Gaussian process regression for tool wear prediction

Article *in* Mechanical Systems and Signal Processing · May 2018

DOI: 10.1016/j.ymssp.2017.11.021

CITATIONS
68

READS
252

3 authors, including:

Dongdong Kong
Shanghai University
15 PUBLICATIONS **263** CITATIONS
SEE PROFILE

Ning Li
14 PUBLICATIONS **273** CITATIONS
SEE PROFILE

Mechanical Systems and Signal Processing 104 (2018) 556–574

Contents lists available at ScienceDirect

Mechanical Systems and Signal Processing

journal homepage: www.elsevier.com/locate/ymssp

Gaussian process regression for tool wear prediction

Dongdong Kong *, Yongjie Chen, Ning Li

School of Mechanical Science & Engineering, Huazhong University of Science and Technology, Wuhan, China

ARTICLE INFO

Article history:
Received 11 August 2017
Received in revised form 16 October 2017
Accepted 14 November 2017
Available online 21 November 2017

Keywords:
Tool wear prediction
Feature fusion
Gaussian process regression

ABSTRACT

To realize and accelerate the pace of intelligent manufacturing, this paper presents a novel tool wear assessment technique based on the integrated radial basis function based kernel principal component analysis (KPCA_IRBF) and Gaussian process regression (GPR) for real-timely and accurately monitoring the in-process tool wear parameters (flank wear width). The KPCA_IRBF is a kind of new nonlinear dimension-increment technique and firstly proposed for feature fusion. The tool wear predictive value and the corresponding confidence interval are both provided by utilizing the GPR model. Besides, GPR performs better than artificial neural networks (ANN) and support vector machines (SVM) in prediction accuracy since the Gaussian noises can be modeled quantitatively in the GPR model. However, the existence of noises will affect the stability of the confidence interval seriously. In this work, the proposed KPCA_IRBF technique helps to remove the noises and weaken its negative effects so as to make the confidence interval compressed greatly and more smoothed, which is conducive for monitoring the tool wear accurately. Moreover, the selection of kernel parameter in KPCA_IRBF can be easily carried out in a much larger selectable region in comparison with the conventional KPCA_RBF technique, which helps to improve the efficiency of model construction. Ten sets of cutting tests are conducted to validate the effectiveness of the presented tool wear assessment technique. The experimental results show that the in-process flank wear width of tool inserts can be monitored accurately by utilizing the presented tool wear assessment technique which is robust under a variety of cutting conditions. This study lays the foundation for tool wear monitoring in real industrial settings.

1. Introduction

Tool wear status is one of the crucial factors for guaranteeing the reliability and stability of manufacturing system since the excessive wear of cutting tools will give rise to sharp increase in cutting forces and even the chatter of machine tools. Besides, tool failure leads to as much as 20% of the downtime in modern manufacturing systems, which results in declining productivity [1]. To solve the above problems, an efficient monitoring system is highly desirable for real-timely and accurately evaluating the wear status of cutting tools so as to make maintenance suggestions. As modern manufacturing systems are becoming more and more complex, the demand for reliable and stable operation also requires the monitoring system can provide early warnings for tool wear, which is investigated in this work.

* Corresponding author at: School of Mechanical Science & Engineering, Huazhong University of Science and Technology, 1037 Luoyu Road, Wuhan, China.
E-mail address: kodon007@163.com (D. Kong).

https://doi.org/10.1016/j.ymssp.2017.11.021
0888-3270/

Many efforts have been devoted to the development of tool wear monitoring which mainly consists of signal acquisition, feature extraction and system identification [2,3]. An integrated tool wear monitoring system is illustrated in Fig. 1. The monitoring scopes for tool wear in machining process mainly include wear width, wear area, wear state and breakage. Commonly used signals and the corresponding sensors that have been proved to be feasible for tool wear monitoring are reviewed in [3], including cutting forces (dynamometer) [5–7,9,13–15], vibration (accelerometer) [8,9,13,15] and acoustic emission (AE sensor) [15] and spindle current or power [5,8,10]. As for system identification, artificial neural networks (ANN) [4–8] and support vector machines (SVM) [9–14] are the most widely used artificial intelligence models for monitoring the flank wear width of cutting tools. Ghosh et al. [5] developed an ANN-based sensor fusion model for tool wear prediction and found that multi-sensor fusion improved the prediction accuracy over their single sensor counterparts. By fusing the cutting forces, torque, cutting conditions and cutting time, Kaya et al. [6] developed a robust ANN-based flank wear monitoring system for milling of Inconel 718. Xu et al. [7] achieved the corner-wear prediction of a high speed steel drill by utilizing back propagation neural network (BPNN) and the features extracted by wavelet packet transform. Salgado and Alonso [10] and Zhang [11] realized tool wear prediction by utilizing least squares support vector machines (LS-SVM) and found that the LS-SVM-based model performed better than the ANN-based model in prediction accuracy. Dutta et al. [12] realized tool flank wear prediction by utilizing ε-support vector regression (ε-SVR) and the texture features obtained from the turned surface images, with a 4.9% maximum predicted error. Kong et al. [14] developed a tool wear predictive model based on kernel principal component analysis (KPCA) and v-support vector regression (v-SVR). They found that the constructed model outperformed the BPNN-based model in prediction accuracy. Wu et al. [15] realized tool wear prediction in milling operations by utilizing random forests (RFs) and found that RFs performed better than ANN and SVR in prediction accuracy. These works [4–15] solved the problem of monitoring the flank wear width of cutting tools in specific areas. However, they cannot provide the uncertainties of the predicted results. The analysis and quantification of uncertainty is of great importance for the improvement of product quality and efficiency, which can be incorporated into modern manufacturing systems.

In this work, Gaussian process regression (GPR) [16] is utilized to construct an effective tool wear predictive model between the signal features and the flank wear width of tool inserts. The GPR model is a flexible and powerful tool that can provide both the predicted results and the confidence interval (uncertainty estimations) which helps to quantify the reliability of predictions [17]. GPR has been widely used in data-driven modelling for various systems. Wang and Chen [18] proposed a multi-response GPR model and demonstrated its superiority through the simulated examples and modelling the response surface of a catalytic reaction process. Zhang et al. [19] proposed a hybrid AR-GPR model for wind speed prediction and found that the proposed model performed better than ANN and SVM. Wang and Chaib-draa [20] proposed a novel KNN-KFGP model to deal with the large-scale or non-stationary datasets and achieved obvious performance improvement in accuracy and efficiency. Aye and Heyns [21] proposed an integrated GPR model for the prediction of remaining useful life of slow speed bearings and achieved lower prediction error. Liu et al. [22] proposed a multi-output GPR model for the fault prognosis of filamentous bulking sludge and showed that the multi-output strategy outperformed the single-output strategy in predic-

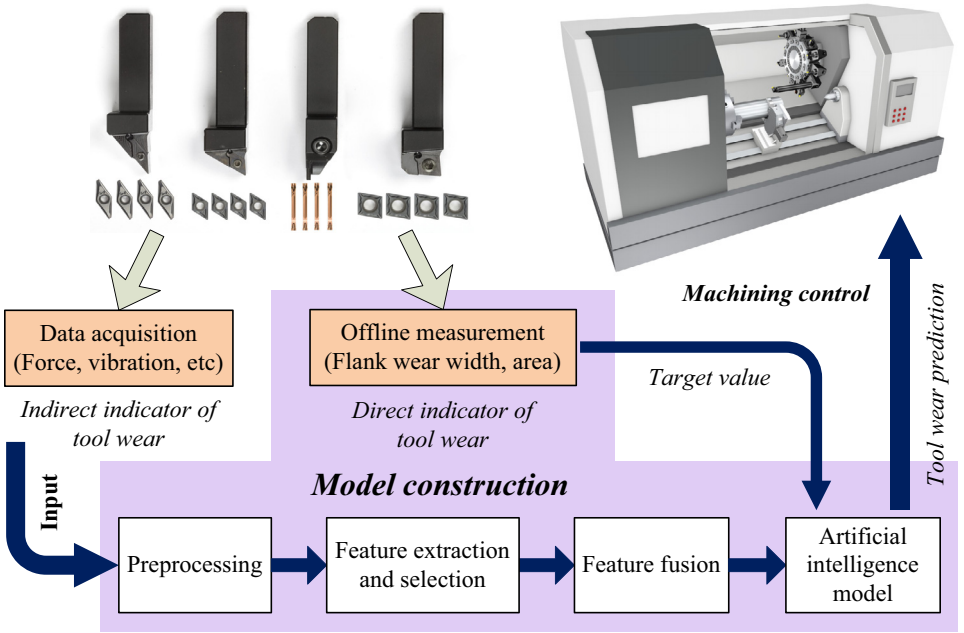

Fig. 1. Flowchart of an integrated tool wear monitoring system.

tion accuracy. Most of the literatures [18–22] focused on the extension and variation of the GPR model so as to improve the prediction accuracy. However, the analysis and amelioration of the confidence interval of the GPR model receives little attention to our knowledge. It should be noted that the existence of Gaussian noises has little impact on the prediction accuracy of GPR since the noises can be modeled quantitatively in the GPR model. Nevertheless, the existence of Gaussian noises can lead to large fluctuation of the confidence interval. If the extracted features that reflect tool wear status are directly sent to GPR, the confidence interval of the GPR-based tool wear predictive model will fluctuate greatly due to the existence of in-process noises which are inevitable in machining process. This will reduce the reliability of the predicted results and undermine the stability of the monitoring process.

In general, dimension-reduction [23,24] is a practical manner to remove noises and enhance the computational efficiency. In this work, the KPCA technique is utilized to fuse the extracted features so as to weaken the negative effects of in-process noises and ameliorate the confidence interval of the GPR model. The basic idea of KPCA is to carry out principal component analysis in a feature space as a substitute for nonlinear principal component analysis in the input space by recurring to nonlinear mapping [25]. Typically, the performance of KPCA lies in the selection of kernel function and the optimization of kernel parameters. The most widely used kernel in KPCA is the radial basis function (RBF). The conventional KPCA_RBF technique achieves not only noise removal but also dimension-reduction which will conduce to enhance the computational efficiency. However, the selection of a suitable kernel parameter for KPCA_RBF is still an open problem that has not been solved completely [26]. It is confirmed that RBF with an extremely large (or small) kernel parameter will under-fit (or over-fit) the given data [27]. Besides, the number of the retained principal components for KPCA_RBF that preserve the valid information as much as possible is hard to be determined. To tackle this problem, the integrated radial basis function (IRBF) [27] is employed in KPCA so as to realize dimension-increment. The proposed KPCA_IRBF technique realizes noise removal and dimension-increment simultaneously, which can preserve and enrich the valid information related to tool wear. In addition, the selectable region of kernel parameters of KPCA_IRBF is much larger than that of KPCA_RBF since IRBF have stronger global learning ability than RBF. Thus, finding a good parameter in KPCA_IRBF is not an intractable problem. Experimental results show that the utilization of KPCA_IRBF will be conducive to making the confidence interval of the GPR model compressed greatly and more smoothed.

This study aims at real-timely and accurately monitoring the in-process flank wear width of tool inserts by utilizing the proposed KPCA_IRBF technique and the GPR model. The paper proceeds as follows. The experimental setup, data collection and related theoretical methods are given in Section 2. Analysis of the presented tool wear assessment technique and the experimental results are given in Section 3. Future research directions are provided in Section 4. Finally, Section 5 concludes this paper.

2. Materials and methods

2.1. Experimental setup and data collection

The cutting tests of dry turning 50# normalized steel with cemented carbide tool inserts are conducted to validate the effectiveness of the presented tool wear assessment technique. The three orthogonal cutting forces in machining process are utilized to follow the tracks of tool flank wear width. Diagrammatic sketch of the experimental setup for tool wear monitoring is illustrated in Fig. 2. Further details of the experimental setup are available in [14]. Machining parameters in the ten sets of cutting tests are listed in Table 1. The original cutting force signals are collected at sampling rate of 20 kHz by using the dynamometer. Besides, the flank wear width *VB* of tool inserts is measured and recorded after each cutting process (at set intervals) by using the Video Measuring System (VMS) until tool breakage ($VB > 0.5$ mm).

The time-domain, frequency-domain and wavelet-domain features as listed in Table 2 are extracted from the original cutting force signals within one second and used as the pre-fusion features. In the time-domain, six types of statistical features are extracted. Mean ($\overline{F_x}, \overline{F_y}, \overline{F_z}$) reflects the central tendency of a signal. Root mean square (RMS) represents the average energy of a signal within a given time interval. Maximum (Max) represents the maximum instantaneous amplitude of a signal. Standard deviation (Std) reflects the degree of fluctuation of a signal around the mean value. Skewness (Ske) reflects the asymmetry of a signal around the symmetry line which refers to the mean value. Kurtosis (Kur) reflects the transient phenomena and stationarity of a signal. In the frequency-domain, two features that quantify the frequency property of a signal are extracted. f_i is the frequency spectrum converted from the time-domain signal (i.e., the original signal) by using the Fast Fourier Transform (FFT), $P(f_i)$ is the power spectral density, N is the length of power spectral. In the wavelet-domain, wavelet packet decomposition [28,29] is utilized to decompose the original signal into multiple sub-bands effectively. Average energy $E_i (1 \leqslant i \leqslant 2^L)$ of the wavelet packet coefficients in each sub-band can provide more detailed information that reflects tool wear status from the perspective of time–frequency domain. $d_{i,t}^{(L)} (t = 1, \cdots, N_i)$ are the wavelet packet coefficients in the i th sub-band, where L is the level of decomposition, N_i is the total number of wavelet packet coefficients in the i th sub-band. In this work, the original cutting force signals are decomposed at level-3 with 'db3' wavelet packets using Shannon entropy. Thus, eight wavelet-domain features can be extracted from each cutting force signal by recurring to wavelet packet decomposition. The features listed in Table 2 are extracted from the three orthogonal cutting forces and altogether 48 signal features are obtained.

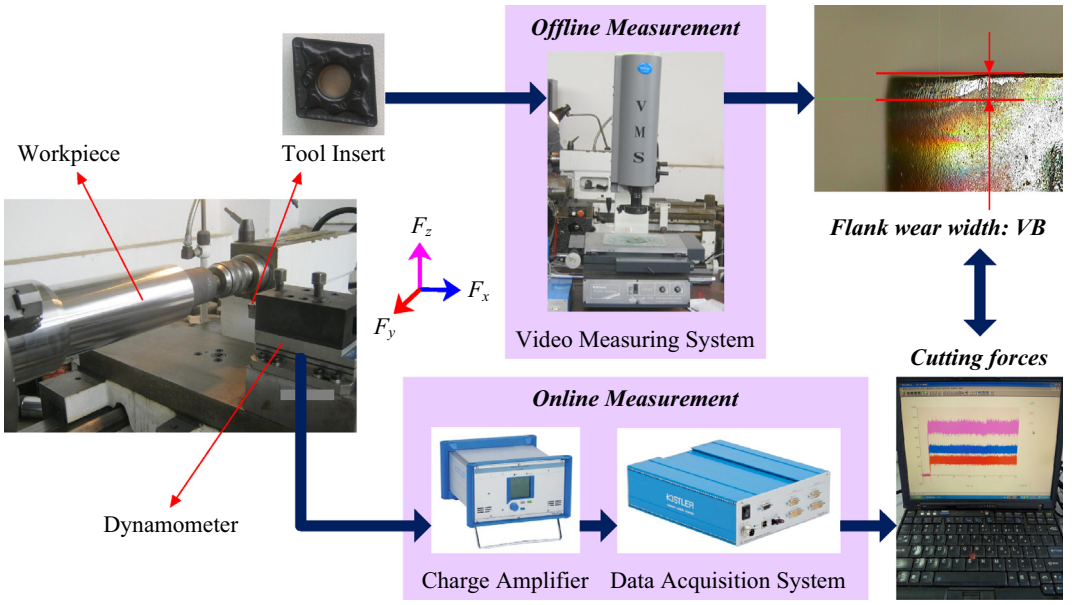

Fig. 2. The experimental setup.

Table 1
Machining parameters in the ten sets of cutting tests.

Test No.	Cutting speed V_c(m/min)	Feed rate f(mm/r)	Cutting depth a_p(mm)	Replications
1	300	0.2	1	2
2	300	0.2	2	2
3	300	0.3	1	2
4	300	0.3	2	2
5	300	0.4	1	1
6	300	0.4	2	1

Table 2
Mathematical equations of the extracted signal features.

Domain	Signal Features	Expression
Time-domain	Mean ($\overline{F_x}, \overline{F_y}, \overline{F_z}$)	$\mu = E(\lvert x_i \rvert)$
	Root mean square (RMS)	$x_{RMS} = \{E(x_i^2)\}^{1/2}$
	Maximum (Max)	$x_{Max} = max(\lvert x_i \rvert)$
	Standard deviation (Std)	$x_{Std} = \sigma = \{E[(\lvert x_i \rvert - \mu)^2]\}^{1/2}$
	Skewness (Ske)	$x_{Ske} = E\{[(\lvert x_i \rvert - \mu)/\sigma]^3\}$
	Kurtosis (Kur)	$x_{Kur} = E\{[(\lvert x_i \rvert - \mu)/\sigma]^4\}$
Frequency-domain	Frequency centroid (FC)	$x_{FC} = \frac{\sum_{i=1}^{N} f_i \cdot P(f_i)}{\sum_{i=1}^{N} P(f_i)}$
	Frequency variance (FV)	$x_{FV} = \frac{\sum_{i=1}^{N} (f_i - x_{FC})^2 \cdot P(f_i)}{\sum_{i=1}^{N} P(f_i)}$
Wavelet-domain	Wavelet packet energy	$E_i = \frac{1}{N_i} \sum_{t=1}^{N_i} (d_{i,t}^{(L)})^2$

The flank wear width *VB* of tool inserts is adopted as the subject investigated in this study. Altogether 215 data files are collected and recorded in the ten sets of cutting tests as shown in Table 1. One data file corresponds to one case number. What's more, it's necessary to perform feature fusion so as to remove the noises and weaken its negative effects. The extracted signal features or the fused features and the corresponding machining parameters make up the feature vectors which are utilized as the input of the GPR-based tool wear predictive model. A total of 430 samples of feature vectors are obtained and equally divided into training and test dataset which do not overlap each other. Also notice that the extracted signal features need to be normalized before feature fusion, which is given by

$$x' = \frac{x - \bar{x}}{\sigma_x} \tag{1}$$

where $\bar{x}$ is the mean value, σ_x is the standard deviation.

2.2. The proposed KPCA_IRBF technique

Kernel principal component analysis (KPCA) is a new method for implementing the nonlinear form of principal component analysis (PCA) by recurring to the 'kernel tricks' [25]. The keystone of KPCA lies in the kernel function in input space which represents the dot product of the mapped vector in feature space. Thus, there is no need to solve the explicit calculation of the mapping from input space to feature space which makes the implementation of KPCA become more operable.

Supposing that the training data are represented by a matrix $X_{n\times m}$. m is the size of training data, n is the dimension of feature vectors. $\Phi : \boldsymbol{x_i} \rightarrow \Phi(\boldsymbol{x_i})$ represents that the data in input space $\boldsymbol{R^n}$ are nonlinearly mapped into the feature space $\boldsymbol{F}$. Assuming that $\Phi(\boldsymbol{x_i})$ satisfies the condition: $\sum_{i=1}^{m}\Phi(\boldsymbol{x_i}) = 0$. The specific implementation process of KPCA is as follows:

Step 1: Calculate the kernel matrix $\boldsymbol{K} = [K_{ij}]_{m\times m}$ by utilizing the kernel function:

$$K_{ij} = K(\boldsymbol{x_i}, \boldsymbol{x_j}) = \langle \Phi(\boldsymbol{x_i}), \Phi(\boldsymbol{x_j}) \rangle \tag{2}$$

where $\boldsymbol{x_i}$ and $\boldsymbol{x_j}$ are the samples of training data in input space.

Step 2: Centralization of the kernel matrix $\boldsymbol{K}$. Under the assumption of $\sum_{i=1}^{m}\Phi(\boldsymbol{x_i}) = 0$, the mapped data should be centralized by using $\overline{\Phi}(\boldsymbol{x_i}) = \Phi(\boldsymbol{x_i}) - (1/m)\sum_{k=1}^{m}\Phi(\boldsymbol{x_k})$. Consequently, the centralized kernel matrix $\boldsymbol{K}_{center}$ can be constructed recurring to $\boldsymbol{K}$ as given by

$$\boldsymbol{K}_{center} = \boldsymbol{K} - 1_{m\times m}\boldsymbol{K} - \boldsymbol{K}1_{m\times m} + 1_{m\times m}\boldsymbol{K}1_{m\times m} \tag{3}$$

where $1_{m\times m}$ represents a $m \times m$ matrix with the elements $1/m$.

Step 3: Carry out the eigenvalue decomposition with respect to the centralized kernel matrix $\boldsymbol{K}_{center}$ by solving:

$$\boldsymbol{K}_{center}\boldsymbol{\alpha} = m\lambda\boldsymbol{\alpha} \tag{4}$$

where $\boldsymbol{\alpha} = (\boldsymbol{\alpha_1}, \boldsymbol{\alpha_2}, \cdots, \boldsymbol{\alpha_m})^T$, $\lambda \neq 0$, $\lambda^* = m\lambda$ are the eigenvalues, $\boldsymbol{\alpha}$ are the eigenvectors.

A brief introduction of the derivation for Eq. (4) is as follows. The covariance matrix $\overline{C}$ of the centralized data $\overline{\Phi}(\boldsymbol{x_i})$ can be constructed as given by

$$\overline{C} = \frac{1}{m}\sum_{i=1}^{m}\overline{\Phi}(\boldsymbol{x_i})\overline{\Phi}^T(\boldsymbol{x_i}) \tag{5}$$

The eigenvalue decomposition with respect to $\overline{C}$ can be given by $\lambda V = \overline{C}V$, λ represents the eigenvalues, V represents the corresponding eigenvectors. Note that the eigenvectors V can be linearly expressed recurring to $\overline{\Phi}(\boldsymbol{x_i})$ as given by

$$V = \sum_{i=1}^{m}\boldsymbol{\alpha_i}\overline{\Phi}(\boldsymbol{x_i}) \tag{6}$$

The inner product of $\overline{\Phi}(\boldsymbol{x_k})$ and $\lambda V = \overline{C}V$ is given by

$$\lambda\langle\overline{\Phi}(\boldsymbol{x_k}), V\rangle = \langle\overline{\Phi}(\boldsymbol{x_k}), \overline{C}V\rangle, k = 1, 2, \cdots, m \tag{7}$$

Eq. (4) is then obtained by inducing Eq. (5) and Eq. (6) into Eq. (7).

Step 4: Standardization of the eigenvectors $\boldsymbol{\alpha}$. The eigenvectors $\boldsymbol{\alpha}$ should be standardized since the eigenvectors V belong to unit vectors, i.e. $V, V = 1$. The standardized eigenvectors $\boldsymbol{\alpha}'$ should satisfy the condition:

$$\langle V, V\rangle = \langle\boldsymbol{\alpha}', \boldsymbol{K}_{center}\boldsymbol{\alpha}'\rangle = \lambda^*\boldsymbol{\alpha}', \boldsymbol{\alpha}' = 1 \tag{8}$$

Step 5: Principal components for the training data. In KPCA, principal components of the vectors $\boldsymbol{x_i}$ can be obtained by carrying out the projections of $\overline{\Phi}(\boldsymbol{x_i})$ onto the eigenvectors V as given by

$$\langle V, \overline{\Phi}(\boldsymbol{x_i})\rangle = \sum_{k=1}^{m}\boldsymbol{\alpha}'_{\boldsymbol{k}}\langle\overline{\Phi}(\boldsymbol{x_i}), \overline{\Phi}(\boldsymbol{x_i})\rangle = \boldsymbol{K}_{center}\cdot\boldsymbol{\alpha}' \tag{9}$$

Step 6: Principal components for the test data. The case for test data $\boldsymbol{x_t}$, a centralized kernel matrix $\boldsymbol{K}^C_{test}$ need to be reconstructed recurring to $K(\boldsymbol{x_t}, \boldsymbol{x_i})$ and $K(\boldsymbol{x_i}, \boldsymbol{x_j})$ as given by

$$\boldsymbol{K}^C_{test} = \boldsymbol{K}_{test} - \boldsymbol{K}_{test}1_{m\times m} - 1_{t\times m}\boldsymbol{K} + 1_{t\times m}\boldsymbol{K}1_{m\times m} \tag{10}$$

where $1_{m\times m}$ represents a $m \times m$ matrix with the elements $1/m$, $1_{t\times m}$ represents a $t \times m$ matrix with the elements $1/m$. Principal components of the vectors $\boldsymbol{x_t}$ is given by

$$\langle V, \overline{\Phi}(\boldsymbol{x_t})\rangle = \boldsymbol{K}^{C}_{test} \cdot \boldsymbol{\alpha}' \tag{11}$$

In general, the reduction ratio of eigenvectors $\boldsymbol{\alpha}'$ directly affects the accuracy of prediction model (such as SVM and ANN) due to the existence of the noises and redundancy. In this study, the cumulative contribution rate λ of principal components is utilized to determine the number of principal components as given by Eq. (12). The eigenvectors corresponding to the first k largest eigenvalues are retained. The $m \times k$ transform matrix constructed by the first k eigenvectors is utilized as the projection direction of $\boldsymbol{K}_{center}$ and $\boldsymbol{K}^{C}_{test}$ to extract the optimal principal components.

$$\sum_{i=1}^{k}\lambda_i^* \Big/ \sum_{i=1}^{m}\lambda_i^* \geqslant \lambda \tag{12}$$

where λ_i^* is the i th eigenvalue of the centralized kernel matrix $\boldsymbol{K}_{center}$, $\lambda_1^* \geqslant \lambda_2^* \geqslant \cdots \geqslant \lambda_m^*$.

It is well known that the performance of KPCA is subject to the selection of kernel function. In this paper, three kernels are investigated, i.e., linear function, radial basis function (RBF) and the integrated radial basis function (IRBF) [27] which are given by

$$K_{Linear}(\boldsymbol{x},\boldsymbol{y}) = \boldsymbol{x}\cdot\boldsymbol{y} \tag{13}$$

$$K_{RBF}(\boldsymbol{x},\boldsymbol{y}) = exp\left(-\frac{||\boldsymbol{x}-\boldsymbol{y}||^2}{2\sigma_0^2}\right) \tag{14}$$

$$K_{IRBF}(\boldsymbol{x},\boldsymbol{y}) = \int_0^{\sqrt{2}\sigma_0} exp\left(-\frac{||\boldsymbol{x}-\boldsymbol{y}||^2}{\sigma^2}\right)d\sigma \tag{15}$$

where σ_0 is the kernel parameter in RBF and IRBF, $\sigma_0 > 0$. Note that when the linear kernel function is utilized, the KPCA technique is equivalent to the PCA technique. The above three kernel-based methods are referred to as PCA, KPCA_RBF and KPCA_IRBF, respectively. PCA and KPCA_RBF belong to the conventional linear and nonlinear dimension-reduction techniques, respectively. However, the proposed KPCA_IRBF is a kind of new nonlinear dimension-increment technique. Besides, the kernel parameter of KPCA_IRBF has a much larger selectable region than that of the conventional KPCA_RBF technique.

2.3. Gaussian process regression

A Gaussian distribution is a probability distribution that describes the random variables such as scalars and vectors. It can be fully specified by a mean and a covariance: $x \sim \mathcal{N}(\mu, \sigma^2)$. On the other hand, a Gaussian process can be considered as a generalization of the Gaussian probability distribution and pours over functions. From the view of function space, a Gaussian process is a collection of random variables, any finite number of which have a joint Gaussian distribution [16].

2.3.1. The fundamentals of GPR

Assuming that $\mathcal{D} = \{(\boldsymbol{x_i}, y_i)|i = 1, 2, \cdots, N\}$ represents the training dataset of the Gaussian model. The feature vectors $\boldsymbol{x_i} \in \boldsymbol{R}^n$ consist of the extracted features or the fused features and the corresponding machining parameters. The observed target values y_i represent the tool flank wear width measured in machining process. $X = \{\boldsymbol{x_i}\}_{i=1}^N$ represents the input matrix of training dataset, $\boldsymbol{y} = \{y_i\}_{i=1}^N$ represents the output vector. A Gaussian process $f(\boldsymbol{x})$ can be fully specified by its mean function $m(\boldsymbol{x})$ and covariance function $k(\boldsymbol{x},\boldsymbol{x}')$. The Gaussian process [16] is denoted as

$$f(\boldsymbol{x}) \sim \mathcal{GP}(m(\boldsymbol{x}), k(\boldsymbol{x},\boldsymbol{x}')) \tag{16}$$

where

$$m(\boldsymbol{x}) = \mathbb{E}[f(\boldsymbol{x})] \tag{17}$$

$$k(\boldsymbol{x},\boldsymbol{x}') = \mathbb{E}[(f(\boldsymbol{x}) - m(\boldsymbol{x}))(f(\boldsymbol{x}') - m(\boldsymbol{x}'))] \tag{18}$$

The mean function $m(\boldsymbol{x})$ represents the expected value of the function $f(\boldsymbol{x})$ at the input point $\boldsymbol{x}$. The covariance function $k(\boldsymbol{x},\boldsymbol{x}')$ can be considered as a measure of the confidence level for $m(\boldsymbol{x})$. In general, the mean function is set to be zero for notational simplicity.

In practical application, the form of mean function $m(\boldsymbol{x})$ and covariance function $k(\boldsymbol{x},\boldsymbol{x}')$ need to be specified so as to define an individual Gaussian process. In this study, the affine mean function and the squared-exponential (SE) covariance function are utilized as given by [30]

$$m(\boldsymbol{x}) = \boldsymbol{a}\boldsymbol{x} + b \tag{19}$$

$$k_{\mathrm{SEiso}}(\boldsymbol{x},\boldsymbol{x}') = cov(f(\boldsymbol{x}),f(\boldsymbol{x}')) = \sigma_f^2 exp\left(-\frac{||\boldsymbol{x}-\boldsymbol{x}'||^2}{2l^2}\right) \tag{20}$$

where l is the characteristic length-scale, σ_f is the standard deviation of the signal. The 'kernel tricks' is utilized in the covariance function, i.e., the covariance of the outputs is expressed by a function of the inputs. It can be seen that the SE covariance function $k_{\mathrm{SEiso}}(\boldsymbol{x},\boldsymbol{x}')$ reaches the maximum σ_f^2 when the inputs satisfy $\boldsymbol{x} \approx \boldsymbol{x}'$ and decreases with the increase of the distance between $\boldsymbol{x}$ and $\boldsymbol{x}'$. Besides, the variation of the SE covariance is also subject to the length-scale l. Thus, the parameter selection of the SE covariance function directly affects the performance of the Gaussian process.

The function values $f(\boldsymbol{x})$ are not accessible in most applications. Actually, only the noisy observations are available, which can be given by

$$\boldsymbol{y} = f(\boldsymbol{x}) + \varepsilon \tag{21}$$

where ε is the additive white noise and assumed to be the independent and identically distributed Gaussian noise: $\varepsilon \sim \mathcal{N}(0,\sigma_n^2)$. σ_n is the standard deviation of the noise. Any finite number of the observed values can also form an individual Gaussian process as given by

$$\boldsymbol{y} \sim \mathcal{GP}(m(\boldsymbol{x}), k(\boldsymbol{x}_i,\boldsymbol{x}_j) + \sigma_n^2\delta_{ij}) \tag{22}$$

where δ_{ij} is the Kronecker delta function. $\delta_{ij} = 1$ holds true if and only if $i = j$ and $\delta_{ij} = 0$ otherwise.

The objective in this work is to predict the function value $\bar{f}_*$ and its variance $cov(f_*)$ given the new test point $\boldsymbol{x}_*$. X_* represents the input matrix of test dataset. N_* represents the size of test dataset. According to the definition of Gaussian process, the observed values and the function values at new test points follow a joint Gaussian prior distribution which can be denoted as

$$\begin{bmatrix} y \\ f_* \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} m(X) \\ m(X_*) \end{bmatrix}, \begin{bmatrix} K(X,X)+\sigma_n^2 I & K(X,X_*) \\ K(X_*,X) & K(X_*,X_*) \end{bmatrix}\right) \tag{23}$$

where $K(X,X)$ is the covariance matrix of training dataset. $K(X_*,X_*)$ is the covariance matrix of test dataset. $K(X,X_*)$ represents the $N \times N_*$ covariance matrix obtained from the training and test dataset, $K(X_*,X) = K(X,X_*)^T$. The posterior distribution over f_* can be derived by imposing restrictions on the joint prior distribution given the training dataset and test points, which is given by

$$f_*|X,\boldsymbol{y},\boldsymbol{x}_* \sim \mathcal{N}(\bar{f}_*, cov(f_*)) \tag{24}$$

where

$$\bar{f}_* = \mathbb{E}[f_*|X,\boldsymbol{y},\boldsymbol{x}_*] = m(X_*) + K(X_*,X)[K(X,X)+\sigma_n^2 I]^{-1}(\boldsymbol{y}-m(X)) \tag{25}$$

$$cov(f_*) = K(X_*,X_*) - K(X_*,X)[K(X,X)+\sigma_n^2 I]^{-1}K(X,X_*) \tag{26}$$

Then, the posterior distribution can be utilized for the prediction of new test input points. $\bar{f}_*$ is the predicted output of the GPR model for test point. Confidence interval (CI) of the predicted output value can be determined by the variance $cov(f_*)$. For instance, the 95% CI can be given by $[\bar{f}_* - 2\times\sqrt{cov(f_*)}, \bar{f}_* + 2\times\sqrt{cov(f_*)}]$. Therefore, the GPR model not only gives the predicted values but also provides the confidence level of the predicted results.

It can be seen that the GPR model is a non-parametric model since the predicted outputs depend only on the inputs and the observed values $\boldsymbol{y}$. In this way, the parameters $\Theta = \{\boldsymbol{a}, b, l, \sigma_f, \sigma_n\}$ are referred to as the hyper-parameters of the GPR model.

2.3.2. Hyper-parameter estimation

When the form of mean function $m(\boldsymbol{x})$ and covariance function $k(\boldsymbol{x},\boldsymbol{x}')$ in the GPR model are determined, the corresponding hyper-parameters are unknown and need to be inferred from the training dataset. Maximum likelihood estimation is the commonly used method for the parameter estimation of the GPR model. The marginal likelihood $P(\boldsymbol{y}|X)$ obtained by utilizing Bayes' rule is given by

$$P(\boldsymbol{y}|X) = \int P(\boldsymbol{y}|f,X)P(f|X)df \tag{27}$$

The optimal hyper-parameters Θ' can be obtained by maximizing the log-marginal likelihood as given by

$$logP(\boldsymbol{y}|X,\Theta) = -\frac{1}{2}(\boldsymbol{y}-\boldsymbol{m})^T[\boldsymbol{K}+\sigma_n^2 I]^{-1}(\boldsymbol{y}-\boldsymbol{m}) - \frac{1}{2}log|\boldsymbol{K}+\sigma_n^2 I| - \frac{N}{2}log2\pi \tag{28}$$

where $\boldsymbol{m} = m(\boldsymbol{x})$, $\boldsymbol{K} = K(X,X)$. The optimal hyper-parameters $\Theta' = arg\max_{\Theta} logP(\boldsymbol{y}|X,\Theta)$ can be solved by utilizing the conjugate gradient algorithm after parameter initialization [16]. Note that the conjugate gradient algorithm is susceptible to the initial value of hyper-parameters.

3. Experimental results and analysis

This work aims at real-timely and accurately monitoring the tool wear in machining process by utilizing the proposed KPCA_IRBF technique and the GPR model. The GPR-based tool wear predictive model is constructed by fitting the target value and the fused features as illustrated in Fig. 1. In general, the prediction accuracy of the GPR model is almost unaffected by Gaussian noises since they can be modeled quantitatively as given by Eq. (22). Therefore, GPR performs better than other data-driven methods, such as artificial neural networks (ANN) and support vector machines (SVM).

3.1. Performance evaluation for prediction accuracy

In this work, v-SVR [31], LS-SVM [9], BPNN [7] and Elman [32] are also utilized to realize tool wear prediction. The feature vectors consist of the 48 extracted features (without feature fusion) and the corresponding machining parameters are directly fed into the five tool wear predictive models (GPR, v-SVR, LS-SVM, BPNN and Elman) so as to reveal the advantage of GPR, i.e., the existence of noises has little impact on the prediction accuracy of GPR.

The hyper-parameters $\Theta = \{\boldsymbol{a}, b, l, \sigma_f, \sigma_n\}$ of the GPR model are unknown and need to be inferred by utilizing the conjugate gradient algorithm given the training dataset. The parameters of the mean function $m(\boldsymbol{x})$ as given by Eq. (19) are initialized as $\boldsymbol{a} = 0$ and $b = 0$. The length-scale parameter of the covariance function $k_{\text{SEiso}}(\boldsymbol{x}, \boldsymbol{x}')$ as given by Eq. (20) is initialized as $l = 1$. The interval ranges of the standard deviation for the signal and the noise are determined by trial and error: σ_f and $\sigma_n \in [0.01, 0.1] \times [0.0001, 0.001]$. The appropriate initial parameters of σ_f and σ_n are determined in terms of the prediction accuracy of GPR by utilizing the grid-search method and k-fold cross-validation as shown in Fig. 3. In k-fold cross-validation, the training dataset are equally divided into k subsets in sorted order. Then, k sets of calculations containing training and test processes need to be carried out. In the i th $(i = 1, \cdots, k)$ calculation, the i th subset is used as the test set and the other $k-1$ subsets are used as the training set. The test set is used to calculate the prediction accuracy of the GPR model trained by the training set. The mean square error (MSE) between the target value y_i and the predicted output $\overline{f}_*(\boldsymbol{x}_i)$ of the GPR model is adopted as the performance evaluation index. The average of the k results from k -fold cross-validation is adopted as the fitness value of the GPR model and serves to select the best grid parameters (σ_f and σ_n). In this work, the parameter k is set to 3 so as to avoid over-fitting and reduce the calculation time effectively. $\sigma_f = 0.08$ and $\sigma_n = 8 \times 10^{-4}$ are selected out from the two-dimension grid space (formed by σ_f and σ_n) by using 3-fold cross-validation. The flowchart of the GPR model for tool wear prediction is shown in Fig. 4.

In the v-SVR model, three hyper parameters (the bound parameter v, the regularization constant C and the RBF kernel parameter γ) make up a three-dimension grid space from which the best parameters are selected and determined $(v = 0.46, C = 3.1623, \gamma = 0.01)$ in terms of prediction accuracy by utilizing the grid-search method and 3-fold cross-validation. In the LS-SVM model, two hyper parameters (the penalty parameter γ and the RBF kernel parameter σ^2) make up a two-dimension grid space from which the best parameters are also determined $(\gamma = 199.5262, \sigma^2 = 79.4328)$ by utilizing the grid-search method and 3-fold cross-validation. In the BPNN and Elman model, the number of neurons assigned in the hidden layer is set to 102 which is twice the dimensionality of feature vectors. The tan-sigmoid and log-sigmoid activation functions are selected for the hidden and output layer, respectively. In the training process, the weight and bias values are updated according to the scaled conjugate gradient method. The maximum number of epochs and the performance goal are set to 10^4 and 10^{-5}, respectively.

Predicted results of the five tool wear predictive models (GPR, v-SVR, LS-SVM, BPNN and Elman) are shown in Fig. 5. The bold red[1] line in each subfigure represents the measured tool flank wear width. The error represents the absolute value of the difference between the predicted and measured tool wear. It can be seen that GPR, v-SVR and LS-SVM can follow the tracks of actual tool wear effectively. BPNN and Elman perform not so well. To compare the performance of the five tool wear predictive models, four kinds of evaluation index are investigated as given by

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|y_i - \bar{f}_i| \tag{29}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \bar{f}_i)^2} \tag{30}$$

$$MAPE = \frac{1}{N}\sum_{i=1}^{N}\frac{|y_i - \bar{f}_i|}{y_i} \times 100\% \tag{31}$$

$$PCC = \frac{\sum_{i=1}^{N}(y_i - \bar{y})(\bar{f}_i - \bar{\bar{f}})}{\sqrt{\sum_{i=1}^{N}(y_i - \bar{y})^2(\bar{f}_i - \bar{\bar{f}})^2}} \tag{32}$$

[1] For interpretation of color in Figs. 5, 8, 10, 13 and 14, the reader is referred to the web version of this article.

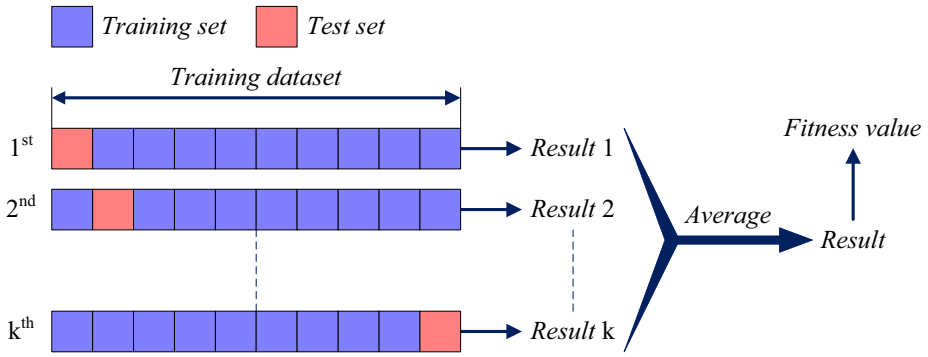

Fig. 3. Flowchart of *k*-fold cross-validation.

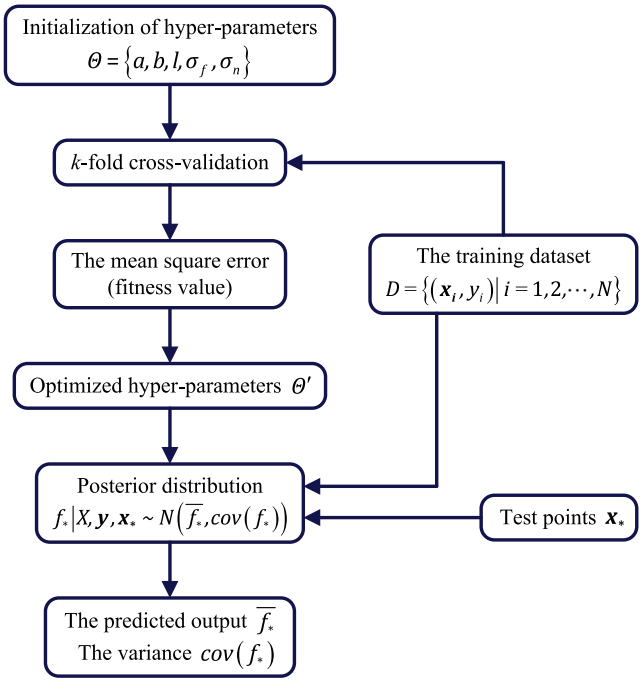

Fig. 4. Parameter optimization of the GPR model for tool wear prediction.

where y_i is the tool flank wear width measured by the Video Measuring System (VMS), $\bar{f}_i$ is the predicted value of one tool wear predictive model at test point $\boldsymbol{x_i}$ which is the corresponding feature vector of y_i. MAE is the mean absolute error, RMSE is the root mean square error, MAPE is the mean absolute percentage error, PCC is the Pearson correlation coefficient. Note that a larger PCC means better performance of the tool wear predictive model. Performance comparison of the five tool wear predictive models under different evaluation indexes is shown in Fig. 6. The number above the column in each subfigure represents the value of evaluation index. It is found that GPR has the minimum MAE/RMSE/MAPE value and the maximum PCC value among the five tool wear predictive models. Therefore, it can be concluded that GPR still can maintain higher prediction accuracy even if there is much noises in the signal features.

3.2. Performance evaluation for confidence interval

The confidence interval (CI) of the predicted results can also be provided by the GPR model. In this work, analysis of the 95% CI of the predicted results form GPR is also carried out. Upper bound and lower bound of the 95% CI are given by $\overline{f_*} + 2 \times \sqrt{cov(f_*)}$ and $\overline{f_*} - 2 \times \sqrt{cov(f_*)}$, respectively. Predicted results of the GPR-based tool wear predictive model without feature fusion are shown in Fig. 7. It can be seen that the 95% CI of the predicted results has a large fluctuation which is not conducive for monitoring the tool wear accurately. Although the existence of the noises in machining process has little

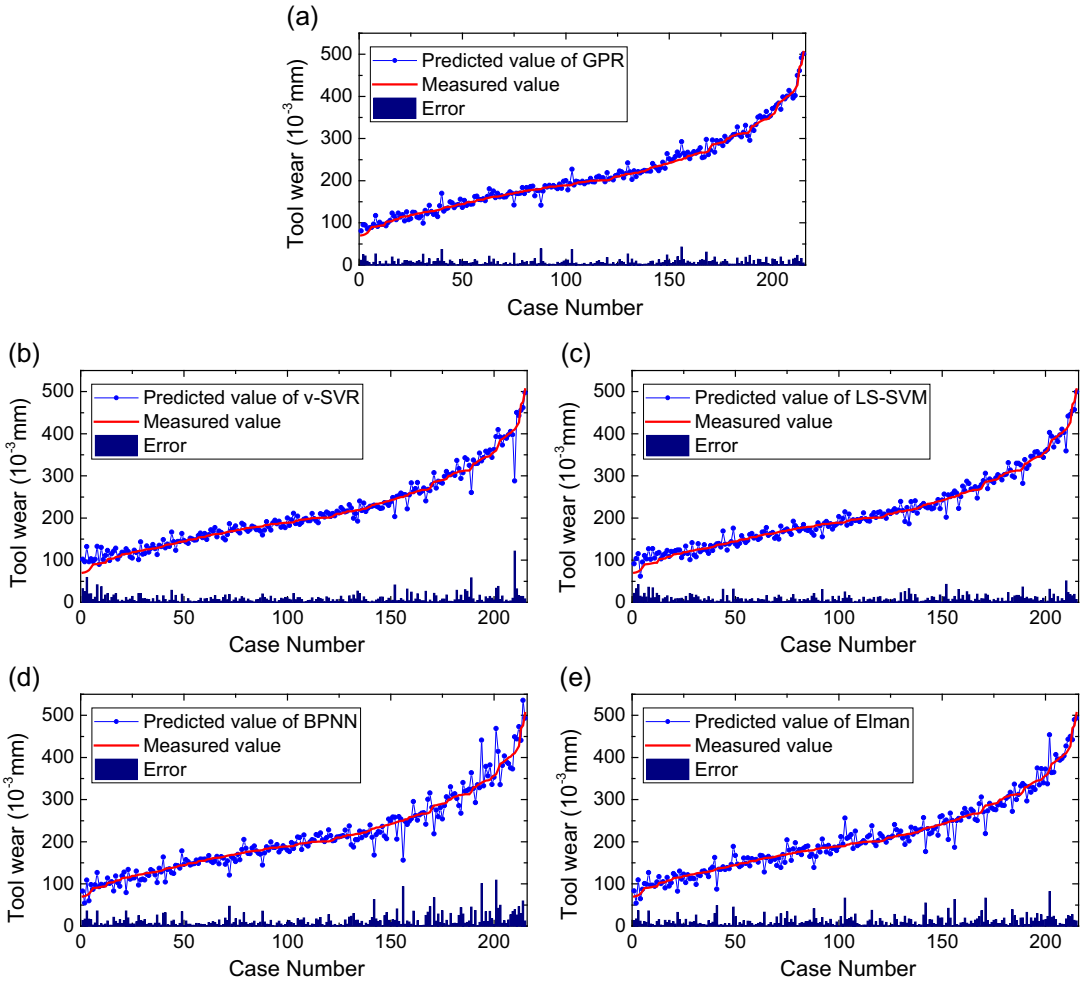

Fig. 5. Predicted results of the five tool wear predictive models (GPR, v-SVR, LS-SVM, BPNN and Elman).

impact on the prediction accuracy of GPR, the confidence interval of the GPR model is greatly affected. In this section, feature fusion techniques are utilized to remove and weaken the negative effects of noises in order to ameliorate the status of the confidence interval. Two evaluation indexes are proposed and investigated so as to quantitatively analyze the reliability and stability of the 95% CI, which are given by

$$CI\_width = \frac{1}{N}\sum_{i=1}^{N} 4 \times \sqrt{cov(f_i)} \tag{33}$$

$$CI\_var = \frac{1}{N}\sum_{i=1}^{N} \left(4 \times \sqrt{cov(f_i)} - CI\_width\right)^2 \tag{34}$$

where $4 \times \sqrt{cov(f_i)}$ is the width of the 95% CI at test point $\boldsymbol{x_i}$. CI_width and CI_var are the average width and the variance of the 95% CI of test dataset, respectively. The less the CI_width value, the more accurate tool wear status can be monitored. The less the CI_var value, the more reliable and stable monitoring process for tool wear can be realized.

Next, three feature fusion techniques (PCA, KPCA_RBF and KPCA_IRBF) are investigated with the aim to ameliorate the status of the confidence interval. The fused features of the three feature fusion techniques and the corresponding machining parameters make up the feature vectors which are fed into the GPR model, respectively. Performance evaluation for the three techniques is carried out under six kinds of evaluation index (MAE, RMSE, MAPE, PCC, CI_width and CI_var).

Firstly, the PCA technique is utilized to fuse the extracted features. Performance evaluation of the GPR-based tool wear predictive model using the fused features of PCA under different cumulative contribution rates (λ) is listed in Table 3. When $\lambda = 100\%$ is selected, the fused feature are the same as the extracted features and the predicted results of GPR are equivalent to that as shown in Fig. 6. The predicted results of GPR are also acceptable when $\lambda = 98\%$ and 99% are selected. However, the

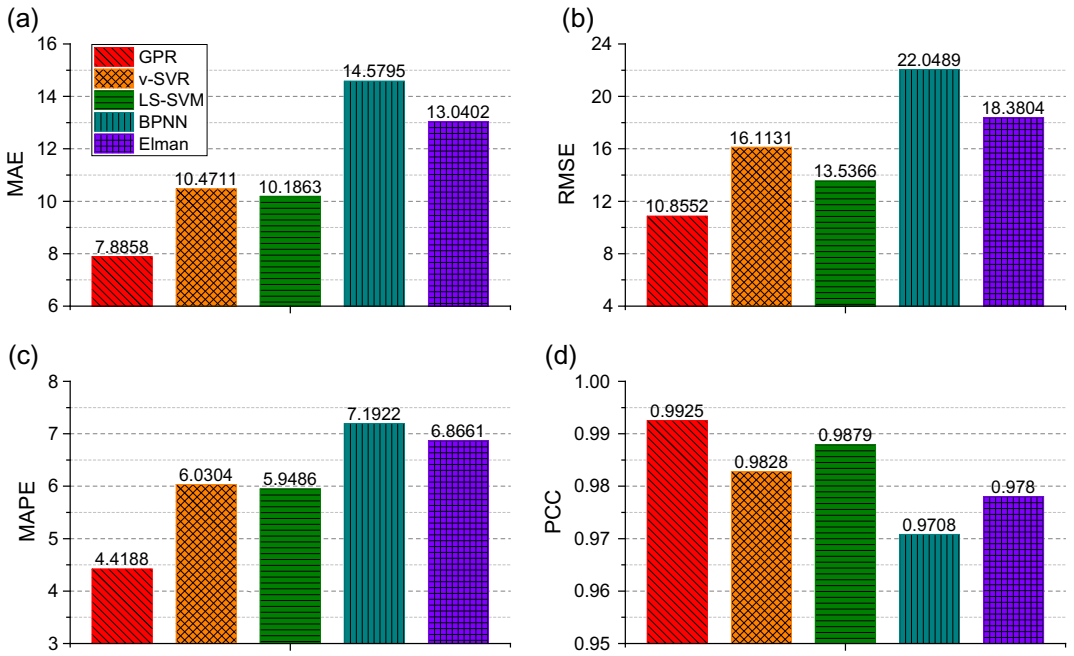

Fig. 6. Performance comparison of the five tool wear predictive models under four kinds of evaluation index.

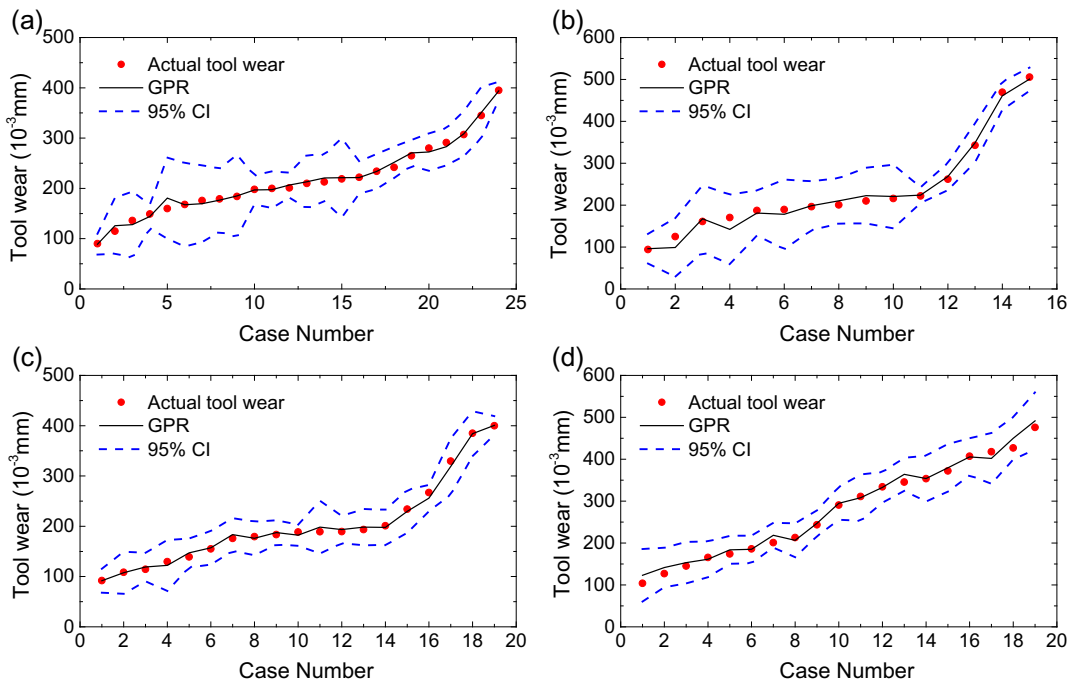

Fig. 7. Predicted results of the GPR-based tool wear predictive model without feature fusion: (a) Replication 2 of Test No. 2; (b) Replication 1 of Test No. 4; (c) Replication 2 of Test No. 4; (d) Test No. 6.

reliability and stability of the 95% CI has not been ameliorated since both the value of CI_width and CI_var have a slight increase. What's more, the prediction accuracy of the GPR model shows a deteriorating trend with the further decrease of λ. Therefore, it can be concluded that the PCA technique can't meet the requirements of ameliorating the status of the confidence interval of the GPR model, under the premise of guaranteeing prediction accuracy.

Table 3
Performance evaluation of the GPR-based tool wear predictive model using the fused features of PCA.

λ (%)	σ_f	σ_n	MAE	RMSE	MAPE	PCC	CI_width	CI_var
95	0.1	1×10^{-3}	9.9306	14.8011	5.7409	0.9854	81.4356	1.72×10^{-3}
96	0.03	3×10^{-4}	9.6339	15.4904	5.3380	0.9840	74.1900	1.42×10^{-3}
97	0.03	1×10^{-3}	9.2760	14.0708	5.0553	0.9868	75.8266	1.09×10^{-3}
98	**0.07**	$\mathbf{6 \times 10^{-4}}$	**7.7521**	**11.2326**	**4.2100**	**0.9917**	**73.7003**	$\mathbf{1.66 \times 10^{-3}}$
99	**0.1**	$\mathbf{8 \times 10^{-4}}$	**8.1125**	**11.8220**	**4.4928**	**0.9909**	**76.3210**	$\mathbf{2.06 \times 10^{-3}}$
100	0.08	8×10^{-4}	7.8858	10.8552	4.4188	0.9925	72.3792	1.36×10^{-3}

Note: The initial parameters of σ_f and σ_n are selected out by utilizing the grid-search method and 3-fold cross-validation.

Secondly, the KPCA techniques are utilized for feature fusion and the threshold of the cumulative contribution rate (λ) is set to 90% so as to remove the random noises and weaken its negative effects. The kernel parameters in a relatively large region $\sigma_0^2 \in [10^1, 10^8]$ are adopted in the two kernel-based methods (KPCA_RBF and KPCA_IRBF). Prediction accuracy of the GPR model may produce very small fluctuations since a few unknown and uncontrollable factors (such as redundance and some non-Gaussian noises) may be introduced during feature fusion. Selectable region of the kernel parameters σ_0^2 in KPCA_RBF and KPCA_IRBF is determined by the MAE value of the GPR model which deviates from the benchmark no more than 5%. The benchmark is the corresponding MAE value (7.8858) obtained from the GPR model without feature fusion. Performance evaluation of the GPR-based tool wear predictive model using the fused features of KPCA_RBF under different kernel parameters is listed in Table 4. It can be seen that the performance of KPCA_RBF largely depends on the selection of kernel parameter and become ineffective with the exponential growth of kernel parameter. Although the status of the confidence interval get some amelioration by using KPCA_RBF at the effective kernel parameters $(\sigma_0^2 = 40, 60, 80)$, the effect is not so satisfactory. Performance evaluation of the GPR-based tool wear predictive model using the fused features of KPCA_IRBF $(\lambda = 90\%)$ with kernel parameters under different orders of magnitude is shown in Fig. 8. The red dashed line in each subfigure represents the threshold which is obtained from the GPR model by using the extracted features. Although the kernel parameters of KPCA_IRBF in the selectable region do not perform so well in terms of the RMSE value, it takes on comparable performance and can be acceptable. The blue dashed line in Fig. 8(e) and (f) represents the maximum CI_width and CI_var value, respectively. It is found that KPCA_IRBF performs better than KPCA_RBF in terms of the CI_width and CI_var value. Most importantly, the proposed KPCA_IRBF technique can be carried out within a much larger kernel parameter region and causes no over-fitting which is very encouraging. A case study that utilizing the KPCA_IRBF technique at the kernel parameter $\sigma_0^2 = 5 \times 10^4$ to fuse the extracted features is carried out. Predicted results of the GPR-based tool wear predictive model using the fused features of KPCA_IRBF under $\sigma_0^2 = 5 \times 10^4$ and $\lambda = 90\%$ are shown in Fig. 9. Performance evaluation shows that, under the premise of guaranteeing the prediction accuracy of the GPR model, the CI_width and CI_var value of the 95% CI is reduced by 22.57% and 68.31%, respectively. In conclusion, the reliability and stability of the monitoring process of the GPR-based tool wear predictive model can be greatly ameliorated by utilization of the proposed KPCA_IRBF technique.

What's more, the proposed KPCA_IRBF is a new nonlinear dimension-increment technique in comparison with the conventional KPCA_RBF technique. Further amelioration of the 95% CI of the GPR model by taking full advantage of the

Table 4
Performance evaluation of the GPR-based tool wear predictive model using the fused features of KPCA_RBF $(\lambda = 90\%)$.

σ_0^2	σ_f	σ_n	MAE	RMSE	MAPE	PCC	CI_width	CI_var
10^1	0.01	4×10^{-4}	10.1961	16.6109	5.9683	0.9816	74.0813	6.36×10^{-4}
2×10^1	0.1	9×10^{-4}	8.6469	13.8278	4.7743	0.9873	69.8071	8.75×10^{-4}
$\mathbf{4 \times 10^1}$	**0.1**	$\mathbf{5 \times 10^{-4}}$	**8.2205**	**12.7165**	**4.4297**	**0.9892**	**67.1678**	$\mathbf{8.16 \times 10^{-4}}$
$\mathbf{6 \times 10^1}$	**0.05**	$\mathbf{1 \times 10^{-3}}$	**7.8211**	**11.0442**	**4.2685**	**0.9919**	**66.9930**	$\mathbf{8.15 \times 10^{-4}}$
$\mathbf{8 \times 10^1}$	**0.09**	$\mathbf{7 \times 10^{-4}}$	**8.2511**	**11.9749**	**4.6092**	**0.9905**	**71.2689**	$\mathbf{8.73 \times 10^{-4}}$
10^2	0.02	9×10^{-4}	8.5401	11.6790	4.8909	0.9909	78.0853	3.04×10^{-4}
2×10^2	0.04	9×10^{-4}	8.7100	11.6380	5.0639	0.9910	74.3964	3.74×10^{-4}
4×10^2	0.04	9×10^{-4}	9.2231	14.0170	4.9474	0.9869	74.9665	1.43×10^{-3}
6×10^2	0.09	5×10^{-4}	9.7432	14.9872	5.2905	0.9850	80.8646	1.26×10^{-3}
8×10^2	0.09	1×10^{-3}	9.7167	14.6977	5.2524	0.9857	85.8077	1.22×10^{-3}
10^3	0.03	4×10^{-4}	12.0374	20.3983	6.4710	0.9721	83.5665	9.51×10^{-4}
2×10^3	0.01	3×10^{-4}	15.3992	26.9250	8.1865	0.9601	97.4775	8.24×10^{-3}
4×10^3	0.07	8×10^{-4}	10.7688	14.4089	6.1777	0.9862	84.0934	4.05×10^{-4}
6×10^3	0.08	5×10^{-4}	11.7958	15.3665	6.7155	0.9843	84.4721	3.21×10^{-4}
8×10^3	0.04	4×10^{-4}	11.8030	15.3497	6.7294	0.9843	83.6633	3.54×10^{-4}
10^4	0.02	9×10^{-4}	13.1643	21.0855	7.1880	0.9704	108.7493	1.31×10^{-3}

Note: The initial parameters of σ_f and σ_n are selected out by utilizing the grid-search method and 3-fold cross-validation.

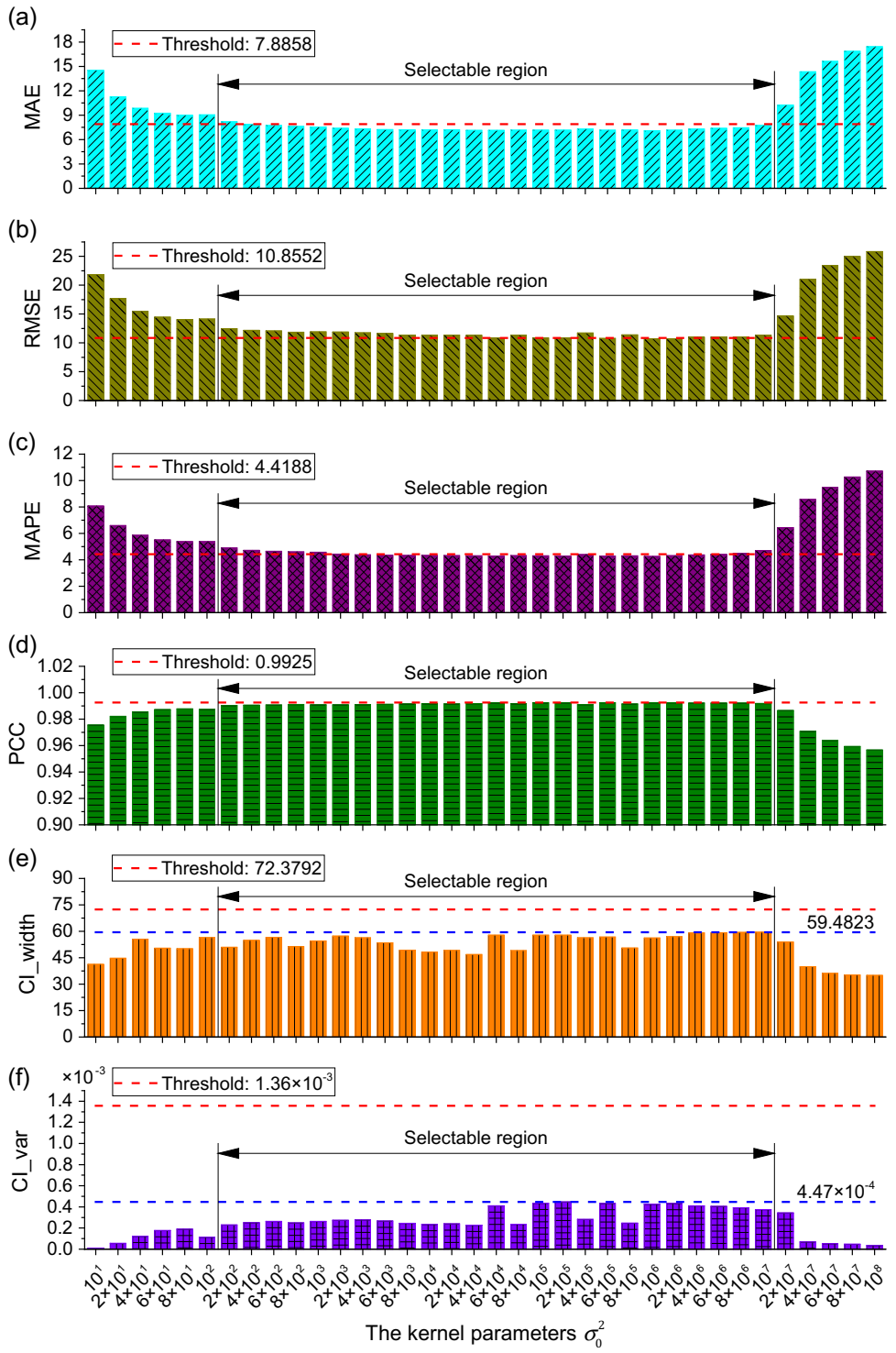

Fig. 8. Performance evaluation of the GPR-based tool wear predictive model using the fused features of KPCA_IRBF ($\lambda = 90\%$).

dimension-increment property of the KPCA_IRBF technique is introduced in the following section. Performance evaluation of the GPR-based tool wear predictive model using the fused features of KPCA_IRBF ($\sigma_0^2 = 5 \times 10^4$) under different cumulative contribution rates (λ) is shown in Fig. 10. The red dashed line in each subfigure represents the threshold which is the same as

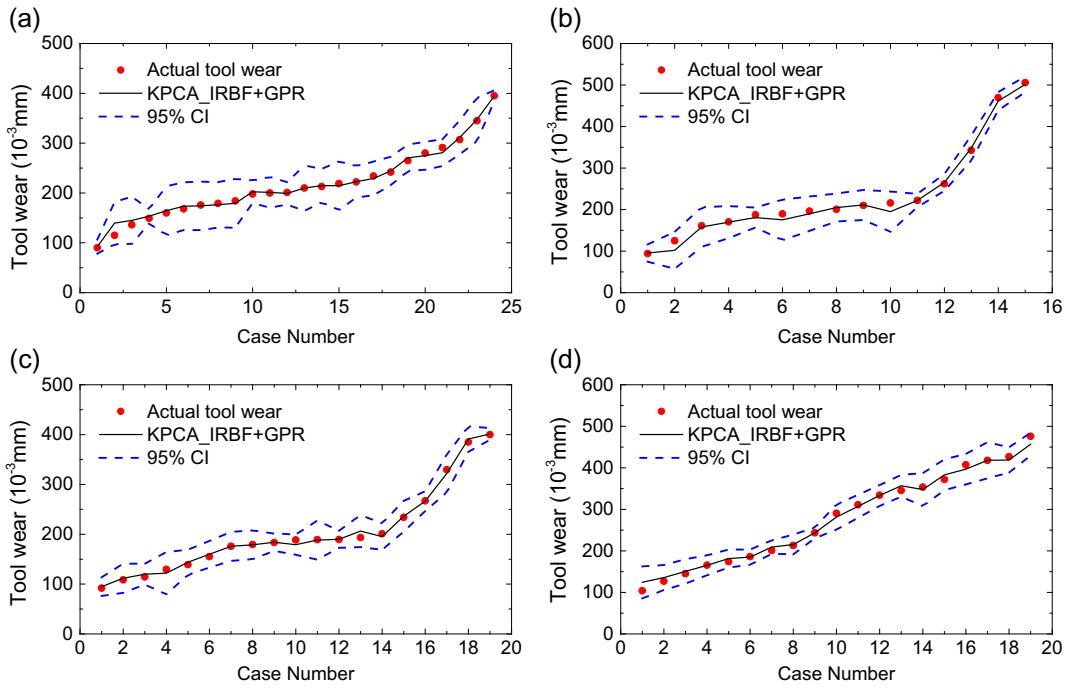

Fig. 9. Predicted results of the GPR-based tool wear predictive model using the fused features of KPCA_IRBF ($\sigma_0^2 = 5 \times 10^4$ and $\lambda = 90\%$): (a) Replication 2 of Test No. 2; (b) Replication 1 of Test No. 4; (c) Replication 2 of Test No. 4; (d) Test No. 6.

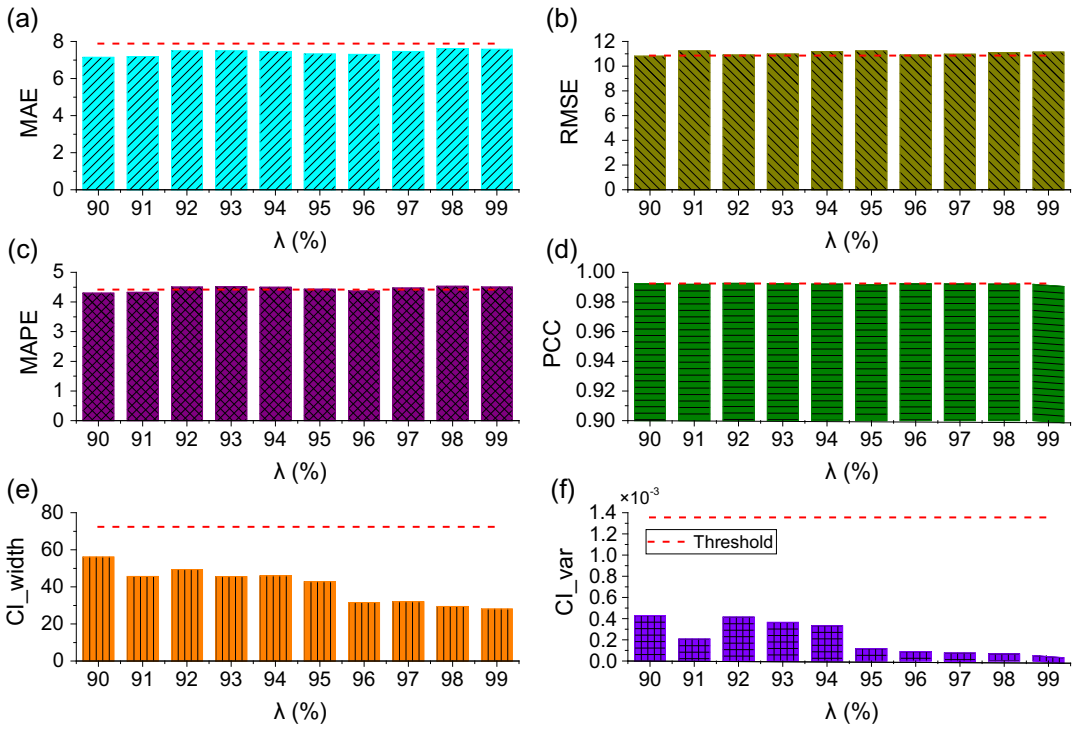

Fig. 10. Performance evaluation of the GPR-based tool wear predictive model using the fused features of KPCA_IRBF ($\sigma_0^2 = 5 \times 10^4$).

that in Fig. 8. It can be seen that the MAE/RMSE/MAPE/PCC value of the GPR model is stable with little fluctuation with the increase of λ. However, the CI_width/CI_var value shows a gradual downward trend with the increase of λ. In comparison with the corresponding threshold as shown in Fig. 10, the reduction (%) of the CI_width and CI_var value at each cumulative

contribution rate (λ) are listed in Table 5. It can be concluded that the dimension-increment property of KPCA_IRBF can help to further ameliorate the reliability and stability of the monitoring process of the GPR-based tool wear predictive model. Predicted results of the GPR-based tool wear predictive model using the fused features of KPCA_IRBF at the kernel parameter $\sigma_0^2 = 5 \times 10^4$ under $\lambda = 95\%$ and $\lambda = 98\%$ are shown in Figs. 11 and 12, respectively. It can be seen that the 95% CI of the GPR model has been greatly reduced and becomes more stable in comparison with Fig. 9. It is found that, under the premise of guaranteeing prediction accuracy, the increase of λ can help to further compress the confidence interval of the GPR model and make it more smoother when the kernel parameter of KPCA_IRBF is selected among the selectable region as shown in Fig. 8. Therefore, it can be concluded that the in-process flank wear width of tool inserts can be monitored more accurately by utilizing the combination of the proposed KPCA_IRBF technique and the GPR model.

3.3. Performance evaluation for time consumption

Intuitively, it may be found that SVM and GPR have similar prediction accuracy as shown in Fig. 5. To further highlight the novelties of the GPR model, a comprehensive analysis of the time consumption of GPR, v-SVR, LS-SVM, BPNN and Elman is also carried out in this work. The detailed time consumption of the five tool wear predictive models is listed in Table 6. CV_time is the time consumption of finding the optimal parameters by utilizing the grid-search method and 3-fold cross-validation. Train_time is the time consumption of model construction by using the determined model parameters and network structure. Test_time is the time consumption of testing time for test dataset.

The GPR model performs the best in terms of CV_time since the interval ranges of the model parameters in GPR is far less than that of v-SVR and LS-SVM. Grid-search is the most commonly used optimization method and easy to implement. However, the CV_time of v-SVR and LS-SVM will show exponential growth with the increase of grid density and even tends to infinity, which is impractical for industrial applications. In general, meta-heuristic algorithms are effective methods for parameter optimization in v-SVR and LS-SVM. As for BPNN and Elman, the determination of network structure and model parameters needs repeated debugging which is quite time consuming.

Table 5
Reduction (%) of the CI_width and CI_var value with the increase of λ.

λ (%)	90	91	92	93	94	95	96	97	98	99
CI_width	22.57	36.95	31.86	37.14	36.58	41.01	56.42	55.80	59.42	61.31
CI_var	68.31	84.65	69.42	73.15	75.42	91.36	93.51	94.42	95.05	96.13

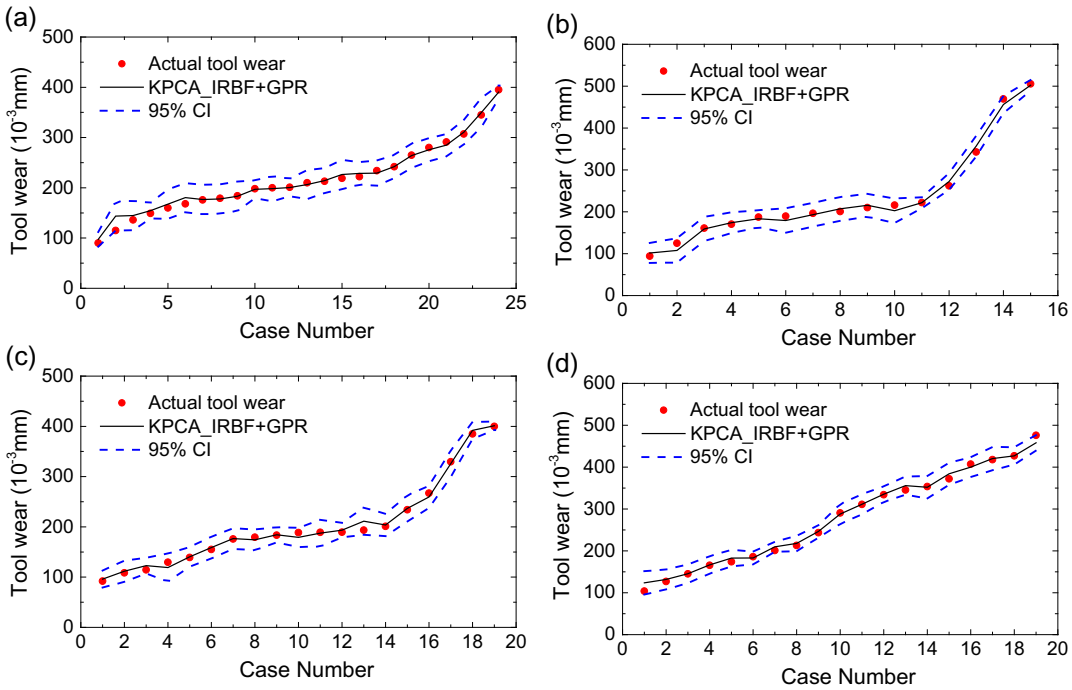

Fig. 11. Predicted results of the GPR-based tool wear predictive model using the fused features of KPCA_IRBF ($\sigma_0^2 = 5 \times 10^4$ and $\lambda = 95\%$): (a) Replication 2 of Test No. 2; (b) Replication 1 of Test No. 4; (c) Replication 2 of Test No. 4; (d) Test No. 6.

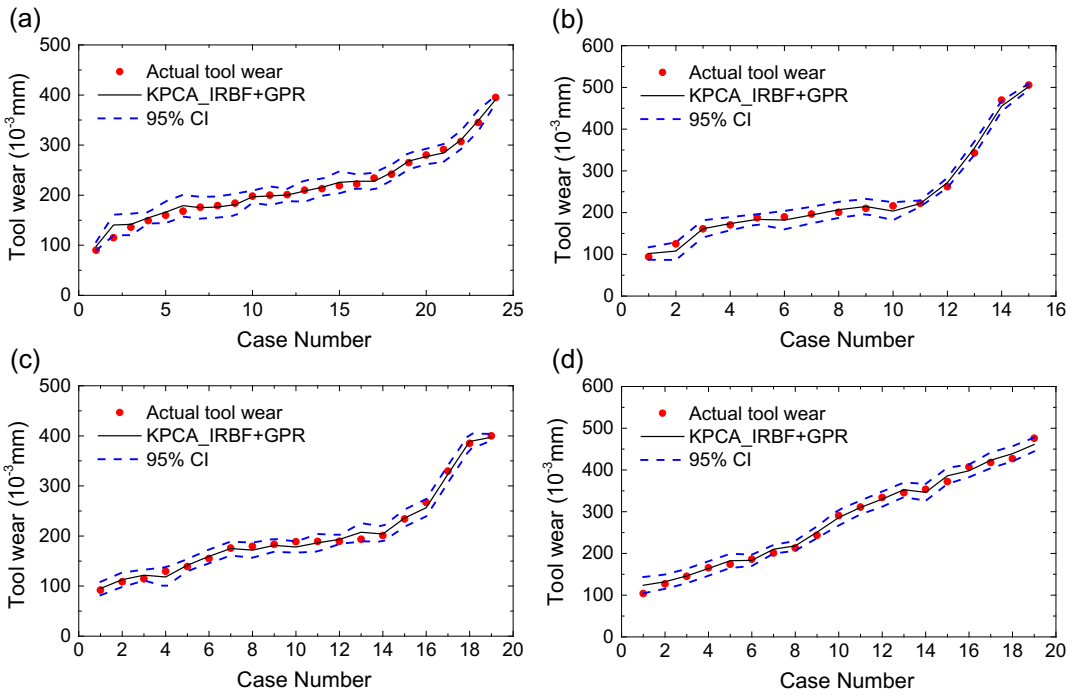

Fig. 12. Predicted results of the GPR-based tool wear predictive model using the fused features of KPCA_IRBF ($\sigma_0^2 = 5 \times 10^4$ and $\lambda = 98\%$): (a) Replication 2 of Test No. 2; (b) Replication 1 of Test No. 4; (c) Replication 2 of Test No. 4; (d) Test No. 6.

Table 6
Time consumption of the five tool wear predictive models (s).

Methods	CV_time	Train_time	Test_time
GPR	386.631	1.342	0.015
v-SVR	3655.557	0.048	0.015
LS-SVM	2597.621	0.078	0.032
BPNN	–	4.150	0.016
Elman	–	6.070	0.016

Besides the prediction accuracy, Test_time is also a key indicator that determines whether on-line monitoring can be realized by using the constructed tool wear predictive model. It can be seen that the Test_time of GPR is comparable to the other four models, which is less than 0.05 s. In summary, the GPR model has the advantages of higher prediction accuracy, less computational efforts (CV_time) for model construction and comparable testing time (Test_time), which make it especially suitable for on-line tool wear monitoring.

4. Future works

Analysis and amelioration of the confidence interval of the GPR model is the highlight in this work. The proposed KPCA_IRBF technique is quite satisfactory in ameliorating the status of the confidence interval of the GPR model and significantly outperforms the conventional KPCA_RBF technique. It provides a new angle and thought for the analysis and quantification of uncertainty, i.e., ameliorating the uncertainty estimations by recurring to dimension-increment techniques.

In the future research, performance comparison between KPCA_IRBF and other conventional dimension-reduction techniques [23,24] will be carried out. Besides, the extension and variation of these conventional dimension-reduction techniques will also be carried out so as to explore new methods that can ameliorate the uncertainty estimations effectively.

5. Conclusions

In this paper, a novel tool wear assessment technique based on the proposed KPCA_IRBF technique and the GPR model is developed for accurately monitoring the in-process flank wear width of tool inserts. The major work is summarized as follows:

(1) GPR performs better than ANN and SVM in terms of prediction accuracy since the Gaussian noises can be modeled quantitatively in the GPR model. Besides, the GPR model can also provide the confidence interval of the predicted results.
(2) As a new nonlinear dimension-increment technique, the proposed KPCA_IRBF outperforms the conventional KPCA_RBF technique in ameliorating the status of the confidence interval of the GPR model. Besides, the kernel parameter of KPCA_IRBF has a much larger selectable region than that of KPCA_RBF. Therefore, the determination of a suitable kernel parameter for KPCA_IRBF is much easier in comparison with KPCA_RBF.
(3) Less computational efforts for model construction and comparable testing time make the GPR model an attractive option for on-line tool wear monitoring.

The effectiveness of the presented tool wear assessment technique is validated through ten sets of cutting tests. The research findings in this work provide important guarantee for accurately monitoring the tool wear in machining process.

Acknowledgements

The authors are grateful to the financial sponsorship from 863 National High-Tech Research and Development Program of China (No. 2013AA041108).

Appendix A

A.1. The dimension-increment property of KPCA_IRBF

In this section, the dimension-increment property of the proposed KPCA_IRBF technique is illustrated with examples. Dimensionality comparison of the fused features of KPCA_IRBF and KPCA_RBF under different sizes of dataset is shown in Fig. 13. The different sizes of dataset are randomly selected from the training dataset as mentioned in Section 2.1. Note that the machining parameters need to be deleted from each dataset. The red dashed line in each subfigure represents the dimensionality of the extracted features. New findings are as follows.

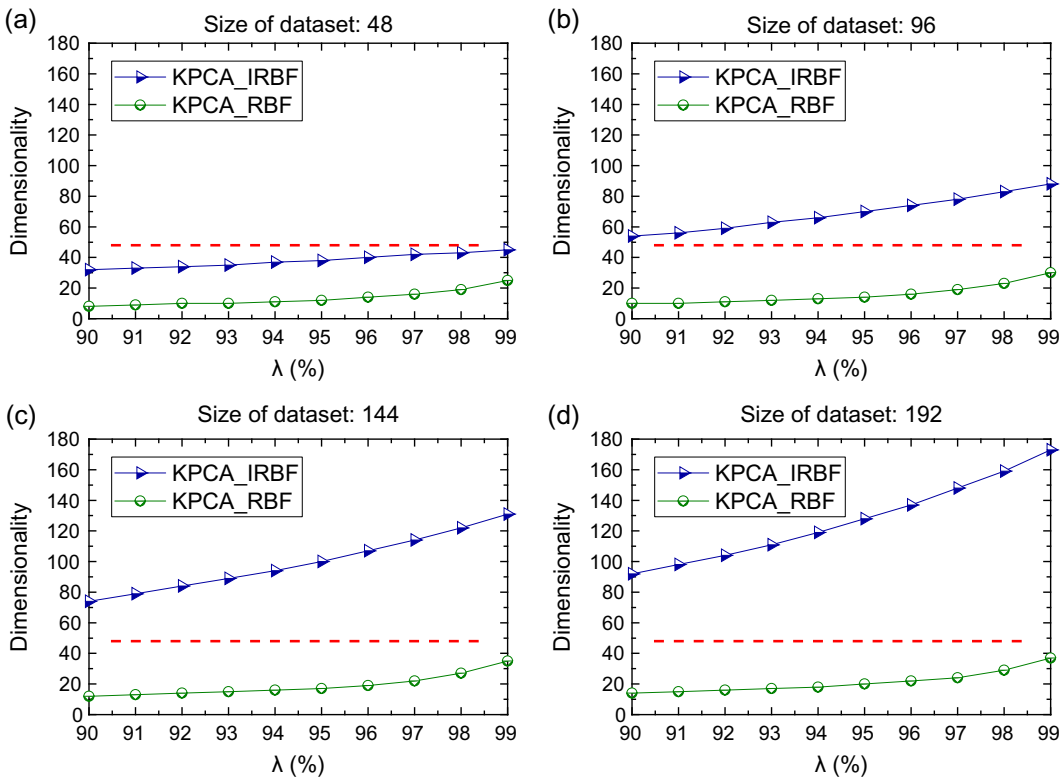

Fig. 13. Dimensionality comparison of the fused features of KPCA_IRBF and KPCA_RBF ($\sigma_0^2 = 10^2$).

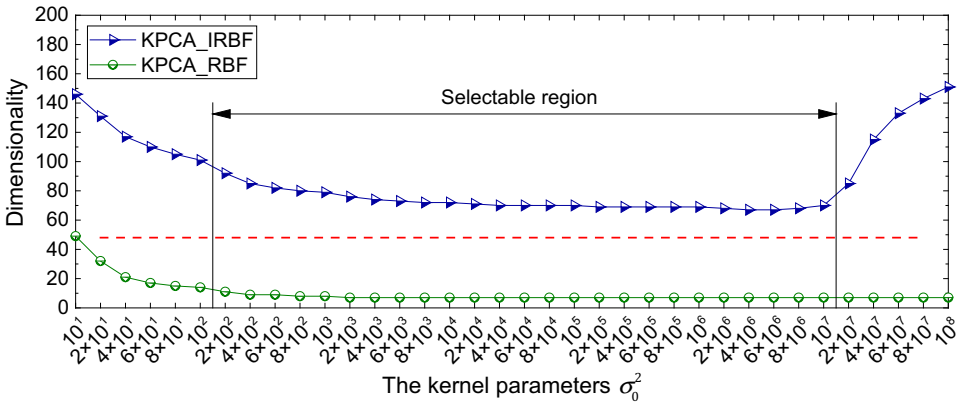

Fig. 14. Dimensionality comparison of the fused features of KPCA_IRBF and KPCA_RBF ($\lambda = 90\%$).

(1) The conventional KPCA_RBF technique realizes dimension-reduction which is not affected by the size of dataset.
(2) The proposed KPCA_IRBF technique realizes dimension-increment when the size of dataset outnumbers the dimensionality of the extracted features.
(3) The dimensionality of the fused features of KPCA_IRBF is larger than that of KPCA_RBF.

A.2. The selection of kernel parameter in KPCA_IRBF

Noting that lots of effective information associated with tool wear status will be lost when the dimensionality of the fused features is reduced too much. Besides, lots of redundant information may arise when the dimensionality of the fused features is retained too much. In both cases, the performance of GPR will be affected.

This section explains why the kernel parameter of KPCA_IRBF has a much larger selectable region than that of KPCA_RBF. The training dataset in this work are adopted as the test data from which the machining parameters need to be deleted. Performance analysis of KPCA_IRBF and KPCA_RBF is carried out. Dimensionality comparison of the fused features of KPCA_IRBF and KPCA_RBF ($\lambda = 90\%$) with kernel parameters under different orders of magnitude is shown in Fig. 14. The red dashed line represents the dimensionality of the extracted features.

It can be seen that the dimensionality of the fused features of KPCA_RBF decreases greatly with the increase of the kernel parameter. A suitable dimensionality is hard to be determined so as to remove noises and preserve the valid information as much as possible. Thus, determination of the kernel parameter in KPCA_RBF needs to be carried out by trial and error.

However, the dimensionality of the fused features of KPCA_IRBF still stabilizes at a certain value with the exponential growth of kernel parameter. The valid information associated with tool wear status is preserved and noises are removed. Selectable region of the kernel parameter reaches five orders of magnitude which is very encouraging. Although KPCA_IRBF is confronted with the problem that an extremely small or large kernel parameter will bring about lots of redundant information, the selection of kernel parameter in KPCA_IRBF can be easily carried out in a much larger selectable region and parameter optimization is not necessary. Thus, the selection of kernel parameter in KPCA_IRBF is much easier in comparison with KPCA_RBF.

References

[1] S. Kurada, C. Bradley, A review of machine vision sensors for tool condition monitoring, Comput. Ind. 34 (1997) 55–72.
[2] R. Teti, K. Jemielniak, G.O. Donnell, D. Dornfeld, Advanced monitoring of machining operations, CIRP Ann-Manuf. Technol. 59 (2010) 717–739.
[3] J.V. Abellan-Nebot, F.R. Subirón, A review of machining monitoring systems based on artificial intelligence process models, Int. J. Adv. Manuf. Tech. 47 (2010) 237–257.
[4] C. Scheffer, H. Engelbrecht, P.S. Heyns, A comparative evaluation of neural networks and hidden Markov models for monitoring turning tool wear, Neural. Comput. Appl. 14 (2005) 325–336.
[5] N. Ghosh, Y.B. Ravi, A. Patra, S. Mukhopadhyay, S. Paul, A.R. Mohanty, A.B. Chattopadhyay, Estimation of tool wear during CNC milling using neural network-based sensor fusion, Mech. Syst. Signal Pr. 21 (2007) 466–479.
[6] B. Kaya, C. Oysu, H.M. Ertunc, Force-torque based on-line tool wear estimation system for CNC milling of Inconel 718 using neural networks, Adv. Eng. Softw. 42 (2011) 76–84.
[7] J. Xu, K. Yamada, K. Seikiya, R. Tanaka, Y. Yamane, Effect of different features to drill-wear prediction with back propagation neural network, Precis. Eng. 38 (2014) 791–798.
[8] F.A. Niaki, L. Feng, D. Ulutan, L. Mears, A wavelet-based data-driven modelling for tool wear assessment of difficult to machine materials, Int. J. Mech. Manuf. Syst. 9 (2016) 97–121.
[9] D. Shi, N.N. Gindy, Tool wear predictive model based on least squares support vector machines, Mech. Syst. Signal Pr. 21 (2007) 1799–1814.
[10] D.R. Salgado, F.J. Alonso, An approach based on current and sound signals for in-process tool wear monitoring, Int. J. Mach. Tool. Manuf. 47 (2007) 2140–2152.
[11] C. Zhang, H. Zhang, Modelling and prediction of tool wear using LS-SVM in milling operation, Int. J. Comput. Integ. Manuf. 29 (2016) 76–91.

[12] S. Dutta, S.K. Pal, R. Sen, On-machine tool prediction of flank wear from machined surface images using texture analyses and support vector regression, Precis. Eng. 43 (2016) 34–42.
[13] J. Wang, J. Xie, R. Zhao, L. Zhang, L. Duan, Multisensory fusion based virtual tool wear sensing for ubiquitous manufacturing, Robot. Cim. Int. Manuf. 45 (2016) 47–58.
[14] D. Kong, Y. Chen, N. Li, S. Tan, Tool wear monitoring based on kernel principal component analysis and v-support vector regression, Int. J. Adv. Manuf. Tech. 89 (2016) 1–16.
[15] D. Wu, C. Jennings, J. Terpenny, R.X. Gao, S. Kumara, A comparative study on machine learning algorithms for smart manufacturing: tool wear prediction using random forests, J. Manuf. Sci. Eng. 139 (2017), 071018–071018–9.
[16] C.E. Rasmussen, C.K.I. Williams, Gaussian Processes for Machine Learning [M], The MIT Press, 2006.
[17] M. Ebden, Gaussian Processes for Regression: A Quick Introduction, August, 2008
[18] B. Wang, T. Chen, Gaussian process regression with multiple response variables, Chemometr. Intell. Lab. 142 (2015) 159–165.
[19] C. Zhang, H. Wei, X. Zhao, T. Liu, K. Zhang, A Gaussian process regression based hybrid approach for short-term wind speed prediction, Energ. Convers. Manage. 126 (2016) 1084–1092.
[20] Y. Wang, B. Chaib-Draa, KNN-based Kalman filter: An efficient and non-stationary method for Gaussian process regression, Knowl.-Based Syst. 114 (2016) 148–155.
[21] S.A. Aye, P.S. Heyns, An integrated Gaussian process regression for prediction of remaining useful life of slow speed bearings based on acoustic emission, Mech. Syst. Signal Pr. 84 (2017) 485–498.
[22] Y. Liu, Y. Pan, D. Huang, Q. Wang, Fault prognosis of filamentous sludge bulking using an enhanced multi-output gaussian processes regression, Control Eng. Pract. 46–54 (2017).
[23] L.J.P.V. Maaten, E.O. Postma, H.J.V.D. Herik, Dimensionality Reduction: A Comparative Review, J. Mach. Learn. Res. 10 (2009).
[24] Z. Liu, Z. Liu, Y. Peng, Dimension reduction of Karhunen-Loeve expansion for simulation of stochastic processes, J. Sound Vib. 408 (2017) 168–189.
[25] B. Schölkopf, A. Smola, K.R. Müller, Nonlinear component analysis as a kernel eigenvalue problem, Neural Comput. 10 (1998) 1299–1319.
[26] J.M. Lee, C.K. Yoo, S.W. Choi, P.A. Vanrolleghem, I.B. Lee, Nonlinear process monitoring using kernel principal component analysis, Chem. Eng. Sci. 59 (2004) 223–234.
[27] A. Tanaka, H. Imai, M. Kudo, M. Miyakoshi, Integrated kernels and their properties, Pattern Recogn. 40 (2007) 2930–2938.
[28] S.G. Mallat, A theory for multiresolution signal decomposition: the wavelet representation, IEEE Trans. Pattern. Anal. 11 (1989) 674–693.
[29] K. Zhu, Y.S. Wong, G.S. Hong, Wavelet analysis of sensor signals for tool condition monitoring: A review and some new results, Int. J. Mach. Tool. Manu. 49 (2009) 537–553.
[30] C.E. Rasmussen, H. Nickisch, Gaussian Processes for Machine Learning (GPML) Toolbox, J. Mach. Learn. Res. 11 (2010) 3011–3015.
[31] B. Schölkopf, A.J. Smola, R.C. Williamson, P.L. Bartlett, New support vector algorithms, Neural Comput. 12 (2000) 1207–1245.
[32] X.H. Shi, Y.C. Liang, H.P. Lee, W.Z. Lin, X. Xu, S.P. Lim, Improved Elman networks and applications for controlling ultrasonic motors, Appl. Artif. Intell. 18 (2004) 603–629.