 (2022)

58. T. Liu, S. Yin, An improved particle swarm optimization algorithm used for BP neural network and multimedia course-ware evaluation. Multimed. Tools Appl. **76**(9), 11961–11974 (2017)