Neural Computing and Applications (2022) 34:3399–3410
https://doi.org/10.1007/s00521-021-05716-1

SPECIAL ISSUE ON MULTI-MODAL INFORMATION LEARNING AND ANALYTICS ON BIG DATA

Condition monitoring and life prediction of the turning tool based on extreme learning machine and transfer learning

Zhan Gao[1] · Qiguo Hu[1] · Xiangyang Xu[1]

Received: 25 October 2020 / Accepted: 5 January 2021 / Published online: 27 May 2021

Abstract
When the turning tool has worn and failed but the failure is not found, if it continues to be used for processing, it will break, and cause the workpiece to be scrapped, and even damage the machine tool. In order to avoid the loss caused by turning tool wear, the remaining useful life (RUL) prediction of turning tool wear has become a hot research topic in recent years. For RUL prediction in turning tools, the traditional machine is difficult to acquire sufficient degradation data and inconsistent data distribution among different turning tools in engineering, and they cannot provide better prediction accuracy to some extent. To solve the above problems, this paper proposes a multi-granularity feature extraction (MGFE) method based on the gray-level co-occurrence matrix (GLCM) and random forest (RF). Moreover, a health indicator (HI) of turning tools in the source domain was obtained. The common representative features in HI sequence of target domain was transferred to source domain and builds the condition monitoring and life prediction system of turning tools based on extreme learning machine and transfer learning. Finally, extreme vector machine (ELM) is used to construct the RUL prediction model. The research results show that the model constructed in this paper is effective in RUL prediction and can significantly improve the prediction accuracy of remaining useful life.

Keywords Extreme learning machine · Transfer learning · Turning tool · Condition monitoring · Life prediction

1 Introduction

At this stage, most manufacturers have two ways to determine the failure of the turning tool: manual experience and supplier recommendations. Generally speaking, as the turning tool wears, many factors will change accordingly, including the vibration generated by the machine tool, the sound made during cutting, the shape of the chip, the power of the machine tool, etc. [1]. At this time, experienced operators can determine whether the turning tool is invalid based on these changes. This way of judging easily causes the following two problems. One is that the tool change is too early. In this case, the turning tool is judged to be invalid before it reaches its service life, which increases production costs for the company. The second is to change the knife too late. In this case, the turning tool is already in a failed state, but it is still used for cutting, which may cause insufficient surface accuracy or damage to the parts. In particular, in the aerospace field, the cost of this situation will be very high, and will seriously affect the production efficiency and economic benefits of the enterprise [2]. At present, domestic manufacturing enterprises have problems of low integration and low efficiency in turning tool management [3]. However, under the promotion of digital processing technology, foreign countries have made great progress in this field, and established a turning tool management system based on failure prediction and health management and enterprise resource planning. Moreover, it is widely used in large enterprises such as Boeing and Airbus, and its degree of intelligence and integration is relatively high [4].

✉ Zhan Gao
gemini_gz@sina.com

✉ Qiguo Hu
swpihqg@cqjtu.edu.cn

[1] Department of Mechatronics and Vehicle Engineering, Chongqing Jiaotong University, Chongqing 400074, China

As the key basic component of the milling machine, the turning tool is the execution end of the milling operation. It is one of the key factors to ensure processing quality, increase production efficiency, and reduce production energy consumption and time costs. However, it is also the most susceptible component to damage and waste. Studies have shown that turning tool damage is the primary factor causing failures in the milling process. The downtime caused by it accounts for 7–20% of the total downtime of the milling machine, and the turning tool and tool change costs account for 3–12% of the total processing cost. Moreover, the indirect costs related to the turning tool management, procurement, inspection, and grinding can be more than 4 times the cost of the turning tool, and its impact on production efficiency can reach 25%. Affected by the complex milling process of discontinuous cutting, the milling cutter is subjected to complex loads such as high pressure, high temperature, vibration, and mechanical shock during the milling process, which makes the turning tool gradually wear, and leads to degradation of cutting performance, and directly affects the processing quality. Once the turning tool wear reaches a certain level, the processing quality will seriously decline. At this time, the turning tool has reached the end of its service life, so the turning tool needs to be changed in time. The life change of turning tool is a minute or hour level change, so the timeliness of turning tool status monitoring is critical. The traditional turning tool replacement strategy is basically an empirical pre-set life index of the turning tool. The life index is expressed by the cutting time or the number of processed pieces. Once the life index is reached, the tool is forced to change. This tool change strategy is simple to implement, but the life cycle has not undergone scientific data analysis, so it is difficult to accurately predict the life of the turning tool [5].

In recent years, domestic and foreign scholars have done a lot of research work on turning tool status recognition and monitoring, and proposed many more effective diagnostic methods, such as recurrent neural network [6], hidden Markov chains [7], relevance vector machines [8], and hybrid networks [9]. This provides a rich technical foundation for TCM research in the milling process.

2 Related work

According to statistics on the 40–50 reasons for the breakage of the turning tool every day in the large machining center workshop, literature [10] found that most of the breakage is caused by the serious wear of the turning tool and not replaced in time. On the other hand, if the tool change time is too early, the turning tool will be replaced before it reaches the blunt standard, which will not only cause waste of the turning tool, but also increase the downtime and affect the processing efficiency. According to statistics, only 50–80% of the effective life of the turning tool is used in engineering applications [11]. Therefore, how to effectively perform the turning tool condition monitoring (TCM) of the milling process, accurately and timely identify the damage degree of the turning tool and monitor the running condition of the turning tool has become an urgent problem to be solved in the intelligent development of milling processing, and it is also one of the main directions of current intelligent processing technology [12].

Foreign scholars have done a lot of research in this field. The literature [13] carried out wavelet analysis on the processing signal, filtered out the eigenvectors such as the root-mean-square value, trained the neural network, and better predicted the remaining service life of the turning tool. The literature [14] used force signal, vibration signal, power signal and sound signal to evaluate the remaining service life of the turning tool under the same model, and found that the effect obtained from the force signal was the best. The literature [15] analyzed the force signal and vibration signal in the milling process in the frequency domain and time–frequency domain, and judged the wear state of the turning tool, and found that the force signal prediction result is better than the vibration signal. The literature [16] selected nine feature quantities based on the signals collected during the experiment and analyzed them, and achieved good prediction results. The literature [17] used Fourier transform to analyze the feature quantity from the force signal collected during turning. The literature [18] extracted the four coefficients of the cutting force signal from the calculation model as feature quantities to predict the service life of the turning tool. Some scholars have also carried out corresponding research. The literature [19] used acoustic emission sensors to monitor the wear status of the turning tool during the cutting process, used wavelet analysis to decompose the signal, and predicted the wear value of the turning tool. The literature [20] analyzed the influence of cutting parameters on the surface roughness of parts by using cutting force signals, and made predictions based on neural networks. Its prediction accuracy is higher than the empirical formula.

During the cutting process of the turning tool, the wear rate does not change linearly with the length of the cutting time, but is affected by a variety of factors, including cutting parameters, workpiece material, and machine tool stiffness. Therefore, the prediction model should have nonlinear prediction performance. PSO-BP neural network fusion prediction algorithm includes BP neural network and particle swarm algorithm. The former has strong nonlinear mapping ability and can perform regression fitting on complex curves with high accuracy, but it performs

poorly at some local extreme points. The latter has strong global optimization performance, can quickly process local extreme points, and has a faster convergence rate. The fusion of the two algorithms can achieve accurate prediction of the life of the turning tool [21]. The literature [22] extracted the characteristic quantities in the cutting force signal, used the PSO-BP model to predict the life of the turning tool, established an objective optimization function, and recommended cutting parameters for lower surface roughness. The literature [23] took the temperature during the cutting process as the characteristic quantity, and established a turning tool wear prediction model based on the PSO-BP fusion algorithm. The prediction accuracy is above 80%. The literature [24] established a turning tool life prediction model, and used a variety of algorithms to predict the life of the turning tool. It is found that the prediction accuracy of the PSO-BP algorithm is the highest. On this basis, considering factors such as cutting parameters, the PSO-BP model is used to predict the remaining service life of the coating turning tool, and good results have been achieved.

The literature [25] collected the time series signals of force and torque, extracted their feature quantities using wavelet analysis, and then estimated the health status of the turning tool based on the HMM model, and achieved good results. Moreover, it used the model to estimate the change characteristics of the force signal, and used wavelet analysis to process the local values, and effectively excavated the wear status information of the current stage of the turning tool. The literature [26] used this model to predict the wear state and remaining service life of the micro-milling cutter during the cutting process, and established a decision-making model. After that, it analyzed the force signal and vibration signal in the milling process in the time–frequency domain, and used the characteristic quantity to predict the wear status of the turning tool on-line based on the HMM model, and achieved good application value. The literature [27] collected the force signal and vibration signal in the high-speed turning process, and used wavelet analysis to extract the characteristics of the signal. As a result, 72 characteristic quantities were screened out. After that, it used the SVR model to predict the surface roughness and turning tool life. The results show that the prediction accuracy has reached more than 75%, so better performance has been achieved, and it is found that the performance of the SVR kernel function is better than other model kernel functions. The literature [28] used the SVR model to predict the wear value of the turning tool. The root-mean-square error between the predicted value and the measured value is 0.0229 mm, and the square correlation coefficient is 0.9628, so the effect is ideal.

3 GLCM

The GLCM is based on estimating the second-order combination conditional probability density of the image, and the texture characteristics are expressed by the gray distribution law between adjacent pixels in the image. Therefore, the GLCM can reflect the texture characteristics of the image to a certain extent. For the convenience of subsequent calculations, we denote the GLCM as W matrix.

In a two-dimensional rectangular image with gray level G, we assume that the number of pixels in the horizontal direction x is N_x, and the number of pixels in the vertical direction y is N_y.

The horizontal spatial domain is:

$$X = \{1, 2, \ldots, N_x\} \tag{1}$$

The vertical spatial domain is:

$$Y = \{1, 2, \ldots, N_y\} \tag{2}$$

The quantized gray scale set is:

$$N = \{1, 2, \ldots, G\} \tag{3}$$

Then, a function mapping relationship such as $f : X \times Y \rightarrow N$ can be used to represent the image.

In the above rectangular image, we arbitrarily select a point pair (a and b are not equal to 0): (x, y) and $(x + a, y + b)$. Then, we set i to be the gray value of point (x, y), and j to be the gray value of point $(x + a, y + b)$. Then, we can get that the gray value of the point pair is (i, j). Then, we fix a and b, and only make x and y in the point (x, y) to take different values in the corresponding space domain to obtain the corresponding (i, j) value of each point pair. Because $G(g)$ is the gray level of the image, there are G^2 combinations of i and j. Finally, we count the frequency of each combination in the entire image as $P(i, j, d, \theta)$. Then, the square matrix $[P(i, j, d, \theta)]_{G\times G}$ is the GLCM, namely:

$$W = [P(i, j, d, \theta)]_{G\times G} \tag{4}$$

The mathematical definition of the GLCM is to start from the image (x, y) with the image gray value i, and count the pixels $(x + a, y + b)$ with the distance $d = \sqrt{a^2 + b^2}$ and the gray value j, and its frequency of appearance is $P(i, j, d, \theta)$. The mathematical expression is:

$$P(i, j, d, \theta) = \left\{ \left[(x, y), (x + a, y + b) \middle| \begin{matrix} f(x, y) = i; \\ f(x + a, y + b) = j \end{matrix} \right] \right\} \tag{5}$$

Among them, θ is the generation direction of the GLCM, which is usually set to $0^\circ, 45^\circ, 90^\circ$ and 135°. The values of a and b should be selected according to the

characteristics of the texture itself. The size contrast of the diagonal of the W matrix and the elements on both sides of the diagonal is a direct reflection of the speed of the texture change.

3.1 Features of GLCM

The GLCM matrix has three main characteristics:

(1) W matrix is a symmetric square matrix

Because W is a symmetric square matrix, the following rules exist when θ changes:

$$\begin{aligned} W(d,0^\circ) &= W^T(d,180^\circ) \\ W(d,45^\circ) &= W^T(d,135^\circ) \\ W(d,90^\circ) &= W^T(d,270^\circ) \end{aligned} \tag{6}$$

and

$$W(d,135^\circ) = W^T(d,135^\circ) \tag{7}$$

Therefore, according to this rule, if we want to get the value of $W(d,\theta)$ in the entire θ coordinate space, we only need to calculate the $W(d,\theta)$ when θ takes $0^\circ, 45^\circ, 90^\circ$ and 135°.

(2) d and g and θ are the three influencing factors of W matrix.

It can be seen from $[P(i,j,d,\theta)]_{G\times G}$ that the dimension of the W matrix is directly related to the gray level g of the image. From the mathematical formula of the W matrix, we can see that the values of θ and d will also directly affect the size of the W matrix. Therefore, before we use the W matrix to extract the texture of the image, the first thing to do is to analyze the texture characteristics of the image itself, and then select θ, d and g suitable for the image, and finally solve the W matrix.

(3) The distribution of the element values of the W matrix indirectly reflects the amount of information and the roughness of the texture of the image.

If the nonzero elements of W are mainly concentrated on the main diagonal, it means that the image in the direction θ has less information and the texture is rougher. If it is contrary to the above hypothesis, it means that the image in the direction θ has a larger amount of information and the texture is more detailed. In fact, the amount of information can also indirectly reflect the contrast of the corresponding area image. If the amount of information is more, the contrast is high, and vice versa.

3.2 Characteristic parameters of GLCM

The normalization can be achieved by the following formula:

$$P(i,j,d,\theta) = P(i,j,d,\theta)/R \tag{8}$$

Among them, R is the normalization constant, equal to the sum of all elements in the gray-level co-occurrence matrix.

Haralick et al. defined 14 characteristic parameters that can be used for texture analysis by analyzing the gray-level co-occurrence matrix in related research documents. The calculation methods and main meanings are as follows.

(1) Angular second moment

$$W_1 = \sum_{i=1}^{g}\sum_{j=1}^{g} P^2(i,j,d,\theta) \tag{9}$$

The second moment of angle $W1$ is also called energy. It can be seen from the above formula that the size of $W1$ is mainly determined by the size of the W matrix, and the size of the W matrix is determined by the size distribution of its element values. Then, we can get the following inference: the smaller the difference between the element values of the W matrix, the closer the element values, the smaller the W matrix, and the smaller the obtained W_1. Moreover, the more similar the element values of the W matrix means that the gray-scale distribution of the image is more uniform, and the texture of the image is more detailed, that is to say, the smaller the image, the more detailed the texture of the image, and vice versa.

(2) Contrast

$$W_2 = \sum_{i=1}^{g}\sum_{j=1}^{g} \left[(i-j)^2 \times P^2(i,j,d,\theta)\right] \tag{10}$$

The contrast is the moment of inertia about the main diagonal in the W matrix. When looking at this problem from a mathematical point of view, we know that when the elements of the W matrix are far away from the diagonal, the contrast between the texture primitives of the image will become strong, and $(i-j)^2$ will become larger. It can be seen from the above formula that the value of W_2 will also increase as $(i-j)^2$ increases. Therefore, the size of W_2 reflects the distribution of element values in the W matrix and the local texture changes of the corresponding image. The larger the W_2, the stronger the contrast of texture primitives and the deeper the texture grooves, which indirectly affects the clarity of the image. Therefore, the following conclusions can be drawn: the clearer the image, the deeper the image texture grooves, the greater the g obtained, and vice versa.

(3) Correlation

$$W_3 = \sum_{i=1}^{g}\sum_{j=1}^{g}[i \times j \times P(i,j,d,\theta) - u_1 \times u_2] \Big/ (d_1 \times d_2) \tag{11}$$

$$u_1 = \sum_{i=1}^{g} i \sum_{j=1}^{g} P(i,j,d,\theta) \tag{12}$$

$$u_2 = \sum_{i=1}^{g} j \sum_{j=1}^{g} P(i,j,d,\theta) \tag{13}$$

$$d_1^2 = \sum_{i=1}^{g} (i-u_1)^2 \sum_{j=1}^{g} P(i,j,d,\theta) \tag{14}$$

$$d_2^2 = \sum_{i=1}^{g} (j-u_2)^2 \sum_{j=1}^{g} P(i,j,d,\theta) \tag{15}$$

It reflects the similarity of the image gray primitives in the x direction and the y direction, what is more, the size distribution of the element values in the W matrix. The more uniform the distribution, the greater the W_3, and vice versa. The size of W_3 is mainly used to determine the main direction of the texture, and its main direction is the direction indicated by θ when W_3 is maximized.

(4) Entropy

$$W_4 = \sum_{i=1}^{g}\sum_{j=1}^{g} P(i,j,d,\theta) \times \log_{10} P(i,j,d,\theta) \tag{16}$$

When the entropy is 0, it means that the image does not contain any texture. Conversely, when the image is full of textures, the entropy will get the maximum value. When thinking about this problem from a mathematical point of view, the above principle is reflected in the formula that when the elements in the W matrix are approximately equal, the entropy obtains the maximum value. In practical applications, $P(i,j,d,\theta) = 0$ will appear, that is, no texture.

(5) Variance

$$W_5 = -\sum_{i=1}^{g}\sum_{j=1}^{g} (i-m)^2 P(i,j,d,\theta) \tag{17}$$

In this formula, m is the mean value of $P(i,j,d,\theta)$.

(6) Sum of average

$$W_6 = \sum_{k=2}^{2g} k \times P_X(k) \tag{18}$$

The sum of average is more suitable for the occasion of feature extraction of grayscale images, and mainly reflects the brightness of the image.

Among them,

$$P_X(k) = \sum_{i=1}^{g}\sum_{j=1}^{g} P(i,j,d,\theta) \quad k = 2,3,\ldots,2g \tag{19}$$

(7) Sum of variance

$$W_7 = \sum_{k=2}^{2g} (k - W_6)^2 P_X(k) \tag{20}$$

The sum of variance has the same meaning as the representation of variance. It represents the period size of the texture and is directly proportional to the period size of the texture.

(8) Inverse difference moment

$$W_8 = \sum_{i=1}^{g}\sum_{j=1}^{g} P(i,j,d,\theta) \Big/ \left[1 + (i-j)^2\right] \tag{21}$$

The inverse moment is a measure of the local change of the image texture, and its size has a certain relationship with the regularity of the image texture. The more regular the texture, the greater the W_8.

(9) Variance of difference

$$W_9 = \sum_{k=0}^{g-1} \left[k - \sum_{k=0}^{g-1} k \times P_Y(k)\right]^2 \times P_Y(k) \tag{22}$$

Among them,

$$P_Y(k) = \sum_{i=1}^{g}\sum_{j=1}^{g} P(i,j,d,\theta) \quad k = 0,1,\ldots,g-1 \tag{23}$$

This characteristic parameter represents the variance of the gray value difference between neighboring pixels. Its size is proportional to the intensity of contrast between texture primitives. That is, the stronger the contrast, the greater the W_9.

(10) Sum of entropy

$$W_{10} = -\sum_{k=2}^{2g} P_X(k) \times \log[P_X(k)] \tag{24}$$

Among them,

$$P_X(k) = \sum_{i=1}^{g}\sum_{j=1}^{g} P(i,j,d,\theta) \quad k = 2,3,\ldots,2g$$

(11) Difference of Entropy

$$W_{11} = -\sum_{k=2}^{2g} P_Y(k) \times \log[P_Y(k)] \tag{25}$$

Among them,

$$P_Y(k) = \sum_{i=1}^{g}\sum_{j=1}^{g} P(i,j,d,\theta)$$

$$k = 0, 1, \ldots, g-1$$

(12) Shadow of clustering

$$W_{12} = -\sum_{i=1}^{g}\sum_{j=1}^{g} [(i-u_1)+(j-u_2)]^3 \times P(i,j,d,\theta) \tag{26}$$

Among them,

$$u_1 = \sum_{i=1}^{g} i \sum_{j=1}^{g} P(i,j,d,\theta)$$

$$u_2 = \sum_{i=1}^{g} j \sum_{j=1}^{g} P(i,j,d,\theta)$$

(13) Prominence of clustering

$$W_{13} = -\sum_{i=1}^{g}\sum_{j=1}^{g} [(i-u_1)+(j-u_2)]^4 \times P(i,j,d,\theta) \tag{27}$$

Among them,

$$u_1 = \sum_{i=1}^{g} i \sum_{j=1}^{g} P(i,j,d,\theta)$$

$$u_2 = \sum_{i=1}^{g} j \sum_{j=1}^{g} P(i,j,d,\theta)$$

(14) Maximal probability

$$W_{14} = \underset{i,j}{MAX}[P(i,j,d,\theta)] \tag{28}$$

Because θ, d, and g are the three main influencing factors of the W matrix, different combinations of the three variables θ, d, and g will generate different GLCM. Through the calculation formulas of the 14 characteristic parameters mentioned above, it can be known that different combinations of θ, d and g will also result in different characteristic parameters.

3.3 Parameter characteristics

In practical applications, comprehensive factors need to be considered. Therefore, when using GLCM, it is generally not to solve all the 14 characteristic parameters, but to seek a balance between the calculation time and the meticulousness of the texture description according to the principle of selection.

Therefore, scholars continue to study this content. Because the direction θ, distance d and gray level g are the three main factors that affect the GLCM, some scholars have proposed to reduce the amount of calculation through image gray-level degradation. However, the reduction in the gray level will result in the reduction in the spatial dependence information of the image texture primitive gray level, which will cause the ability of the feature parameters to characterize the texture to be greatly reduced, so this improved method is not advisable. When gray-scale degradation is undesirable, some scholars choose to do in-depth research on direction θ and distance d. In practical applications, the combination of θ and d of d=1 is generally selected when θ is equal to $0°, 45°, 90°$ and $135°$ respectively. According to the symmetry characteristics of the GLCM, it can be seen that this method is obviously desirable.

Through the above analysis, this paper only selects the four characteristic parameters of energy, contrast, correlation, and entropy. When calculating, in terms of direction θ and distance d, this paper chooses d=1 and $\theta = 0°, 45°, 90°, 135°$, which are commonly used by most researchers. The texture feature extraction steps of GLCM moments are as follows:

Step 1 The algorithm solves the angular second moment, which is also called energy:

$$W_1 = \sum_{i=1}^{g}\sum_{j=1}^{g} P^2(i,j,d,\theta)$$

Step 2 The algorithm solves the contrast.

$$W_2 = \sum_{i=1}^{g}\sum_{j=1}^{g} \left[(i-j)^2 \times P^2(i,j,d,\theta)\right]$$

Step 3 The algorithm solves the correlation.

$$W_3 = \sum_{i=1}^{g}\sum_{j=1}^{g} [i \times j \times P(i,j,d,\theta) - u_1 \times u_2] \Big/ (d_1 \times d_2)$$

Among them,

$$u_1 = \sum_{i=1}^{g} i \sum_{j=1}^{g} P(i,j,d,\theta)$$

$$u_2 = \sum_{i=1}^{g} j \sum_{j=1}^{g} P(i,j,d,\theta)$$

$$d_1^2 = \sum_{i=1}^{g} (i-u_1)^2 \sum_{j=1}^{g} P(i,j,d,\theta)$$

$$d_2^2 = \sum_{i=1}^{g} (j - u_2)^2 \sum_{j=1}^{g} P(i,j,d,\theta)$$

Step 4 The algorithm solves the entropy.

$$W_4 = \sum_{i=1}^{g} \sum_{j=1}^{g} P(i,j,d,\theta) \times \log_{10} P(i,j,d,\theta)$$

Step 5 The algorithm calculates the average of the above four texture parameters to obtain a four-dimensional feature moment of the corresponding image.

4 A MGFE method modeling for feature extraction

A MGFE method based on GLCM and RF is proposed to extract feature. The original vibration data is transformed from multiple granularities to a single granularity, and the machine learning algorithm RF is used to select effective features that affect the actual problem from the high-latitude data. Starting from a single feature level of high-latitude and multi-granularity feature data, it is transformed from a multi-granularity dimension to a single-granularity dimension, which can effectively solve the computational complexity.

The steps of multi-granularity feature extraction based on GLCM and RF expressed as:

(1) Normalize the data to be selected for feature selection

Each data point X_i and feature x_j can be expressed as:

$$X_i = \{x_1, x_2, \ldots x_m\}\ i = 1, 2, \ldots, n) \tag{29}$$

$$x_j = \{x_{j1}, x_{j2}, \ldots x_{jl}\}\ (j = 1, 2, \ldots, m) \tag{30}$$

Then, each data point can be expressed as the following matrix A:

$$A = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1l} \\ x_{21} & x_{22} & \cdots & x_{2l} \\ \vdots & \vdots & & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{ml} \end{bmatrix} \tag{31}$$

Among them, each row of matrix A represents a feature of data point X. In this paper, each feature is collectively referred to as a multi-granularity feature.

(2) Develop a GLCM modeling to reduct dimensionality of multi-granularity features.

Construct a dataset C with normalized data, $C = \{(X_1, y_1), (X_2, y_2), \ldots, (X_n, y_n)\}$, y_i stands for Category label. The GLCM model we chose includes three convolutional layers, a flatten layer, and a fully connected layer. The data after training through GLCM is recorded as c_i.

(3) Use random forest (RF) to perform feature selection on the data to be feature selection.

Construct a dataset C', $C' = \{(X_1, c_1), (X_2, c_2), \ldots, (X_n, c_n)\}$, c_i is the data after training through GLCM. Use RF to train the training set C' and get output m features $F_i(i = 1, 2, \ldots, m)$ with its importance $I_i(i = 1, 2, \ldots, m)$. Moreover, we can choose features according to importance I_i.

For RUL predictions under different working conditions, it is significant to consider the timing relationship between the data before and after the degradation sequence. The extracted public features should also have better trends and monotonicity. By calculating the sequence similarity between the depth features of each dimension in the target domain and HI, the features whose trend is close to that of HI are extracted as common features, so as to realize the migration of degradation information. The dynamic time warping (DTW) distance is used to measure sequence similarity. This measure method with dynamic programming ideas, which can effectively solve the problem of graphic translation, is suitable for measuring the shape similarity of unequal length sequences. Therefore, it is suitable for measuring the depth characteristics and HI of each dimension.

Assuming that number of turning tool in the target domain is N, $F_i(i = 1, 2, \ldots, m)$ represent m dimension features in target domain of turning tool degradation from beginning to end. After training, we can get common feature set $G_i(i = 1, 2, \ldots, m)$. The specific steps is as follows:

Step 1 Calculate the DTW distance between each dimension feature b of feature matrix F_i and the health index HI, and obtain the similarity matrix $R_i = [r_1, r_2, \ldots, r_m]^T$ between the ith turning tool and HI.

Step 2 The similarity matrix is calculated for all turning tool in the target domain, and the similarity matrix $R_i' = [R_1, R_2, \ldots, R_N]$ is obtained.

Step 3 A weight matrix $W = [w_1, w_2, \ldots, w_M]$ is constructed, where w_j represents the weight coefficient of each one-dimensional feature. The calculation method is as follows:

$$w_j = \sum_{i=1}^{N} M - idx_i \tag{32}$$

where idx_i represents the similarity matrix after similarity vector R_i is arranged in ascending order. The higher the similarity and the corresponding weight, the smaller the R_i value.

Step 4 The weight matrix is sorted in descending order, and the importance of the features with large weight value is relatively high. Therefore, the importance order

of each feature is selected, and features are added to it successively. The number of features that minimize the prediction error is selected, and finally the common feature set $G_i(i = 1, 2, \ldots, m)$ is obtained.

Finally, the public features are taken as input and the RUL value corresponding to the same point features at the same time is taken as output to construct ELM regression prediction model. We combine MGFE and ELM to build regression prediction model based on acquired common sensitive feature set to avoid the end-to-end model in small sample problems that may exist on the phenomenon of learning. The calculation formula of RUL is as follows:

$$\mathrm{RUL} = T_1 - T_2 \tag{33}$$

where T_1 is the sampling time before the damage and T_2 is the time when the damage is considered complete. The flowchart of our model is shown in Fig. 1.

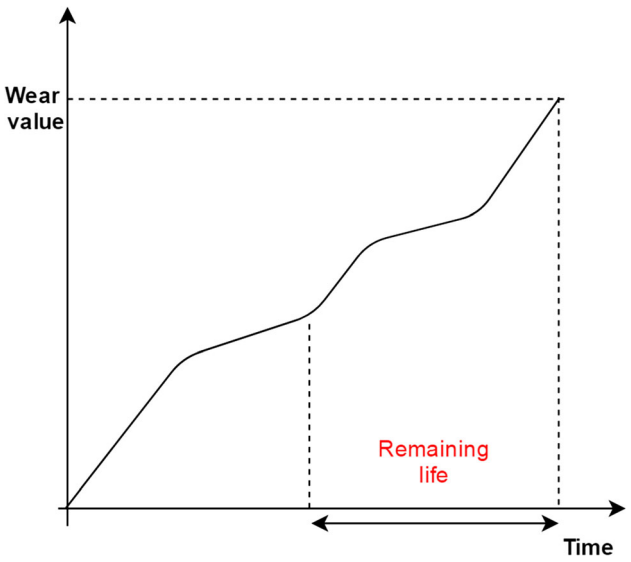

Fig. 2 Schematic diagram of remaining service life of turning tool

5 Model building of RUL prediction

The real-time prediction of the remaining service life of turning tools is of great significance. Therefore, by developing a relatively high economic and adaptable remaining service life prediction model, real-time monitoring of the remaining service life of the turning tool is of great significance to manufacturing companies. The prediction model of the remaining service life of the turning tool is shown in Fig. 2. What is more, when the signal characteristics in the current time period are known, the remaining service life of the turning tool is predicted according to the algorithm in the model.

We take two kinds of comparisons to verify the effect of the RUL prediction:

(1) Compare with two representative HI build methods. Method 1 [29] uses CNN to extract depth features and reduce dimension to construct HI (HI-CNN). Method 2 extracted RMS, FFT, wavelet packet and other traditional features from the original monitoring signal through wavelet packet decomposition, and used reduction algorithm Laplacian eigenmaps (LE) for dimensionality reduction to obtain the final HI (HI-LE).

(2) Compare the RUL prediction results between the proposed method and different regression algorithms with and without MGFE features. To make a comprehensive comparison, we also introduce two RUL prediction methods [30–32] for comparison. In this section, the prediction results of each intermediate step of the proposed method are provided.

In general, the termination condition of the remaining life is set as, when the amplitude of the vibration signal reaches 20 g. However, the data sampling of some turning tool do not reach 20 g, and the sampling has been stopped. In determining the turning tool that do not meet the breakdown threshold of residual life of time, as shown in Fig. 3, if directly to the end of the sample point as the

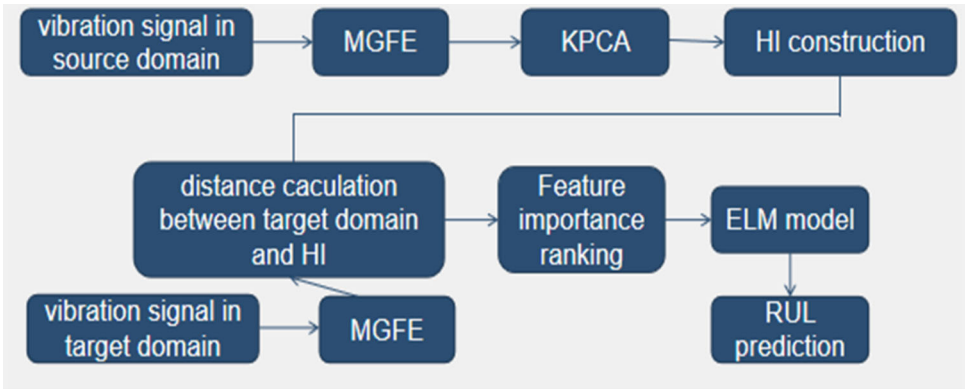

Fig. 1 Flowchart of method

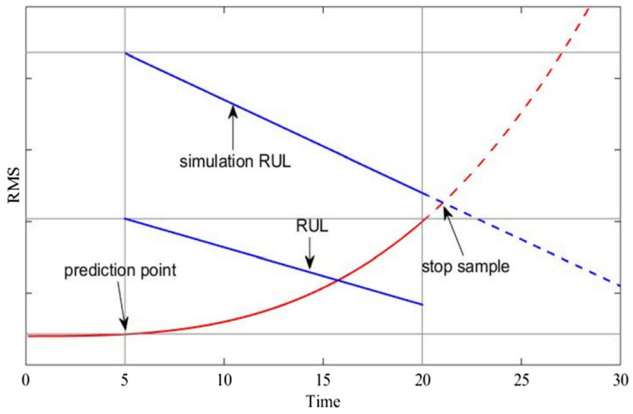

Fig. 3 Determined residual life with simulated fault pattern and finite sample determined

residual life of existing, the scope of its residual life will be very small. In many turning tool model, the scope of its residual life obviously inconsistent with the other turning tool will cause greater error. If the existing sample is used to predict its next trend, and when its trend reaches the defined 20 g value, part of the remaining life value can be obtained for training. And the remaining life curve of this part is consistent with the curve reaching 20 g in range, so it can be used for good remaining life prediction.

Therefore, it is necessary to carry out polynomial fitting for the samples that do not reach the trend. In general, the order of polynomial fitting in this case is 2 to 3. Because of the trend of the remaining life itself being a monotonically increasing curve, the curve fitted after the third order will become more complex and unable to reflect the following trend.

The experiments were carried out in a certain transmission gear processing factory. Milling contains turning tool data under three different working conditions. The data came from the turning tool of the transmission gear machining experiment for three times. The material and geometric characteristics of the machining objects were the same, but the performance of the tool was different. The machining parameters of working conditions 1 were the spindle speed of 200 r/min and the feed rate of 20 mm/min. In working conditions 2, the machining parameters were the spindle speed of 200 r/min and the feed rate of 15 mm/min. In working conditions 3, the machining parameters were the spindle speed of 200 r/min and the feed rate of 10 mm/min. Among them, working conditions 1 and 2 each contain four turning tools, and working conditions 3 contain three turning tools.

In the experiment of this section, working condition 1 is used as the source domain and working condition 2 as the target domain. Among them, four turning tools under working condition 2 take turns as the target turning tool to be predicted, and the other turning tools are used as the available data under working condition 2 to implement algorithm 1 and extract the common features of the two working conditions.

HI build results because the RUL prediction is mainly for early failure occurs after the degradation process of prediction. First, MGFE is used to extract the 100-dimensional features of the source domain data. For convenience to display, three-dimensional features are randomly selected, as shown in Fig. 4. It can be seen that MGFE features show an obvious trend with the degradation process of turning tools. On this basis, KPCA is used to reduce the feature to one dimension to construct the source domain HI, as shown in Fig. 5. At the same time, the moving average method is used to smooth the sample. Since the transfer learning model in this paper focuses on the overall trend of the sample, 20 features are selected for smoothing. If the value is too small, it is easy to be interfered by the local fluctuation of the data, which is not conducive to the construction of the overall trend. Therefore, it can be seen that the HI obtained has the overall trend and monotonicity, and shows obvious changes to the early failures, which proves that it has better fault sensitivity.

Compare our HI build methods with two representative HI build methods, results shown in Figs. 7, 8 and 9. In order to intuitive compare, curve of RMS after corresponding calculation was drawn on figures. Although CNN can also extract depth features, the degradation trend is not obvious. However, the traditional features of HI-LE are not sensitive to early failures. The trend of the method proposed in this paper is the closest to that of RMS compared with HI-CNN and HI-LE.

In order to further compare the effect, we adopted the HI evaluation indexes *Corr* and *Mic* [33], which are widely used at present, to make a numerical comparison of the three kinds of HI as shown in Figs. 6, 7 and 8, and the

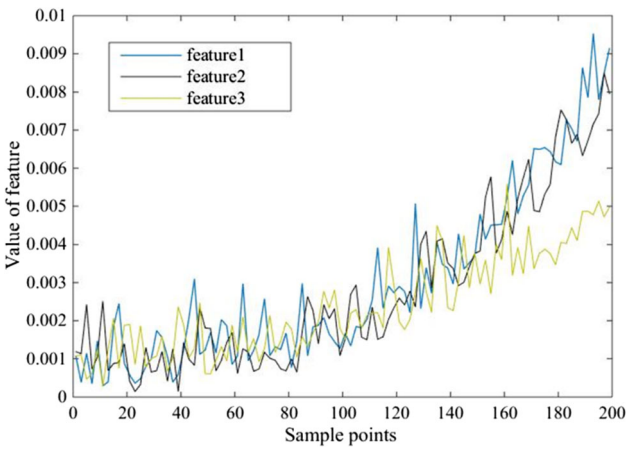

Fig. 4 Schematic diagram of MGFE characteristic trend of turning tool in source domain

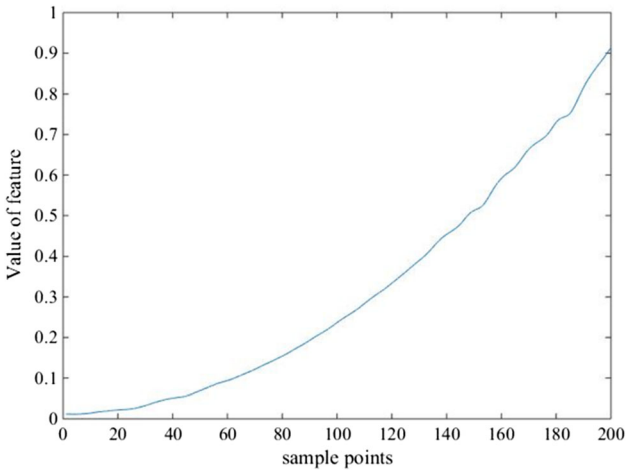

Fig. 5 Smooth source domain turning tool is HI

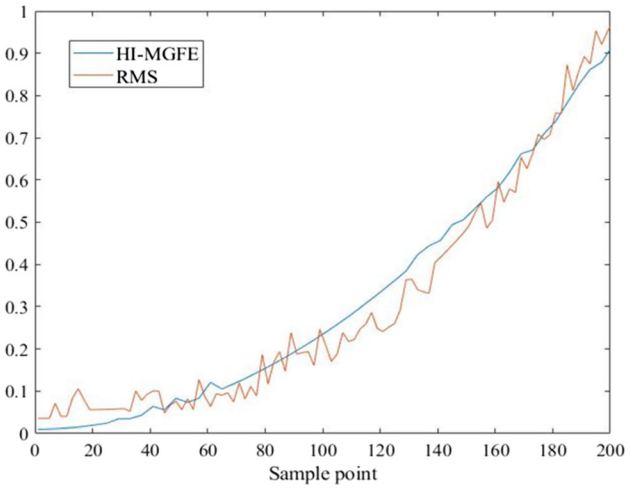

Fig. 6 HI-MGFE and RMS

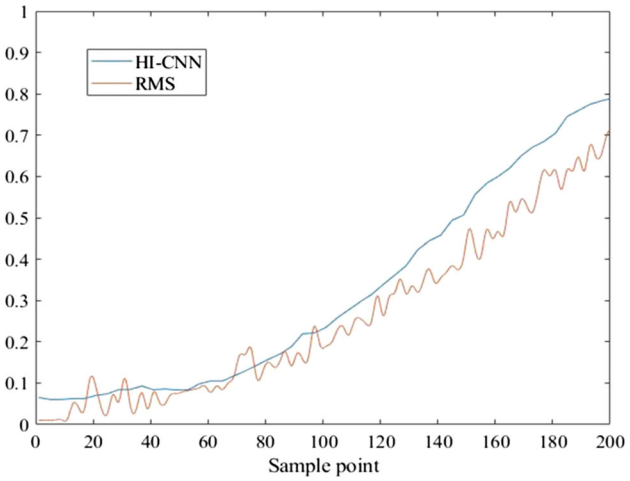

Fig. 7 HI-CNN and RMS

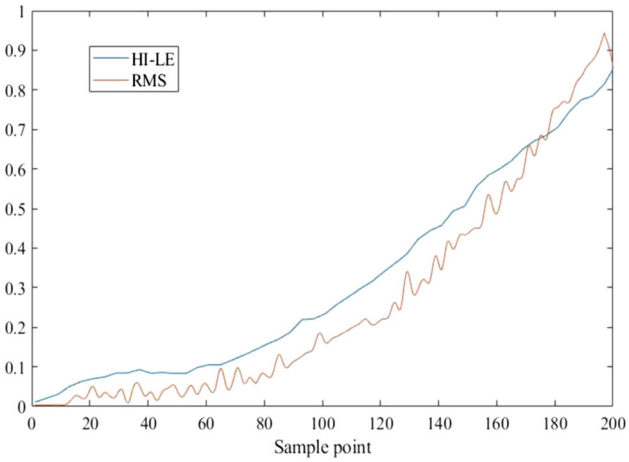

Fig. 8 HI-LE and RMS

Table 1 Evaluation index scores of three HI construction methods

	HI-MGFE	HI-CNN	HI-LE
Corr	0.95	0.95	0.94
Mic	0.94	0.85	0.91

results are shown in Table 1. *Corr* represents the correlation of HI with time, expressed as:

$$Corr = \frac{|\sum_{t=1}^{T}(F_t - \widetilde{F})(l_t - \widetilde{l})|}{\sqrt{\sum_{t=1}^{T}(F_t - \widetilde{F})^2 (l_t - \widetilde{l})}} \tag{34}$$

where F_t and l_t, respectively, represent the eigenvalue and time value of the sample at time t. *Mic* can further measure the correlation between HI and time, which is widely used to measure the nonlinear correlation of data, expressed as:

$$Mic = \max\left(\frac{\max\left(\sum_{F_t \in F, T_t \in T} P(F_t, T_t) \log \frac{P(F_t, T_t)}{\sum_{F_t \in F} P(F_t, T_t) \sum_{F_t \in T} P(F_t, T_t)}\right)}{\log \min F_t, T_t}\right) \tag{35}$$

where F_t and T_t, respectively, represent the eigenvalue and time value of the sample at time t, $B(n) = N^{0.6}$. n represents the number of datasets.

Table 1 shows that the *Mic* value of the method proposed in this paper in Table 1 is significantly higher than that of the other two methods, which verifies the rationality of using the temporal series depth feature to construct HI.

In this paper, turning tool 2 and in the target domain is selected as turning tool to be predicted for effect demonstration, as shown in Fig. 9. We compare transfer learning and non-transfer learning. Non-transfer learning represents direct feature extraction of MGFE to data from working

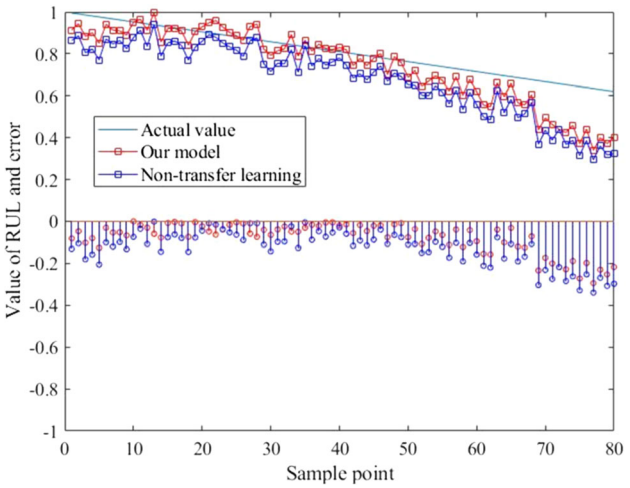

Fig. 9 Prediction results of turning tool 2 under working condition 2

Table 2 Average error of RUL prediction in the target domain

	Liu's method [30]	Wu's method [31]	Deutsch's method [32]	Our method
RMSE	45.31	50.54	46.80	40.89

conditions 1 and 2, and then predicts on the target turning tool to be tested. It can be seen that the common features have a stronger ability to express the degradation trend with transfer learning, and it has better prediction accuracy.

Finally, this paper selects three RUL prediction methods for experimental comparison, and the comparison results are shown in Table 2. We use the average RUL prediction error for the turning tool in the target domain. Among them, Liu et al. [30] used fault feature selection and SVM to build a prediction model. Wu et al. [31] extracts dynamic differential features from a raw signal and then utilizes LSTM to conduct RUL prediction from the perspective of temporal information. Deutsch et al. [32] used the time series of fault features as the input of the deep confidence network to establish a RUL prediction model. We use squared error (RMSE) to evaluate prediction performance, RMSE expressed as:

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{n} (\hat{y}_i - y_i)}{n}} \tag{36}$$

It can be seen that the method in this paper achieves the minimum prediction error on RMSE.

6 Conclusion

This paper takes turning tool wear online monitoring and remaining life prediction as the goal to establish a turning tool wear monitoring and remaining life prediction model based on extreme learning machine and transfer learning. In order to predict the residual life of turning tool under different working conditions, we introduce the monitoring data of the same turning tool under different working conditions, this method can effectively solve the problem that the prediction result of RUL is limited when the existing turning tool data is insufficient. The HI constructed by using the MGFE can effectively describe the degradation trend and has good monotony and fault sensitivity, which can effectively improve the RUL prediction effect of the target turning tool.

Acknowledgements Information of funding: Subsidized by Chongqing Basic Science and Research Project (cstc2015jcyjbx0133), National Natural Science Foundation of China (NSFC 51375519), National Natural Science Foundation of China (NSFC 51975078).

References

1. Karandikar JM, Abbas AE, Schmitz TL (2014) Tool life prediction using Bayesian updating. Part 2: turning tool life using a Markov Chain Monte Carlo approach. Precis Eng 38(1):18–27
2. Yin Z, Huang C, Yuan J et al (2015) Cutting performance and life prediction of an Al_2O_3/TiC micro–nano-composite ceramic tool when machining austenitic stainless steel. Ceram Int 41(5):7059–7065
3. Karandikar JM, Abbas AE, Schmitz TL (2014) Tool life prediction using Bayesian updating. Part 1: milling tool life model using a discrete grid method. Precis Eng 38(1):9–17
4. Benkedjouh T, Medjaher K, Zerhouni N et al (2015) Health assessment and life prediction of cutting tools based on support vector regression. J Intell Manuf 26(2):213–223
5. Krolczyk GM, Nieslony P, Legutko S (2015) Determination of tool life and research wear during duplex stainless steel turning. Arch Civ Mech Eng 15(2):347–354
6. Qin Yi, Xiang S, Chai Yi, Chen H (2020) Macroscopic-microscopic attention in LSTM networks based on fusion features for gear remaining life prediction. IEEE Trans Ind Electron 67(12):10865–10875
7. Dong M, He D (2007) A segmental hidden semi-Markov model (HSMM)-based diagnostics and prognostics framework and methodology. Mech Syst Signal Process 21(5):2248–2266
8. Shen ZJ, Chen XF, He ZJ et al (2013) Remaining life predictions of rolling bearing based on relative features and multivariable support vector machine. J Mech Eng (in Chinese) 49(2):183–189
9. Mao W, He J, Zuo MJ (2019) Predicting remaining useful life of rolling bearings based on deep feature representation and transfer learning. IEEE Trans Instrum Meas PP(99):1–1
10. Mosallam A, Medjaher K, Zerhouni N (2016) Data-driven prognostic method based on Bayesian approaches for direct remaining useful life prediction. J Intell Manuf 27(5):1037–1048
11. Shihab SK, Khan ZA, Mohammad A et al (2014) A review of turning of hard steels used in bearing and automotive applications. Prod Manuf Res 2(1):24–49

12. Kim DM, Bajpai V, Kim BH et al (2015) Finite element modeling of hard turning process via a micro-textured tool. Int J Adv Manuf Technol 78(9–12):1393–1405
13. Ahmadzadeh F, Lundberg J (2014) Remaining useful life estimation. Int J Syst Assur Eng Manag 5(4):461–474
14. Gupta MK, Sood PK (2017) Surface roughness measurements in NFMQL assisted turning of titanium alloys: an optimization approach. Friction 5(2):155–170
15. Das SR, Dhupal D, Kumar A (2015) Study of surface roughness and flank wear in hard turning of AISI 4140 steel with coated ceramic inserts. J Mech Sci Technol 29(10):4329–4340
16. Kumar R, Sahoo AK, Mishra PC et al (2018) Comparative investigation towards machinability improvement in hard turning using coated and uncoated carbide inserts: part I experimental investigation. Adv Manuf 6(1):52–70
17. Mia M, Dhar NR (2017) Optimization of surface roughness and cutting temperature in high-pressure coolant-assisted hard turning using Taguchi method. Int J Adv Manuf Technol 88(1–4):739–753
18. Leo Kumar SP, Jerald J, Kumanan S et al (2014) A review on current research aspects in tool-based micromachining processes. Mater Manuf Process 29(11–12):1291–1337
19. Fernández-Valdivielso A, López de Lacalle LN, Urbikain G et al (2016) Detecting the key geometrical features and grades of carbide inserts for the turning of nickel-based alloys concerning surface integrity. Proc Inst Mech Eng Part C J Mech Eng Sci 230(20):3725–3742
20. Bensouilah H, Aouici H, Meddour I et al (2016) Performance of coated and uncoated mixed ceramic tools in hard turning process. Measurement 82:1–18
21. Gupta M, Kumar S (2015) Investigation of surface roughness and MRR for turning of UD-GFRP using PCA and Taguchi method. Eng Sci Technol Int J 18(1):70–81
22. Madariaga A, Esnaola JA, Fernandez E et al (2014) Analysis of residual stress and work-hardened profiles on Inconel 718 when face turning with large-nose radius tools. Int J Adv Manuf Technol 71(9–12):1587–1598
23. Frangopol DM, Soliman M (2016) Life-cycle of structural systems: recent achievements and future directions. Struct Infrastruct Eng 12(1):1–20
24. Javed K, Gouriveau R, Zerhouni N et al (2014) Enabling health monitoring approach based on vibration data for accurate prognostics. IEEE Trans Ind Electron 62(1):647–656
25. Sun S, Brandt M, Mo JPT (2014) Evolution of tool wear and its effect on cutting forces during dry machining of Ti–6Al–4V alloy. Proc Inst Mech Eng Part B J Eng Manuf 228(2):191–202
26. Khorasani AM, Yazdi MRS (2017) Development of a dynamic surface roughness monitoring system based on artificial neural networks (ANN) in milling operation. Int J Adv Manuf Technol 93(1–4):141–151
27. Pervaiz S, Rashid A, Deiab I et al (2014) Influence of tool materials on machinability of titanium-and nickel-based alloys: a review. Mater Manuf Process 29(3):219–252
28. Suresh P, Marimuthu K, Ranganathan S et al (2014) Optimization of machining parameters in turning of Al–SiC–Gr hybrid metal matrix composites using grey-fuzzy algorithm. Trans Nonferrous Met Soc China 24(9):2805–2814
29. Guo L, Lei YG, Li NP et al (2018) Machinery health indicator construction based on convolutional neural networks considering trend burr. Neurocomputing 292(31):142–150
30. Liu Z, Zuo MJ, Qin Y (2016) Remaining useful life prediction of rolling element bearings based on health state assessment. Proc Inst Mech Eng C J Mech Eng Sci 230(2):314–330
31. Wu Y, Yuan M, Dong S, Lin L et al (2018) Remaining useful life estimation of engineered systems using vanilla lstm neural networks. Neurocomputing 275:167–179
32. Deutsch J, He D (2017) Using deep learning-based approach to predict remaining useful life of rotating components. IEEE Trans Syst Man Cybern Syst 48(1):11–20
33. Reshef DN, Reshef YA, Finucane HK et al (2011) Detecting-novel associations in large data sets. Science (New York, NY) 334(6062):1518–1524