

Article

A Multisensor Fusion Method for Tool Condition Monitoring in Milling

Yuqing Zhou ^{1,2}  and Wei Xue ^{2,*}

¹ College of Mechanical Engineering, Zhejiang University of Technology, Hangzhou 310014, China; zhoyuq@wzu.edu.cn

² College of Mechanical and Electrical Engineering, Wenzhou University, Wenzhou 325035, China

* Correspondence: xw_wzu@163.com; Tel.: +86-577-8868-9169

Received: 29 September 2018; Accepted: 8 November 2018; Published: 10 November 2018



Abstract: Tool fault diagnosis in numerical control (NC) machines plays a significant role in ensuring manufacturing quality. Tool condition monitoring (TCM) based on multisensors can provide more information related to tool condition, but it can also increase the risk that effective information is overwhelmed by redundant information. Thus, the method of obtaining the most effective feature information from multisensor signals is currently a hot topic. However, most of the current feature selection methods take into account the correlation between the feature parameters and the tool state and do not analyze the influence of feature parameters on prediction accuracy. In this paper, a multisensor global feature extraction method for TCM in the milling process is researched. Several statistical parameters in the time, frequency, and time–frequency (Wavelet packet transform) domains of multiple sensors are selected as an alternative parameter set. The monitoring model is executed by a Kernel-based extreme learning Machine (KELM), and a modified genetic algorithm (GA) is applied in order to search the optimal parameter combinations in a two-objective optimization model to achieve the highest prediction precision. The experimental results show that the proposed method outperforms the Pearson’s correlation coefficient (PCC) based, minimal redundancy and maximal relevance (mRMR) based, and Principal component analysis (PCA)-based feature selection methods.

Keywords: tool condition monitoring; milling process; multisensor fusion; kernel extreme learning machine; genetic algorithm

1. Introduction

Milling is a very common and efficient cutting operation that uses a rotary milling cutter with one or more teeth to intermittently cut workpieces into flat surfaces, grooves, threads, and many complex geometric components. Highly efficient milling processes are suitable for mass production and have been used widely in the manufacturing industry. Tools are considered the pillars of the milling process [1], and tool breakage is a major cause of unscheduled stops in a machining environment. Tool breakage has negative direct (capital) and indirect (time loss) effects. The downtime of a milling machine due to tool failure accounts for 7–20% of total downtime [2,3], and the cost of tools and tool changes account for 3–12% of the total processing cost [4]. As the conditions of tools vary over time, the timeliness of detecting tool damage is critical, and it requires an appropriate tool replacement strategy. Conventionally, tool change strategies are subjective and regular and the time period is determined by the experience of the operator. While early replacement of a workable tool will waste tools and increase downtime, late replacement of a worn tool will result in lower quality workpieces and increased production costs [5]. Research [6,7] has determined that tools are typically only used for 50–80% of their effective lifespan. Thus, tool condition monitoring (TCM) has become an important challenge in the milling processes that schedules activities based on the result of condition

measurements without interrupting normal machine operations [8]. TCM systems are developed in order to generate better surface quality and to extend tool life by diagnosing cutting tool deficiencies with appropriate signal processing and pattern recognition techniques, thereby reducing losses due to tool wear or failure. An accurate and reliable TCM system can reduce costs by 10–40% by reducing downtime and fully utilizing tools [7,9].

TCM in milling processes has been studied for over 30 years using two types of methods: Direct monitoring and indirect monitoring. Direct monitoring methods use optical equipment and machine vision technology to directly monitor the tool; for example, microscopes are used to capture tool images and evaluate the tool's state with image analysis technology [10]. Direct methods are advantageous because they do not affect the machining process and they have high recognition accuracy under certain conditions. However, direct methods are not suitable for the manufacturing shop [11,12] because (1) the required equipment and the software are expensive, which could increase manufacturing costs, and (2) the recognition accuracy is easily disturbed by the cutting fluid and cutting chips. Therefore, indirect monitoring methods have been widely adopted. One or more sensors are used to measure a signal associated with the tool state, and the tool state is estimated by analyzing the measured signal. Compared with direct methods, indirect methods are cheaper and more adaptable to practical applications.

Indirect TCM is a data-driven method; it uses single or multiple sensors to monitor the milling process and synthesize information provided by the sensor signal to determine the best estimates for the tool state through a mechanism based on training [13]. Indirect TCM can be divided into two phases: Model training and online monitoring. The model training phase provides a training sample and then trains the monitoring model. The online monitoring phase monitors the milling process and estimates the cutting tool's condition in real time. The model training phase consists of three modules: Sensor configuration, feature extraction, and monitoring model. The sensor configuration module provides an alternative sensory signal; the feature extraction module extracts features in the sensory signal that are related to the tool condition (wear, breakage, etc.). The monitoring model module builds a decision support model for online monitoring, and the online monitoring phase consists of two modules: Online sensory monitoring and decision-making. The sensor configuration from online monitoring is determined from the model training. If the sensors are not changed, then the configuration is the same in both online monitoring and the model training. If the sensors are changed, then the configuration from online monitoring will not consider sensors that are irrelevant to the tool condition.

Multisensors can provide more information related to the tool condition, but they increase the risk that effective information will be overwhelmed by redundant information. Thus, obtaining the most effective feature information from multisensor signals is currently a hot topic. In this paper, a multisensor global fusion method for TCM in the milling process is researched. Several statistical parameters in the time, frequency, and time–frequency domains of multiple sensors are selected as an alternative parameter set. The monitoring model is executed by the Kernel extreme learning machine (KELM), and the genetic algorithm (GA) is applied in order to search the optimal parameter combinations to achieve the highest prediction precision. The remainder of this paper is organized as follows: Section 2 reviews the literature of multisensor feature extraction in TCM for the milling process. Section 3 describes the theoretical framework and the learning algorithm of our proposed method. Section 4 describes an experiment with an open data set of the tool condition in order to compare the proposed method and several current feature extraction methods. Conclusions and suggestions for future work are given in Section 5.

2. Literature Review

Based on the literature, many sensors have been used in TCMs to obtain process signals [14–17], such as cutting force, vibration, current, and acoustic emission sensors. However, due to the uncertainty and limitations of single sensor monitoring, TCMs based on a single sensor cannot achieve good

performance or robustness. Multiple sensor-based methods can enhance the richness of information that contains potential tool wear levels [18]. Although multisensor setups provide more redundant information, they can reduce the overall uncertainty of the measurement and improve the resolution and accuracy of the TCM system [4,19,20].

Therefore, multisensor monitoring has become a research trend. According to the statistics in Reference [21], the number of studies using multisensor-based methods has been gradually increasing for TCMs in the milling process. Recently, several studies found that the prediction accuracy of TCM is not positively related to the number of sensors and the number of feature parameters. The primary purpose of the feature extraction module is to extract feature parameters closely related to the tool state from the signals and to significantly reduce the dimensions of the original information. The result of the feature extraction module greatly affects the performance of the monitoring model. Too many feature parameters will greatly increase the model computation and affect the timeliness of online monitoring. In addition, irrelevant and redundant feature parameters have a negative impact on the performance of the monitoring model, and a few appropriate feature parameters can generate a more accurate and robust model [22,23].

The feature extraction methods for TCM in the milling process can be divided into four classifications, (1) time-domain-based method; (2) frequency-domain-based method; (3) wavelet analysis (time–frequency domain) based method; and (4) multidomain-based method. The methods based on the time domain extract feature information related to the tool state from the time dimension of the signal, including the time series analysis, several statistical parameters, and other information. The time series analysis includes Auto-Regression(AR), Autoregressive Moving Average(ARMA), time domain averaging (TDA), and other information [3]. The statistical parameters include the root mean square error (RMSE), maximum/minimum, average, standard deviation, and kurtosis [24]. Frequency-domain-based methods extract feature information related to the tool state from the frequency dimension of the signal, including the frequency structure and harmonic component of the signal. These methods convert sensory signals from the time domain into the frequency domain with a fast Fourier transform and then extract feature parameters, such as the power spectrum, peak-to-peak amplitude, and tooth frequency [25]. The time-and-frequency-domain-based methods can only provide feature information from a single perspective, and both assume that the signal is stationary, which is not suitable for non-stationary signals in milling processes [26].

To overcome the shortcomings of the time-and-frequency-domain-based methods, a time–frequency analysis method based on a wavelet transform (WT) has been used for feature extraction in milling TCM. In WT-based methods, a discrete wavelet transform (DWT) [27], a continuous wavelet transform (CWT) [28,29], and a wavelet packet transform (WPT) [30] have been applied in order to extract a series of wavelet coefficients to reflect the tool state.

Multidomain-based methods select some parameters (such as the time domain, frequency domain, and time–frequency domain) to compose a candidate feature parameter set and apply certain feature selections or dimensional reduction methods to some feature parameters that are strongly related to the tool state. The advantage of multidomain methods is that they provide more candidate feature parameters related to the tool state and reduce the risk of losing important information, which is important for improving the performance of TCM. In the TCM model training phase, there is limited knowledge and experience to guide the selection of parameters. If the parameters closely related to the tool states are not selected for the candidate feature set, then the performance of the monitoring model could easily decrease. In the multidomain-based methods, although the number of feature parameters of the candidate feature set is clearly increased, the feature set is reduced to a low dimension by the feature selection or dimension reduction algorithm. Then, one needs to calculate the feature parameters after a dimension reduction in the online monitoring phase, which barely affects the operation speed of the model.

Currently, there are two kinds of multidomain feature extraction methods: Feature fusion and feature selection. Feature fusion methods obtain new parameters from preset sensor feature parameter

sets by linear or nonlinear mapping, and the new feature parameters are used as the input of the monitoring model [31,32]. For example, Wang G.F. et al. [33] used the local preserving projection (LPP) algorithm to establish new fusion parameters and reduce the feature parameter dimension. Wang J. et al. [34] fused 54 feature parameters in the time-domain and the frequency-domain, as well as wavelet coefficients of 11 new feature parameters, using the kernel Principal component analysis (PCA) algorithm. The advantage of the feature fusion methods is they can use all preset feature parameters to provide information that is more comprehensive and that reduces feature dimensionality. However, in the online monitoring stage, all the sensors used in the training phase are needed and all the preset feature parameters are calculated, which could increase the maintenance cost and the computation for online monitoring. The feature selection methods select a few effective feature parameters from the preset candidate feature parameters. For example, Zhang et al. [22] selected 13 parameters significantly related to tool wear values using Pearson's correlation coefficient (PCC) from 144 feature parameters extracted in the time, frequency, and time–frequency domains of multiple sensory signals. In Reference [35], the minimal redundancy and maximal relevance (mRMR) algorithm is utilized to select the most prominent features. Sohyung et al. [36] constructed 135 feature parameters in the time and frequency domains of multiple sensory signals and selected 25 feature parameters by employing the entropy correlation algorithm. Liu et al. [4] extracted 138 feature parameters (including time domain, frequency domain, and wavelet energy) as a candidate feature parameter set and used the fast correlation filter (FCBF) algorithm to establish the smallest redundant feature set, including 19 feature parameters. Feature selection methods can reduce the number of input parameters of the monitoring model by screening out the characteristic parameters that have a strong correlation with the tool state, but they also reduce the computation of the online monitoring stage. However, most of the current feature selection methods take into account the correlation between the feature parameters and the tool state and do not analyze the influence of the feature parameters on prediction accuracy. The parameters related most strongly to the tool state cannot generate the highest prediction accuracy. It is necessary to find out the global optimal combination of the sensor feature parameters for prediction accuracy.

3. Theoretical Framework

3.1. The Framework of TCM

In this paper, a multisensor global fusion method for TCM is constructed, whose structure is shown in Figure 1. The operation of this method is composed of three steps. The first step is multisensor signal acquisition. Dynamic signals from multiple sensors (such as force, vibration, and current) are collected in order to depict the characteristic of the milling process. The second step is global feature extraction. A few statistical parameters in the time, frequency, and time–frequency domains of multiple sensors are extracted as candidate parameters. The KELM is applied as the monitoring model and the base of prediction precision. Two objective functions include the prediction error and the number of parameters, and the optimal sensor feature parameters are sought in order to realize the smallest prediction error and the least number of parameters through a modified GA. The last step is to monitor the tool state and judge the tool wear value using KELM, in which the inputs are the optimal sensor feature parameters.

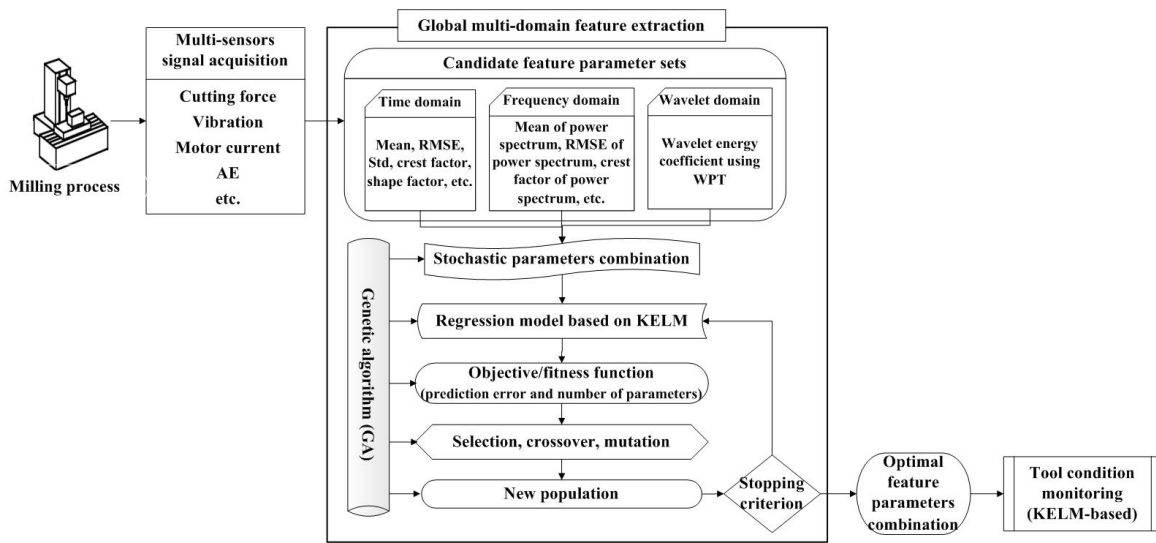


Figure 1. Framework of the proposed tool condition monitoring method. AE: Acoustic emission, RMSE: Root mean square error, WPT: Wavelet packet transform, KELM: Kernel extreme learning machine.

3.2. Kernel Extreme Learning Machine

An ELM first proposed by Huang et al. [37,38] is proposed for the single hidden-layer feed forward neural networks. A Kernel extreme learning machine (KELM) is an application of ELM with a kernel function that has the advantage of a fast learning speed and high efficiency compared to other learning algorithms, such as Support vector machine(SVM), back propagation neural networks(BPNN), and least square SVM(LS-SVM) [39].

The KELM theory tends to reach not only the smallest training error, but also the smallest norm of the output weights. This objective function for the regression task can be expressed as follows:

$$\min_{\gamma, \epsilon} \frac{1}{2} \|\beta\|_F^2 + \frac{C}{2} \sum_{i=1}^n \|\epsilon_i\|^2 \quad s.t. \quad \epsilon_i = y_i - \beta^T f(x_i), \forall i, \tag{1}$$

where $(X, Y) = \{(x_i, y_i), i = 1, \dots, n\}$ is the training sample set, n is the number of training samples, $f(x_i) = \{f(x_{i1}), \dots, f(x_{iL})\}$ is the hidden-layer output vector with respect to x_i , $f(\bullet)$ is a form of feature mapping that maps the input data from the original dimension space to the L -dimensional hidden-layer feature space, $\beta = \{\beta_1, \dots, \beta_L\}^T$ is the vector of the output weights between the hidden layer of the L nodes and the output node, ϵ_i is the training error of the i -th training sample, and C is the regularization parameter that trades off the norm of output weights and training errors. $\|\bullet\|_F$ is the Frobenius norm.

According to Reference [37], the optimal $\hat{\beta}$ that minimizes Equation (1) can be efficiently solved as:

$$\hat{\beta} = \Phi^T \left(\frac{I}{C} + \Phi \Phi^T \right)^{-1} Y, \tag{2}$$

where F is the hidden-layer output matrix, $F = \begin{bmatrix} f(x_1) \\ \vdots \\ f(x_n) \end{bmatrix} = \begin{bmatrix} f(x_{11}) & \Lambda & f(x_{1L}) \\ \vdots & \Lambda & \vdots \\ f(x_{n1}) & \Lambda & f(x_{nL}) \end{bmatrix}$, I is an identity

matrix, and $Y = \{y_1, \Lambda, y_n\}^T$ is the dependent value vector in the training samples. The prediction score y'_x on test point x is determined by:

$$y'_x = \varphi(x) \hat{\beta}. \tag{3}$$

Since the signal of the multisensors in TCM is high-dimension, nonlinear, and heterogeneous, the feature mapping $\phi(\cdot)$ is unknown. It is necessary to define a kernel matrix for ELM using Mercer's conditions:

$$\Psi = \Phi\Phi^T = \{\varphi_{ij}\}, \varphi_{ij} = \phi(x_i)\phi(x_j) = k(x_i, x_j). \quad (4)$$

Then, the prediction score on test point x in Equation (3) can be rewritten as:

$$y'_x = \phi(x)\Phi^T \left(\frac{I}{C} + \Phi\Phi^T \right)^{-1} Y = \begin{bmatrix} k(x, x_1) \\ \mathbf{M} \\ k(x, x_n) \end{bmatrix}^T \left(\frac{I}{C} + \Psi \right)^{-1} Y. \quad (5)$$

In this context, similar to SVM, the feature mapping $\phi(\cdot)$ does not need to be known, and a common kernel function can be used, e.g., a Gaussian kernel, linear kernel, or polynomial kernel. In addition, the dimensionality L of the feature space (number of hidden nodes) does not need to be given. It is noteworthy that the kernel matrix Ψ is only related to the input data x_i and the number of training samples [39].

3.3. Global Feature Extraction

Different from most of the current feature selection methods that consider the correlation (linear/nonlinear) between the feature parameters and the tool state, the prediction accuracy of the TCM method was taken into account in this paper.

In this paper, we want to reach not only the smallest prediction error, but also the lowest number of parameters. These two objective functions can be expressed as follows:

$$\min \frac{1}{m} \sqrt{\sum_{t=1}^m (Y_{X_t} - Y'_{X_t})^2} \quad \& \quad \min \sum_{q=1}^Q \delta_q, \quad (6)$$

where Y_{X_t} is the dependent value with respect to X_t , Y'_{X_t} is the predicted value with respect to X_t using the KELM, δ_q denotes the state of the q -th feature parameter, and $\delta_q = 1$ indicates that the q -th feature parameter is included in the input set of KELM regression model; otherwise, $\delta_q = 0$, and Q is the number of candidate feature parameters. The first objective function calculates the RMSE of the selected data, and the second objective function calculates the number of parameters used in the KELM regression model.

Mathematically, the problem of Equation (6) is formulated as a combinatorial optimization problem. Here, the function used for optimization is the generalization performance of the predictive model, represented by the error of a training data set, and the design variables are the inclusion (1) or the exclusion (0) of the candidate parameters. Therefore, the two-objective optimization problem in Equation (6) can be transformed into a single objective optimization, in which the second objective is embedded in the optimization with the first objective, while remaining the best solution with the fewest parameters in each optimization iteration. An exhaustive selection of candidate feature parameters would evaluate many different combinations ($=2^Q$). This process becomes impracticable when Q is large. The GA algorithm is an intelligent optimization method for function optimization based on the mechanics of natural genetics and biological evolution, which is capable of solving the global optimization of complex problems [40]. In this paper, GA was used to optimize Equation (6), and the global feature extraction algorithm-based GA is shown in Figure 2.

In Figure 2, the fitness function is the RMSE with the training set using KELM. Obviously, a high RMSE indicates a low fitness, and those individuals with greater fitness will have a greater probability of being selected for recombination. The selection operator divides into two parts. The first part is the best chromosome that has a minimal RMSE and the lowest number of parameters, and that can solve the second objective in Equation (6). The other part selects the individuals according to their fitness level using the roulette wheel method. The crossover operator and the mutation operator

recombine the selected individuals in order to generate a new population with a preset crossover rate P_c (the chance of crossover being applied to a chromosome) and mutation rate P_m (the chance of a chromosome being mutated). Finally, the iterations are completed when the number of iterations reaches the preset maximum iteration value.

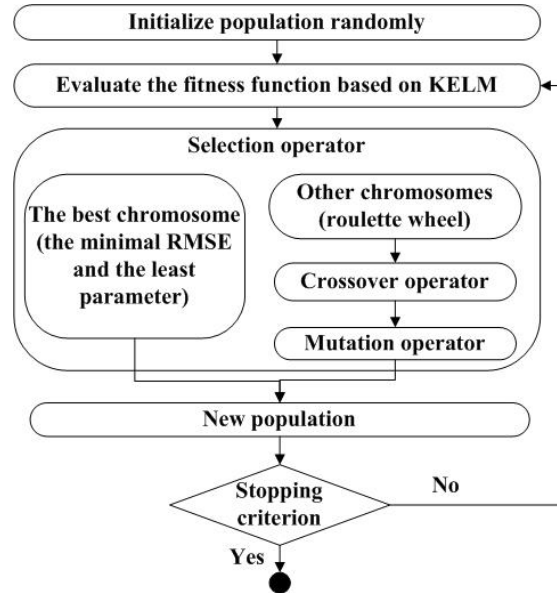


Figure 2. The process of global feature extraction algorithm based genetic algorithm (GA).

4. Experiments

4.1. Descriptions of Datasets

In the present work, the TCM prediction problem reported in the 2010 Prognostics and Health Management (PHM) data challenge [41] was used to verify the performance of the proposed method. The experimental setup is shown in Figure 3 [42], and this dataset was collected from a high-speed Computerized numerical control (CNC) milling machine (Type RFM 760 from Roders Tech Co. Ltd in Germany) under dry milling operations. The operation parameters in this experiment are shown in Table 1.

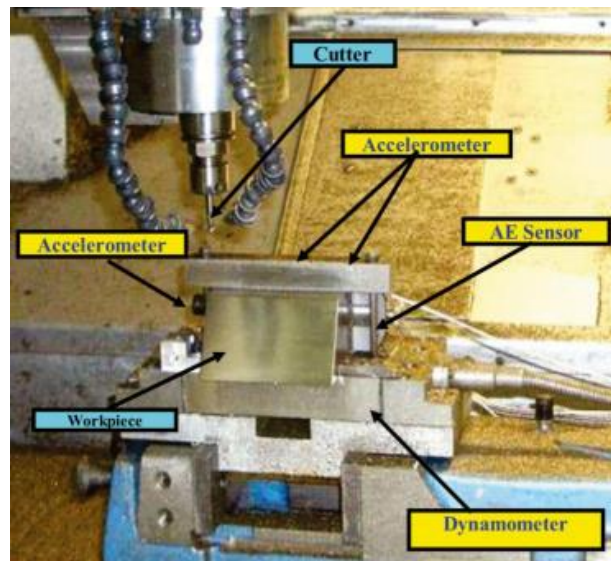


Figure 3. The experimental setup.

Table 1. Operation parameters in the experiment.

Operation Parameter	Value
CNC machine	Roders Tech RFM 760
Workpiece material	Inconel 718 (Jet engines)
Cutter	3-flute ball nose
Spindle speed	10,400 RPM
Feed rate	1555 mm/min
Y depth of cut (radial)	0.125 mm
Z depth of cut (axial)	0.2 mm
Number of sensors	5
Number of sensor channels	7
Sampling data	50 KHz/channel

The workpiece surface was machined to have a slope with 60° to accommodate the 2-flute ball nose cutter. A Kistler quartz 3-component platform dynamometer was mounted between the workpiece and the machining table to measure the cutting forces in the form of charges, which were then converted to voltages by the Kistler charge amplifier. Three Kistler piezoelectric accelerometers were mounted on the workpiece to measure the machine tool vibrations of the cutting process in the X, Y, and Z directions, respectively. A Kistler acoustic emission (AE) sensor was mounted on the workpiece to monitor the high frequency stress wave generated by the cutting process. Therefore, the sensory data consisted of seven channels: The force in three directions, the vibration in three directions, and the AE-RMS. A DAQ NI PCI1200 was adopted to perform in-process measurements, including the force and vibration in three directions (x, y, and z), with a continuous sampling frequency of 50 KHz during the tool wear test. The corresponding flank wear of each individual flute was measured offline using a LEICA MZ12 microscope after finishing each surface. Finally, three individual cutter records named C1, C4, and C6 were selected as our dataset, and each record contained 315 data samples. Two tests, C1 and C4, were used as the training sample, and C6 was used as the testing set.

4.2. Candidate Parameter Sets

To overcome the drawback of features in a single domain, which lose some useful information related to the tool condition, in this study, the multidomain features of multisensor signal were extracted in the time, frequency, and time–frequency domains. According to previous papers [34,43,44] and our experimental studies [45], a few dimensional and dimensionless statistical feature parameters in the time, frequency, and time–frequency (wavelet) domains were chosen.

As shown in Table 2, nine statistical feature parameters related to the tool state from the time dimension of the sensor signal were extracted as candidate parameters, including three dimensional features—the average value, the root mean square, and the standard deviation—and six dimensionless features—the crest factor, the shape factor, the waveform, the kurtosis factor, the skewness factor, and the margin factor.

As shown in Table 3, eight statistical feature parameters related to the tool state from the frequency domain of the sensor signal were extracted as candidate parameters, including two dimensional features—the mean and the root mean square of the power spectrum—and six dimensionless features—the crest factor of the power spectrum, the modified equivalent bandwidth, the high–low ratio of the power spectrum, the stabilization ratio, the skewness of the bandpower, and the kurtosis of the bandpower.

Table 2. Nine time domain statistical parameters.

Domain	Indexes	Formula
Time	Average Value T_{avg}	$T_{avg} = \sum_{j=1}^n x_j/n$
	Root mean square T_{rms}	$T_{rms} = \sqrt{\sum_{j=1}^n x_j^2/n}$
	Standard Deviation T_{sd}	$T_{sd} = \frac{1}{n-1} \sqrt{\sum_{j=1}^n (x_j - T_{avg})^2}$
	Crest Factor T_{cf}	$T_{cf} = \max\{ x_j \}/T_{rms}$
	Shape factor T_{sf}	$T_{sf} = T_{rms}/T_{avg}$
	Waveform T_{wf}	$T_{wf} = n \cdot T_{rms}/(\sum_{j=1}^n x_j)$
	Kurtosis Factor T_{ku}	$T_{ku} = (\sum_{j=1}^n (x_j - T_{avg})^4)/(n \cdot T_{sd}^4) - 3$
	Skewness Factor T_{sk}	$T_{sk} = (\sum_{j=1}^n (x_j - T_{avg})^3)/(n \cdot T_{sd}^3)$
	Margin factor T_{mf}	$T_{sk} = n^2 \cdot \max\{ x_j \}/(\sum_{j=1}^n \sqrt{ x_j })^2$

Table 3. Eight frequency domain statistical parameters ¹.

Domain	Indexes	Formula
Frequency	Mean of power spectrum F_{mps}	$F_{mps} = \sum_{j=1}^n P_j/n$
	Root mean square of power spectrum F_{rms}	$F_{rms} = \sqrt{\sum_{j=1}^n P_j^2/n}$
	Crest factor of power spectrum F_{cf}	$F_{cf} = \max\{P_i\}/F_{rms}$
	Modified equivalent bandwidth F_{meb}	$F_{meb} = \sqrt{(\sum_{j=1}^n (f_j - \bar{f})^2 P_j)/(\sum_{j=1}^n P_j)}$
	High–low ratio of power spectrum F_{hlps}	$F_{hlps} = (\sum_{j=n/4}^{n/2} P_i)/(\sum_{j=1}^{n/4} P_i)$
	Stabilization ratio F_{sr}	$F_{sr} = (\sum_{j=1}^n f_j^2 P_i)/(\sqrt{\sum_{j=1}^n P_i} \sqrt{\sum_{j=1}^n f_j^4 P_i})$
	Skewness of bandpower F_{sb}	$F_{sb} = (\sum_{j=1}^n (P_i - F_{mps})^3)/(\sum_{j=1}^n (P_i - F_{mps})^2)^{3/4}$
	Kurtosis of bandpower F_{kb}	$F_{sb} = n \cdot (\sum_{j=1}^n (P_i - F_{mps})^4)/(\sum_{j=1}^n (P_i - F_{mps})^2)^2$

¹ f_i is the frequency signal with x_i by FFT, P_i is the power spectrum of f_i , $\bar{f} = \sqrt{\sum_{i=1}^n f_i/n}$.

In the time–frequency domain, Wavelet transform (WT) can be used to extract candidate feature parameters. The Wavelet packet transform (WPT) conducts a multilevel band division over the entire signal band, which not only inherits the advantages of the good time–frequency localization from the WT, but it also further decomposes the high-frequency band to increase the frequency resolution [4,22,34]. Thus, the WPT was applied in order to extract the time–frequency domain features in this paper, and the wavelet energy feature is the energy of a 3-level wavelet packet decomposition using db1, which corresponds to the wavelet coefficient with a higher energy that is related to the

characteristic frequency of the machine [46]. The average energies for each frequency band of the signal were calculated with the following equation:

$$E_j = \frac{1}{n} \sum_{k=1}^n (d_{j,k})^2 = \frac{1}{n} \sum_{k=1}^n \left(\int w_{j,k}(t)x(t)dt \right)^2, \quad (7)$$

where $d_{j,k}$ ($j = 1, 2, \dots, 2^L; k = 1, 2, \dots, n$) denotes the wavelet packet coefficients of signal $x(t)$ and $w_{j,k}(t)$ are the wavelet packets localized at $2^j k$ in the scale 2^j , and L is the level of the WPT (here, $L = 3$). Taking the C1 sample as an example to show the wavelet time–frequency diagrams through WPT using db1, due to limited space, three wavelet time–frequency diagrams with the X-dimension force, X-dimension Vibration, and AE are listed in Figures 4–6. It can be found that the maximal wavelet coefficients of the X-dimension force in the C1 training sample are enhanced with the increase of tool wear, while that of other two sensors showed no obvious trend of change.

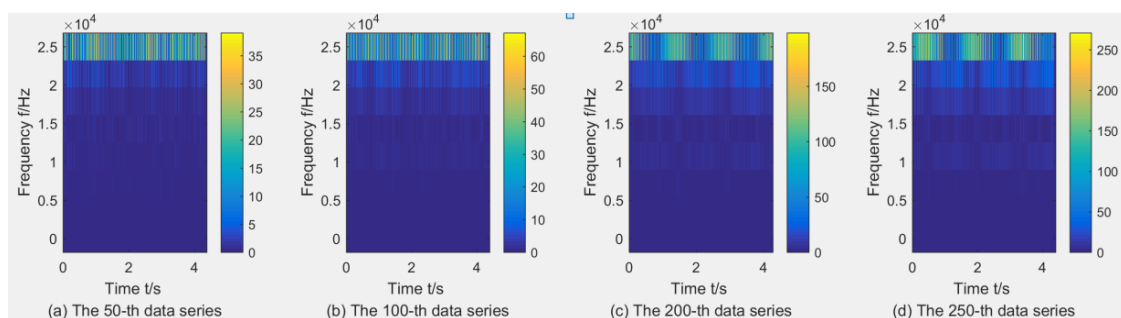


Figure 4. Wavelet time–frequency diagrams for the X-dimension force in the C1 training sample through WPT using db1

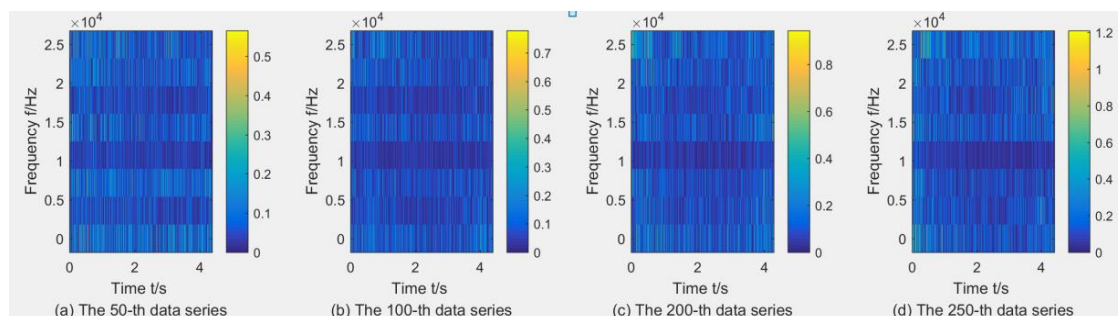


Figure 5. Wavelet time–frequency diagrams for the X-dimension Vibration in the C1 training sample through WPT using db1.

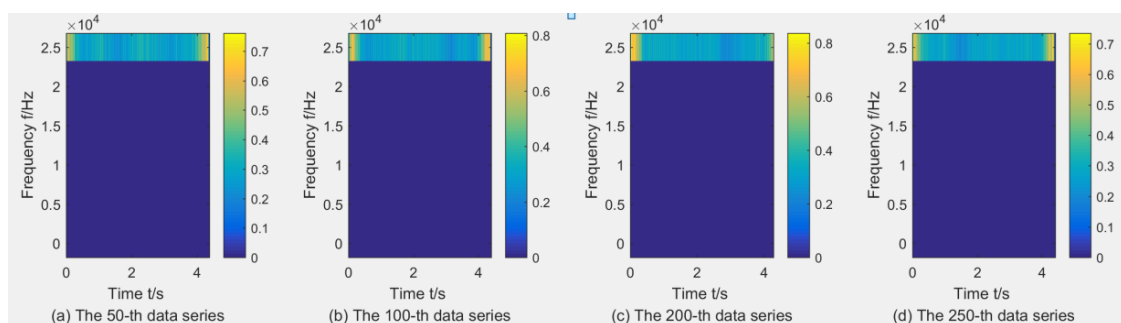


Figure 6. Wavelet time–frequency diagrams for the AE in the C1 training sample through WPT using db1.

Finally, there were $(9 + 8 + 8) \times 7 = 175$ features to be extracted from the seven sensor channels in the time, frequency, and time–frequency domains.

4.3. Results and Discussion

In the proposed GA–KELM-based global feature extraction algorithm, the input parameters used are summarized in Table 4. The optimal process of the objected function was run 10 times, and the iterations of optimal solutions are shown in Figure 7. It can be shown that the optimal result was obtained when the iteration was run about 1000 times.

Table 4. Input parameters in the proposed GA–KELM-based algorithm.

Parameters	Value
Size of the population for every generation	50
Crossover rate P_c	0.6
Mutation rate P_m	0.05
Number of iterations	1500
Regularization parameter C	6
Kernel function	Gaussian kernel
Hyperparameter of kernel	2

Eleven feature parameters were selected as the optimal feature parameter set, which involved two sensor channels, T_{rms} , F_{cf} , F_{sb} , F_{kb} , E_1 , E_2 , E_3 , E_5 , and E_6 in the X-dimension Force, and T_{sd} , and T_{cf} in the Z-dimension Force. Figures 8–18 show the scatter diagrams between these feature parameters selected and tool wear value (VB) with the C6 testing set. It can be seen that four feature parameters have no obvious correlation with the tool state, that is, F_{cf} , F_{sb} , F_{kb} in the X-dimension Force, and T_{cf} in the Z-dimension Force. This demonstrates that the parameters related most strongly to the tool state could not obtain the highest prediction accuracy, that is to say, not every parameter is strongly related to tool state in the optimal parameters corresponding to prediction accuracy.

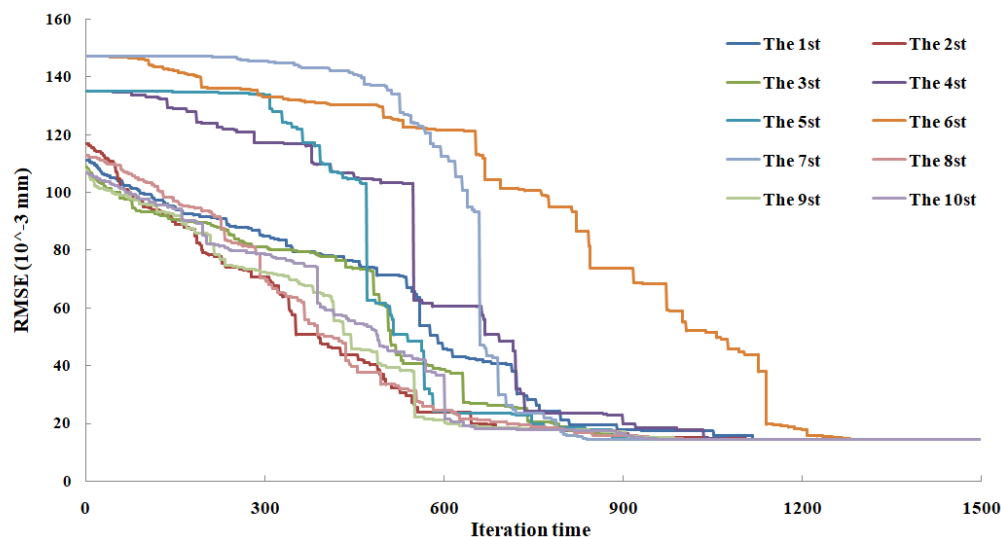


Figure 7. The optimal process of the objected function.

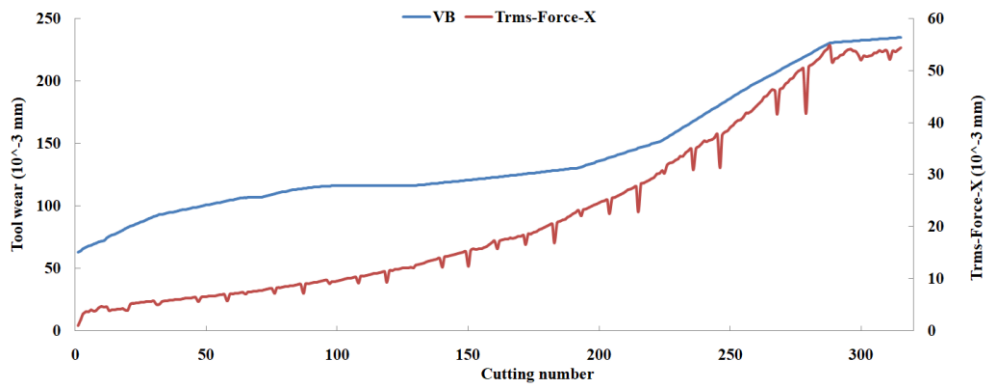


Figure 8. The scatter diagram between T_{rms} in X-dimension Force and tool wear value.

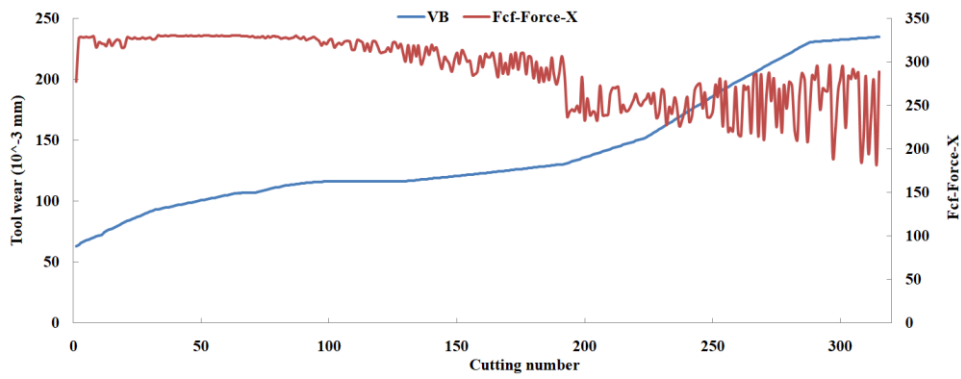


Figure 9. The scatter diagram between F_{cf} in X-dimension Force and tool wear value.

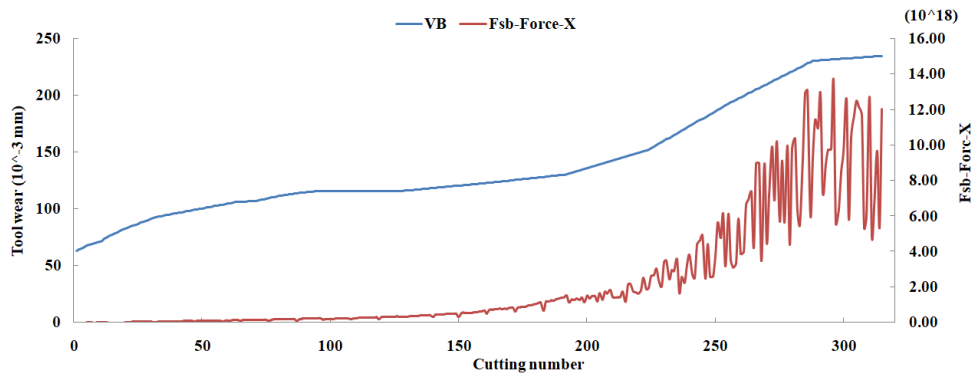


Figure 10. The scatter diagram between F_{sb} in X-dimension Force and tool wear value.

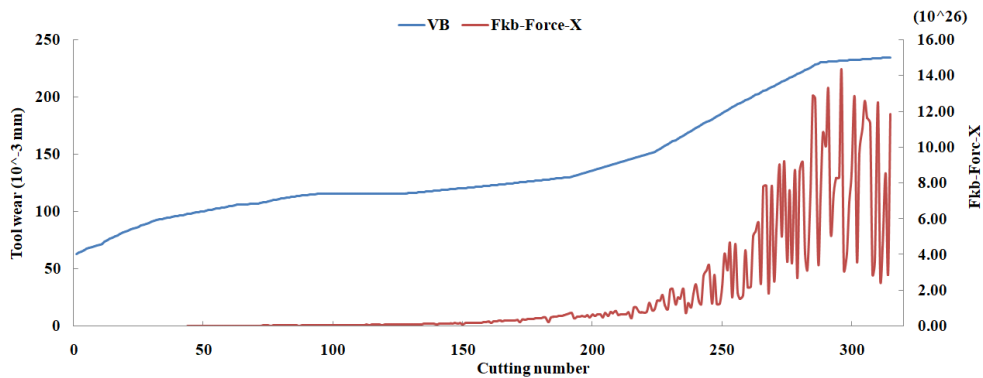


Figure 11. The scatter diagram between F_{kb} in X-dimension Force and tool wear value.

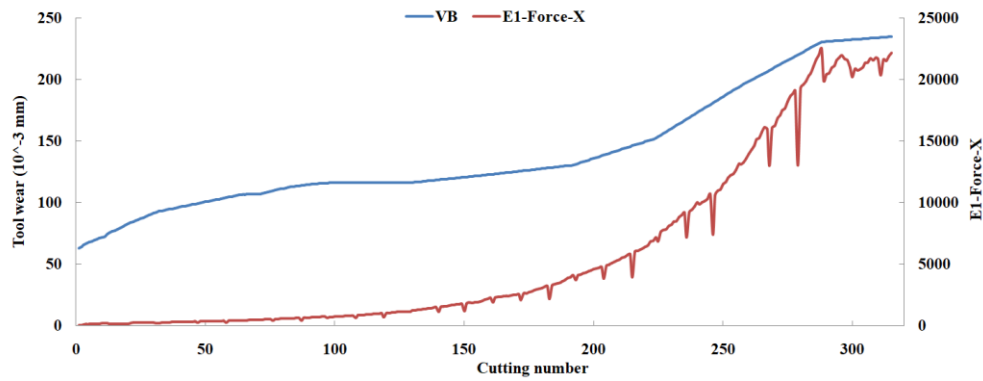


Figure 12. The scatter diagram between E_1 in X-dimension Force and tool wear value.

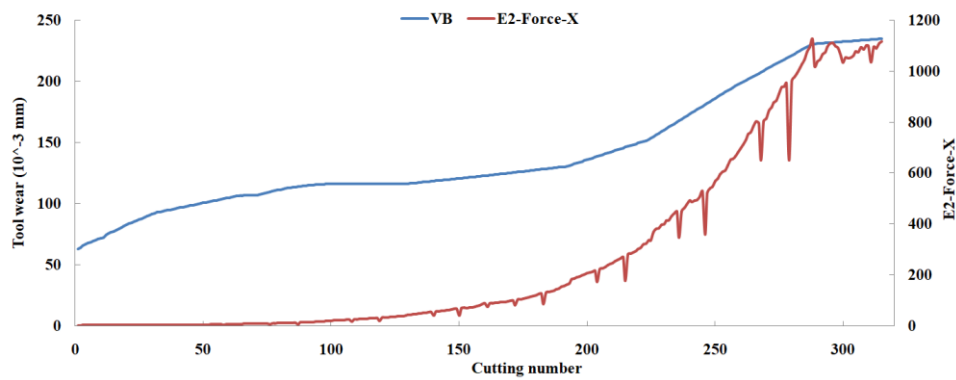


Figure 13. The scatter diagram between E_2 in X-dimension Force and tool wear value.

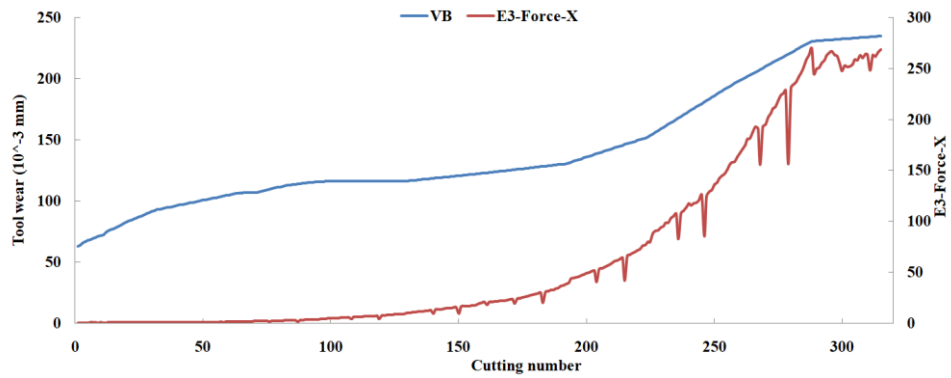


Figure 14. The scatter diagram between E_3 in X-dimension Force and tool wear value.

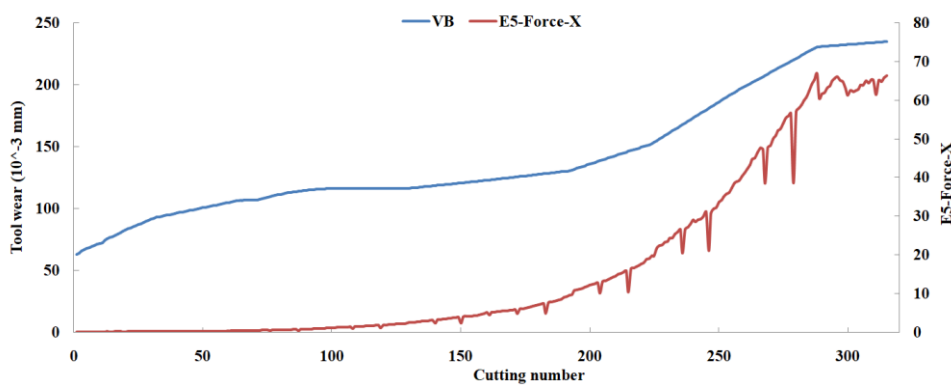


Figure 15. The scatter diagram between E_5 in X-dimension Force and tool wear value.

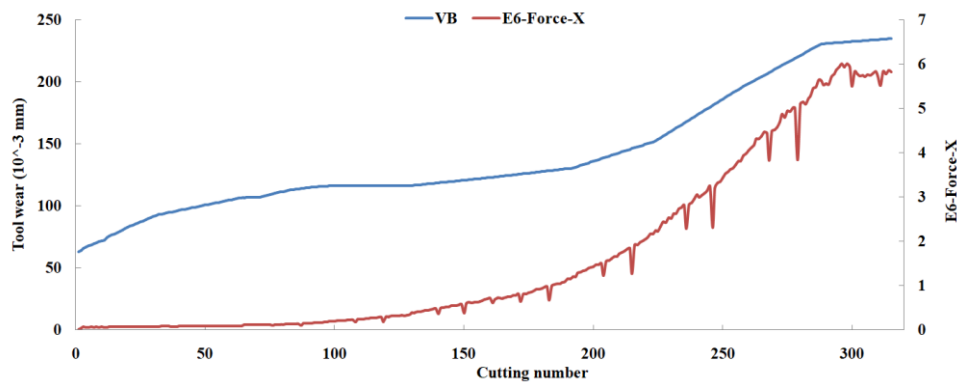


Figure 16. The scatter diagram between E_6 in X-dimension Force and tool wear value.

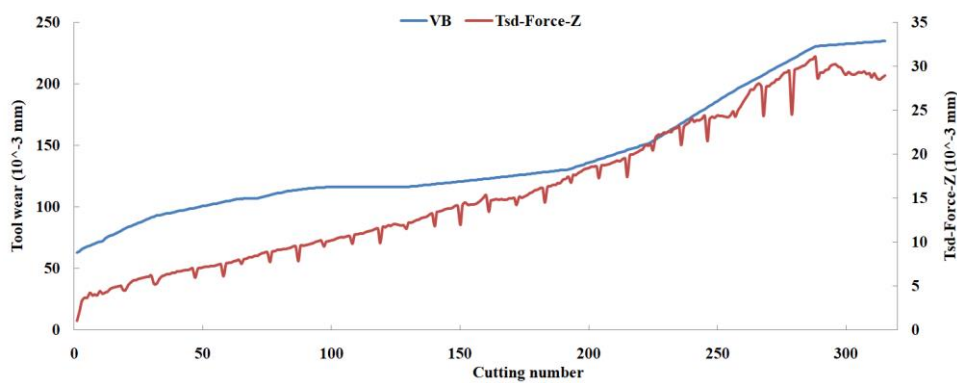


Figure 17. The scatter diagram between T_{sd} in Z-dimension Force and tool wear value.

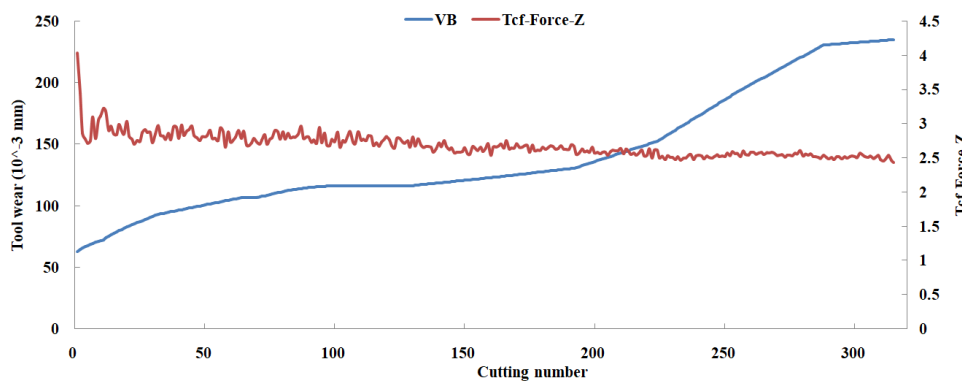


Figure 18. The scatter diagram between T_{cf} in Z-dimension Force and tool wear value.

The predicted tool wear values with the C6 testing set are shown in Figure 19. It was found that the overall error of the predicted tool wear was small, and the proposed method could reveal the trend of the tool wear state well. In addition, three benchmark methods, PCC-based, mRMR-based, and PCA-based, were compared with the proposed method. The PCC [24,47,48], mRMR [34,35], and PCA [31,32] methods are often used to extract feature parameters or principal components in TCM according to linear or nonlinear correlation. In the PCC-based method, these feature parameters of the correlation coefficient with tool wear values above 0.8 were selected as the inputs of the KELM in this study, and 33 parameters were retained in the final features. In the mRMR-based method, 19 parameters were selected in the final features by considering the maximum relevance and the minimum redundancy criterion simultaneously. In the PCA-based method, principal components were obtained through solving the eigenvalues of the correlation matrix of training data. The cumulative contribution rate of the first 14 principal components reached 90.23%, and then the first 14 principal components were selected as the inputs of the KELM. The predicted tool wear values of the three

methods with the C6 testing set are shown in Figure 16, and several performance indexes (RMSE and R^2 with the truth data, the number of selected parameters, and the number of sensor channels involved) of the four methods are shown in Table 5. It can be shown that the proposed method outperforms the other three methods in terms of prediction accuracy and the number of selected parameters.

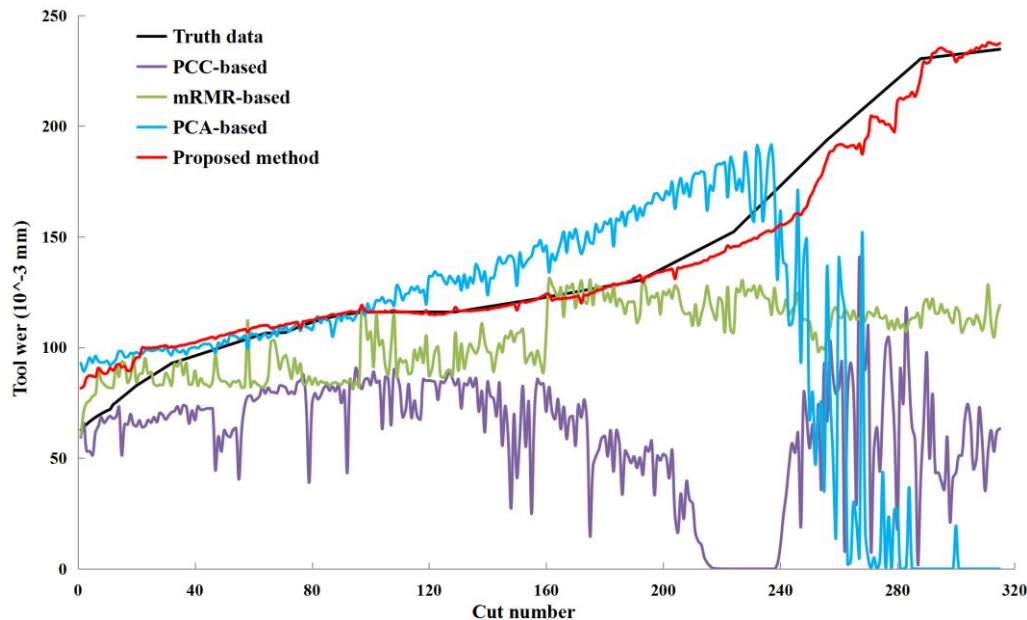


Figure 19. Predicted tool wear values of several methods with the C6 testing set.

Table 5. Several performance indexes of three methods on the testing set.

Methods	RMSE	R^2	Number of Selected Parameters	Number of Sensor Channels Involved
The PCC-based method	98.339	−0.198	33	5
The PCA-based method	94.665	−0.5768	175	7
The mRMR-based method	53.268	0.598	19	5
The proposed method	24.711	0.988	11	2

5. Conclusions

The present study proposed a multisensor global fusion method for TCM in the milling process. Several statistical parameters in the time, frequency, and time–frequency domains of multiple sensors were selected as an alternative parameter set. The GA was applied in order to search the optimal parameter combinations to achieve the highest prediction precision, while the monitoring model was executed by a KELM. The experimental results show that the proposed method outperforms the PCC-based and mRMR-based methods. The RMSE and R^2 values for the truth data, the number of selected parameters, and the number of sensor channels of the proposed method were better than those of the other two methods.

Author Contributions: Software, Validation, and Writing—Original Draft Preparation, Y.Z.; Methodology, Writing—Review and Editing, and Project Administration, W.X. All authors have read and approved the final manuscript.

Funding: This research was funded by the National Science Foundation of China (Grant No. 51405346), the Zhejiang Provincial Natural Science Foundation of China (No. LY17E050005), and the Wenzhou City Public Industrial Science and technology project of China (No. G20160015, G20170009).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Javed, K.; Gouriveau, R.; Li, X.; Zerhouni, N. Tool wear monitoring and prognostics challenges: A comparison of connectionist methods toward an adaptive ensemble model. *J. Intell. Manuf.* **2016**, 1–18. [[CrossRef](#)]
2. Vetrichelvan, G.; Sundaram, S.; Kumaran, S.S.; Velmurugan, P. An investigation of tool wear using acoustic emission and genetic algorithm. *J. Vib. Control* **2014**, *21*, 3061–3066. [[CrossRef](#)]
3. Bhattacharyya, P.; Sengupta, D.; Mukhopadhyay, S. Cutting force-based real-time estimation of tool wear in face milling using a combination of signal processing techniques. *Mech. Syst. Signal Process.* **2007**, *21*, 2665–2683. [[CrossRef](#)]
4. Liu, C.; Wang, G.F.; Li, Z.M. Incremental learning for online tool condition monitoring using ellipsoid artmap network model. *Appl. Soft Comput.* **2015**, *35*, 186–198. [[CrossRef](#)]
5. Aliustaoglu, C.; Ertunc, H.M.; Oca, H. Tool wear condition monitoring using a sensor fusion model based on fuzzy inference system. *Mech. Syst. Signal Process.* **2009**, *23*, 539–546. [[CrossRef](#)]
6. Konstantinos, S.; Athanasios, K. Reliability assessment of cutting tool life based on surrogate approximation methods. *Int. J. Adv. Manuf. Technol.* **2014**, *71*, 1197–1208.
7. Karandikar, J.; Mcleay, T.; Turner, S.; Schmitz, T. Tool wear monitoring using naïve bayes classifiers. *Int. J. Adv. Manuf. Technol.* **2015**, *77*, 1613–1626. [[CrossRef](#)]
8. Gao, R.; Wang, L.; Teti, R.; Dornfeld, D.; Kumara, S.; Mori, M.; Helu, M. Cloud-enabled prognosis for manufacturing. *CIRP Ann.* **2015**, *64*, 749–772. [[CrossRef](#)]
9. Rehorn, A.G.; Jiang, J.; Orban, P.E. State-of-the-art methods and results in tool condition monitoring: A review. *Int. J. Adv. Manuf. Technol.* **2005**, *26*, 693–710. [[CrossRef](#)]
10. Dutta, S.; Kanwat, A.; Pal, S.K.; Sen, R. Correlation study of tool flank wear with machined surface texture in end milling. *Measurement* **2013**, *46*, 4249–4260. [[CrossRef](#)]
11. Ghosh, N.; Ravi, Y.B.; Patra, A.; Mukhopadhyay, S.; Paul, S.; Mohanty, A.R.; Chattopadhyay, A.B. Estimation of tool wear during CNC milling using neural network-based sensor fusion. *Mech. Syst. Signal Process.* **2007**, *21*, 466–479. [[CrossRef](#)]
12. Drouillet, C.; Karandikar, J.; Nath, C.; Journeaux, A.C.; Mansori, M.E.; Kurfess, T. Tool life predictions in milling using spindle power with the neural network technique. *J. Manuf. Process.* **2016**, *22*, 161–168. [[CrossRef](#)]
13. Chryssolouris, G.; Domroese, M.; Beaulieu, P. Sensor Synthesis for Control of Manufacturing Processes. *J. Eng. Ind. ASME* **1992**, *114*, 158–174. [[CrossRef](#)]
14. Nouri, M.; Fussell, B.K.; Ziniti, B.L.; Linder, E. Real-time tool wear monitoring in milling using a cutting condition independent method. *Int. J. Mach. Tools Manuf.* **2015**, *89*, 1–13. [[CrossRef](#)]
15. Azmi, A.I. Monitoring of tool wear using measured machining forces and neuro-fuzzy modelling approaches during machining of gfrp composites. *Adv. Eng. Softw.* **2015**, *82*, 53–64. [[CrossRef](#)]
16. Zhang, H.; Zhao, J.; Wang, F.; Li, A. Cutting forces and tool failure in high-speed milling of titanium alloy tc21 with coated carbide tools. *Proc. Inst. Mech. Eng. Part B J. Eng. Manuf.* **2015**, *229*, 20–27. [[CrossRef](#)]
17. Stavropoulos, P.; Papacharalampopoulos, A.; Vasiliadis, E.; Chryssolouris, G. Tool wear predictability estimation in milling based on multi-sensorial data. *Int. J. Adv. Manuf. Technol.* **2016**, *82*, 509–521. [[CrossRef](#)]
18. Snr, D.E.D. Sensor signals for tool-wear monitoring in metal cutting operations—A review of methods. *Int. J. Mach. Tools Manuf.* **2000**, *40*, 1073–1098.
19. Rizal, M.; Ghani, J.A.; Nuawi, M.Z.; Che, H.C.H. A review of sensor system and application in milling process for tool condition monitoring. *Res. J. Appl. Sci. Eng. Technol.* **2014**, *7*, 2083–2097. [[CrossRef](#)]
20. Chen, S.L.; Jen, Y.W. Data fusion neural network for tool condition monitoring in cnc milling machining. *Int. J. Mach. Tools Manuf.* **2000**, *40*, 381–400. [[CrossRef](#)]
21. Zhou, Y.Q.; Xue, W. Review of tool condition monitoring methods in milling processes. *Int. J. Adv. Manuf. Technol.* **2018**, *96*, 2509–2523. [[CrossRef](#)]
22. Zhang, C.; Yao, X.; Zhang, J.; Jin, H. Tool condition monitoring and remaining useful life prognostic based on a wireless sensor in dry milling operations. *Sensors* **2016**, *16*, 795. [[CrossRef](#)] [[PubMed](#)]

23. Suprock, C.A.; Roth, J.T. Methods for on-line directionally independent failure prediction of end milling cutting tools. *Mach. Sci. Technol.* **2007**, *11*, 1–43. [[CrossRef](#)]
24. Huang, P.B.; Ma, C.C.; Kuo, C.H. A PNN self-learning tool breakage detection system in end milling operations. *Appl. Soft Comput.* **2015**, *37*, 114–124. [[CrossRef](#)]
25. Cuka, B.; Kim, D.W. Fuzzy logic based tool condition monitoring for end-milling. *Robot. Comput.-Integrated Manuf.* **2017**, *47*, 23–36. [[CrossRef](#)]
26. Zhu, K.; Wong, Y.S.; Hong, G.S. Wavelet analysis of sensor signals for tool condition monitoring: A review and some new results. *Int. J. Mach. Tools Manuf.* **2009**, *49*, 537–553. [[CrossRef](#)]
27. Madhusudana, C.K.; Kumar, H.; Narendranath, S. Face milling tool condition monitoring using sound signal. *Int. J. Syst. Assurance Eng. Manag.* **2017**, *2017*, 1–11. [[CrossRef](#)]
28. Sevilla-Camacho, P.Y.; Robles-Ocampo, J.B.; Jauregui-Correa, J.C.; Jimenez-Villalobos, D. FPGA-based reconfigurable system for tool condition monitoring in high-speed machining process. *Measurement* **2015**, *64*, 81–88. [[CrossRef](#)]
29. Liu, C.; Li, Y.; Zhou, G.; Shen, W. A sensor fusion and support vector machine based approach for recognition of complex machining conditions. *J. Intell. Manuf.* **2016**, 1–14. [[CrossRef](#)]
30. Wang, M.; Wang, J. CHMM for tool condition monitoring and remaining useful life prediction. *Int. J. Adv. Manuf. Technol.* **2012**, *59*, 463–471. [[CrossRef](#)]
31. Grasso, M.; Albertelli, P.; Colosimo, B.M. An Adaptive SPC Approach for Multi-sensor Fusion and Monitoring of Time-varying Processes. *Procedia Cirp* **2013**, *12*, 61–66. [[CrossRef](#)]
32. Wang, G.; Zhang, Y.; Liu, C.; Xie, Q.; Xu, Y. A new tool wear monitoring method based on multi-scale pca. *J. Intell. Manuf.* **2016**, *7*, 1–10. [[CrossRef](#)]
33. Wang, G.F.; Yang, Y.W.; Zhang, Y.C.; Xie, Q.L. Vibration sensor based tool condition monitoring using ν , support vector machine and locality preserving projection. *Sens. Actuators A Phys.* **2014**, *209*, 24–32. [[CrossRef](#)]
34. Wang, J.; Xie, J.; Zhao, R.; Zhang, L.; Duan, L. Multisensory fusion based virtual tool wear sensing for ubiquitous manufacturing. *Robot. Comput. Integr. Manuf.* **2017**, *45*, 47–58. [[CrossRef](#)]
35. Wang, G.; Yang, Y.; Li, Z. Force sensor based tool condition monitoring using a heterogeneous ensemble learning model. *Sensors* **2014**, *14*, 21588–21602. [[CrossRef](#)] [[PubMed](#)]
36. Cho, S.; Binsaeid, S.; Asfour, S. Design of multisensor fusion-based tool condition monitoring system in end milling. *Int. J. Adv. Manuf. Technol.* **2010**, *46*, 681–694. [[CrossRef](#)]
37. Huang, G.B.; Zhou, H.; Ding, X.; Zhang, R. Extreme learning machine for regression and multiclass classification. *IEEE Trans. Syst. Man Cybern. Part B* **2012**, *42*, 513–529. [[CrossRef](#)] [[PubMed](#)]
38. Huang, G.B. What are Extreme Learning Machines? Filling the Gap Between Frank Rosenblatt’s Dream and John von Neumann’s Puzzle. *Cognit. Comput.* **2015**, *7*, 263–278. [[CrossRef](#)]
39. Chen, X.; Peng, X.; Li, J.; Peng, Y. Overview of Deep Kernel Learning Based Techniques and Applications. *J. Netw. Intell.* **2016**, *1*, 83–98.
40. Maslov, I.V.; Gertner, I. Multi-sensor fusion: An Evolutionary algorithm approach. *Inf. Fusion* **2006**, *7*, 304–330. [[CrossRef](#)]
41. The Prognostics and Health Management Society, 2010 Conference Data Challenge. Available online: <https://www.phmsociety.org/competition/phm/10> (accessed on 1 June 2018).
42. Benkedjouh, T.; Zerhouni, N.; Rechak, S. Tool wear condition monitoring based on continuous wavelet transform and blind source separation. *Int. J. Adv. Manuf. Technol.* **2018**, *97*, 3311–3323. [[CrossRef](#)]
43. Zhou, Y.; Liu, X.; Li, F.; Bingtao, S.; Wei, X. An online damage identification approach for numerical control machine tools based on data fusion using vibration signals. *J. Vib. Control* **2015**, *21*, 2925–2936.
44. Zhao, R.; Yan, R.; Wang, J.; Mao, K. Learning to Monitor Machine Health with Convolutional Bi-Directional LSTM Networks. *Sensors* **2017**, *17*, 273. [[CrossRef](#)] [[PubMed](#)]
45. Gao, C.; Xue, W.; Ren, Y.; Zhou, Y. Numerical control machine tool fault diagnosis using hybrid stationary subspace analysis and least squares support vector machine with a single sensor. *Appl. Sci.* **2017**, *7*, 346. [[CrossRef](#)]
46. Zhao, R.; Wang, D.; Yan, R.; Mao, K.; Shen, F.; Wang, J. Machine Health Monitoring Using Local Feature-based Gated Recurrent Unit Networks. *IEEE Trans. Ind. Electron.* **2017**, *65*, 1539–1548. [[CrossRef](#)]

47. Chen, B.; Chen, X.; Li, B.; He, Z.; Cao, H.; Cai, G. Reliability estimation for cutting tools based on logistic regression model using vibration signals. *Noise Vib. Bull.* **2011**, *25*, 2526–2537. [[CrossRef](#)]
48. Wang, G.; Grosu, R. Milling-tool wear-condition prediction with statistic analysis and echo-state networks. *Challenges Technol. Innov.* **2017**, 149–153.



© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).