Tool wear predicting based on weighted multi-kernel relevance vector machine and probabilistic kernel principal component analysis

Guohao Song1,2 · Jianhua Zhang1,2 · Yingshang Ge1,2 · Kangyi Zhu1,2 · Zhensheng Fu1,2 · Luchuan Yu3

Received: 25 March 2022 / Accepted: 9 July 2022 / Published online: 7 September 2022
© The Author(s), under exclusive licence to Springer-Verlag London Ltd., part of Springer Nature 2022

Abstract
This paper proposes a novel tool wear predicting method based on a weighted multi-kernel relevance vector machine (WMKRVM) and the integrated radial basis function–based probabilistic kernel principal component analysis (PKPCA_IRBF). The proposed WMKRVM model is constructed using the optimized standard single kernel RVM and its weight parameters. As a new dimension increment technique, PKPCA_IRBF can extract the noise information of the cutting force signal feature and incorporate the noise information into the model. Moreover, PKPCA_IRBF is first proposed to fuse the cutting force signal feature to improve the confidence interval provided by the WMKRVM model. Compared to the traditional PKPCA_RBF method, PKPCA_IRBF has a broader range of kernel parameter selection intervals and higher model accuracy. The cutting experiment is carried out to validate the effectiveness of the proposed tool wear predicting technique. Experimental results show that the proposed tool wear predicting technique can accurately monitor the tool wear width with strong robustness under various cutting conditions, laying the foundation for application in the industrial field.

Keywords Feature fusion · Multi-kernel relevance vector machine · Probabilistic kernel principal component analysis · Tool wear monitoring

1 Introduction
The workpiece quality and economy of the manufacturing system are the most susceptible to tool wear status. Thus, tool condition monitoring (TCM) and prediction are indispensable in an intelligent manufacturing system. To guarantee the stability of the intelligent manufacturing system, an efficient monitoring system for real-time and precise evaluation of the tool wear status is highly desirable.

Many remarkable achievements have emerged in tool condition monitoring and prediction with recent sensor technology and machine learning theory advancements. The researchers [1–4] review methods used for TCM and provide a comprehensive survey of sensor technologies, signal processing, and decision-making system for process monitoring. By extensively reviewing papers in different areas, researchers summarized the development trend of TCM. They gave suggestions for the development of TCM in the future, thus promoting the cross-fertilization of concepts and techniques within TCM research. Recently, artificial intelligence (AI) technology, such as the hidden Markov model, long short-term memory (LSTM), and artificial neural networks, have been applied with considerable success. Liu et al. [5] propose a novel approach for tool wear monitoring based on a novel switching hidden semi-Markov model to represent the equipment’s degradation process. Han et al. [6] research the sticky hierarchical Dirichlet process hidden Markov model. Huang et al. [7] combine the advantages of bidirectional LSTM networks and particle filter and mitigate their limitations. Sun et al. [8] use the LSTM network to establish a residual convolutional neural network to predict multiple flank wear values. Xu et al. [9] propose an intelligent model-adaptive neuro-fuzzy inference system to estimate the tool wear. Shi et al. [10] present a novel deep
The International Journal of Advanced Manufacturing Technology (2022) 122:2625–2643

learning data-driven modeling framework for TCM in an ultraprecision machine. Wang et al. [11] present a multiscale principal component analysis method to realize online tool wear monitoring of the milling process. Huang et al. [12] propose new tool wear predicting method based on multi-domain feature fusion by a deep convolutional neural network. Multi-domain (including time domain, frequency domain, and time–frequency domain) features are extracted from multisensory signals as health indicators of tool wear condition. A predicting model using advanced machine learning methods with a multi-feature multi-model ensemble and dynamic smoothing scheme is developed by Shen et al. [13]. However, these methods focus on improving the test dataset’s accuracy and lack the prediction result’s uncertainty estimation or probability.

Relevance vector machine (RVM) is a sparse probability model based on kernel function and originally was derived and experimented on binary classification [14]. Researchers have successfully applied RVM to many engineering fields since it has lower requirements on training samples and is robust to noise samples [15, 16]. However, traditional single-kernel RVM has difficulty capturing the structure of nonhomogeneous data from multiple data sources or different data types. Therefore, researchers choose multiple types of kernel functions to build multi-kernel models to exploit the advantages of different types of kernel functions. Chen et al. [17] presented an early fault diagnosis method for rolling bearings based on a multi-kernel RVM. Experimental results show that the proposed method has higher prediction accuracy than the traditional single kernel RVM.

Inspired by the multi-kernel model, this paper proposes a new tool wear predicting model based on the weighted multi-kernel relevance vector machine (WMKRVM). In this work, the weight parameter and the hyperparameters of each kernel function are optimized by the sparrow search algorithm (SSA) [18]. Moreover, this work proposes the integrated radial basis function-based probabilistic kernel principal component analysis (PKPCA_IRBF) to improve the confidence interval (CI) of the WMKRVM model.

Dimension-fusion techniques, including kernel principal component analysis (KPCA), and isometric feature mapping, have been widely used for feature selection and fusion [19]. Feature fusion can effectively compress and remove noise and redundancy from the data, which is vital for improving the precision and reliability of the decision-making system. In this research, the probabilistic kernel principal component analysis (PKPCA) technique [20] is used to fuse the extracted feature to improve the WMKRVM model’s performance. Kim et al. [21] proposed a multivariate process monitoring method based on probabilistic principal component analysis and applied it to process monitoring. Neil [22] proposes a novel dual probabilistic principal component analysis (DPPCA). The DPPCA model has the advantage that the linear mappings from the embedded space can easily be nonlinearized through Gaussian processes. The basic idea of PKPCA is to quantify the randomness of process variables contaminated by random noise. Typically, the kernel function determines the performance of PKPCA. Since it is hard to determine the effective and noisy information preserved by the radial basis function (RBF), this research adopts the integrated radial basis function (IRBF) [23] as the kernel function of PKPCA to fuse features. The proposed PKPCA_IRBF technique can enrich noise while preserving effective tool wear information. Experimental results show that the utilization of PKPCA_IRBF will be conducive to improving the confidence interval of the WMKRVM model.

This paper is organized as follows: In the first section, the experimental and data collection is introduced. The theory of PKPCA_IRBF, WMKRVM, and SSA is shown in the second section. In the third section, the proposed model and the experimental results are analyzed. Finally, conclusions are given in the fourth section.

2 Experimental setup

The cutting process is conducted on the CNC lathe (PRE-CION CKD6150H) to validate the effectiveness of the proposed tool wear predictive model. A CNC lathe, cutting force monitoring system, experimental material, HD CCD video microscope, and cutting tool constitute the experimental setup, as shown in Fig. 1. The cutting force monitoring system is composed of a dynamometer, charge amplifier, data acquisition card, and computer with DynoWare.

The software DynoWare collects the cutting force signal during the machining process in real time at a sampling rate of 20 kHz and displays it on a computer. The experimental material is 50# normalized steel (HB160 ~ 197) with a cutting diameter of 150 ~ 80 mm and a cutting length of 140 mm. The cemented carbide indexable insert (Sandvik CION CKD6150H) is selected as the cutting tool and connected to the dynamometer via a tool holder (Sandvik PCLNR 2525 M 12). The tool wear width VB at half of the cutting depth, measured by an HD CCD video microscope, quantifies the tool wear degree. The whole service life of the cutting tool is the time from new tool to tool breakage (VB > 0.5 mm). HD CCD video microscope measures the tool wear width VB and records it. The laboratory instruments used in this research are shown in Table 1. The experiment is carried out under different machining conditions, as shown in Table 2, with varying feed rates and cutting depths.

Three orthogonal cutting forces contain rich information about the change of tool wear width during machining. The time domain, frequency domain, and wavelet domain feature listed in Table 3 are extracted from the original orthogonal
cutting forces signals within 5 s as feature reflecting tool wear. Root mean square (RMS) reflects the effective value of the signal in a given time interval. The dispersion of the cutting force signal and the difference between the original force signal and the expected value are measured by variance. The maximum value (Max) represents the maximum instantaneous amplitude of the signal within a specific time range. Skewness measures the direction and degree of the skew of the probability density curve distribution of the data, that is, the degree of asymmetry of the data compared to the average value. Peak to peak (PP) represents the difference between the highest and lowest values of the data in a period and describes the size of the range of data change. Kurtosis represents the maximum size of the probability density distribution curve of the signal at the average value.

The kurtosis visually reflects the peak sharpness of the curve. The higher the kurtosis, the greater the thickness at the bottom of the curve, which leads to more significant variance. The reason is that the signal has a certain number of differences that are significantly larger or smaller than the average value.

**Fig. 1** The experimental setup for cutting

**Table 1** The type of laboratory instrument

| Instrument                  | Type            | Application                                                                 |
|-----------------------------|-----------------|----------------------------------------------------------------------------|
| HD CCD video microscope     | BD-100          | Measure and record tool wear                                               |
| Dynamometer                 | Kistler 9257B   | Measure cutting force when turning, cutting, grinding                      |
| Multichannel charge amplifier| Kistler 5080A   | Suitable for force measurements when a wide measuring range or high-quality signal is needed |
| Data acquisition card       | Kistler 5697A   | Measures highly-dynamic processes                                          |
| DynoWare                    | Kistler 2825D   | A general-purpose data acquisition and analysis software                    |
average. Wavelet packet decomposition effectively decomposes the original signal into multiple sub-band. Based on wavelet transform, the detailed sub-band is further decomposed. Finally, the optimal signal decomposition path is calculated by minimizing the cost function, which decomposes the original input signal. As a sparse basis, the Daubechies wavelet introduces the smoothing error of the signal. It has reasonable regularity, the signal reconstruction process is smooth, the order of $N$ can control the localization ability in the frequency domain, and the flexibility is high. And the packet decomposition process uses the Daubechies wavelet function basis.

This paper adopts the tool wear width $VB$ of the cutting tool as the object of research. As shown in Table 2, all data files are collected and recorded in the twelve cutting tests. One data file corresponds to one cutting experiment. In addition, features need to be fused to attenuate the noise to reduce its negative impact. This work combines the extracted original signal or fused features with the corresponding processing parameters to form feature vectors fed into the proposed model. All signal features are mixed and divided into training and test datasets to avoid errors brought by individual experiments. All feature vector samples are obtained and divided into training and test datasets at random at a ratio of 9:1, which do not intersect.

Table 2  Machining condition of the experiment

| Test | Cutting speed $V_c$ (m/min) | Feed rate $f$ (mm/r) | Cutting depth $a_p$ (mm) | Repetition |
|------|---------------------------|---------------------|-------------------------|------------|
| 1    | 300                       | 0.2                 | 1                       | 3          |
| 2    | 300                       | 0.3                 | 1                       | 3          |
| 3    | 300                       | 0.4                 | 1                       | 3          |
| 4    | 300                       | 0.2                 | 2                       | 3          |

The extracted signal features need to be normalized before the feature fusion:

$$x' = \frac{x - \bar{x}}{\sigma_x}$$  \hspace{1cm} (1)

where $\bar{x}$ is the mean value of cutting force, $\sigma_x$ is the standard deviation.

3 Theoretical background

3.1 Probabilistic kernel principal component analysis

Probabilistic principal component analysis was proposed by Tipping and Bishop within the maximum likelihood framework to analyze kernel principal components in a probabilistic manner [23, 24]. Probabilistic analysis supposes that the data in the feature space follow a special factor analysis model which relate a $p$ dimensional observation data $x$ to a latent $q$ dimensional variable $z$ as:

$$x = Wz + \mu + \varepsilon$$  \hspace{1cm} (2)

Where $z \sim \mathcal{N}(0, I_q)$, $\varepsilon \sim \mathcal{N}(0, \sigma^2 I_p)$. $W \in R^{p \times q}$ is the loading matrix. $\mu$ is the mean vector. Then, the marginal distribution of $x \sim \mathcal{N}((\mu, \Sigma)$ obtained by

$$p(x) = \int p(x|z)p(z)dz = (2\pi)^{(-d/2)}|\Sigma|^{-1/2}\exp\{-1/2(x-\mu)^T\Sigma^{-1}(x-\mu)\}$$  \hspace{1cm} (3)

The model covariance is computed by $\Sigma = WW^T + \sigma^2 I$. The log-likelihood of the observation data is expressed as follows:
\[ L = \ln p(x|\mu, W, \sigma^2) = -\frac{N}{2} \{ p\ln(2\pi) + \ln|\Sigma| + Tr(\Sigma^{-1}S) \} \]

where \( S = \frac{1}{N} \sum_{n=1}^{N} (x_n - \mu)(x_n - \mu)^T \). Based on the maximum likelihood estimates, the weight matrix \( W \) and the noise variance \( \sigma^2 \) are expressed as follows:

\[ W = U_q(\Lambda_q - \sigma^2 I)^{1/2} R \]

\[ \sigma^2 = \frac{1}{p-q} \sum_{j=q+1}^{p} \lambda_j \]

where \( S = U_q\Lambda_q U_q^T \) and \( R \) is an arbitrary \( q \times q \) orthogonal rotation matrix. \( \lambda_{q+1}, \ldots, \lambda_q \) are the smallest eigenvalue of \( S \). Based on the Bayesian theory, the associated posterior distribution of the latent variable \( z \) given a specific observed data \( x \) is expressed as follows:

\[ z|x \sim N(M^{-1}W^T(x - \mu), \sigma^2M^{-1}) \]

where \( M = \sigma^2I + WW^T \in \mathbb{R}^{p\times p} \) is the posterior covariance. In the case of PKPCA, a nonlinear expectation maximization (EM) algorithm has been used for parameter learning. The EM algorithm iterates two steps, expectation (E-step) and maximization (M-step) until convergence, and local maxima of the data likelihood can be guaranteed.

\[ \bar{W} = HKHW(M^{-1}W^THKHW + \sigma^2I)^{-1} \]

\[ \hat{\sigma}^2 = \frac{1}{p} tr((HKH - HKHW\hat{M}^{-1}W)^T) \]

where \( H = I - \frac{1}{p} 1_p 1_p^T \). \( I \) and \( 1_p \) denote the \( p \times p \) identity matrix and the \( p \times 1 \) vector of one, respectively. It is important to note that it works with the kernel matrix \( K \) only.

Selection of the kernel function directly affects the performance of the fused feature obtained from PKPCA. In this paper, the integrated radial basis function (IRBF) is adopted as the kernel function of PKPCA, referred to as the PKPCA_IRBF technique and is expressed as follows:

\[ K(x, y) = \int_{0}^{\sqrt{2\sigma_0}} \exp\left(-\frac{|x - y|^2}{\sigma^2}\right)d\sigma \]

where \( \sigma_0 \) is the kernel parameter.

### 3.2 Weighted multi-kernel relevance vector machine

RVM is a sparse learning classifier proposed by Tipping based on the Bayesian inference. Supposing that a dataset is \( \{x_i, y_i\}_{i=1}^{m} \), \( x_i \in \mathbb{R}^d \), \( y_i \in \mathbb{R} \), \( m \) is the number of cutting data samples, which follows the probabilistic formulation \( p(y_i|x_i) = N(y_i|f(x_i), \sigma^2) \):

\[ y_i = f(x_i) + \epsilon_i \]

where the additive noise \( \epsilon_i \sim N(0, \sigma^2) \). The regression function of the RVM model corresponds to those implemented by SVM, which means the target values of \( y_i \) are the linear weighted sum of the nonlinear basis function:

\[ f(x) = \sum_{i=1}^{m} w_i K(x, x_i) + w_0 \]

where \( K(x, x_i) \) is the kernel function of RVM, \( w_i \) is the model weight value, and \( w_0 \) is the bias. The corresponding likelihood function is defined as:

\[ p(y|\omega, \sigma^2) = \prod_{i=1}^{m} N(y_i|f(x_i), \sigma^2) = (2\pi \sigma^2)^{-N/2} \exp\left(-\frac{1}{2\sigma^2}||y - \Phi\omega||^2\right) \]

where \( y = [y_1, y_2, \ldots, y_m]^T \) is the target vector of cutting data, and \( w = [w_1, w_2, \ldots, w_m]^T \) is the weight vector correspond to \( \Phi = [\phi(x_1), \phi(x_2), \ldots, \phi(x_m)]^T \), which is the “design” matrix, wherein \( \phi(x_i) = (1, k(x_i, x_1), k(x_i, x_2), \ldots, k(x_i, x_m))^T \).

The introduction of an individual hyperparameter \( \alpha_i \) for each weight parameter \( \omega_i \) is the key feature of the RVM. According to the Bayesian theory, the traditional method is that direct execution of maximum-likelihood estimation of \( \omega \) and \( \sigma^2 \) will lead to severe overfitting. To overcome this problem, a zero-mean Gaussian prior distribution over \( \omega_i \) with a different precision \( \alpha_i \) is imposed to constrain their variation. Thus, the prior distribution of \( \omega \) is expressed as follows:

\[ p(\omega|\alpha_\sigma) \propto \prod_{i=1}^{m} N(\omega_i|0, \alpha_i^{-1}) \]

\[ p(\omega|\alpha_\sigma) \propto (2\pi)^{-m/2} \prod_{i=1}^{m} \alpha_i^{1/2} \exp\left(-\frac{\omega_i^2}{2}\right) \]

where \( \alpha \) is a vector of \( m + 1 \) hyperparameter. Consequently, the posterior distribution over the weight \( w \) conditioned on the cutting data is expressed as follows:

\[ p(w|y, \alpha, \sigma^2) \propto \left( \frac{p(y|\omega, \sigma^2)p(w|\alpha)}{p(y|\alpha, \sigma^2)} \right) = \left( \frac{(2\pi)^{-m/2} \prod_{i=1}^{m} \alpha_i^{1/2} \exp\left(-\frac{\omega_i^2}{2}\right)}{p(y|\alpha, \sigma^2)} \right) \]

\[ \exp\left(-\frac{1}{2\sigma^2}(w - \mu)^T\Sigma^{-1}(w - \mu)\right) \]

where \( p(y|\alpha, \sigma^2) = \int p(y|\omega, \sigma^2)p(w|\alpha)dw \) is convolution of Gaussian. The variance matrix \( \Sigma \) and the weight mean matrix \( m \) over the posterior weight \( w \) follow a Gaussian distribution and are expressed as follows:

\[ \Sigma = (\sigma^{-2} \Phi^T \Phi + A)^{-1} \]

\[ \mu = \sigma^{-2} \Sigma \Phi^T y \]

where \( A = \text{diag}(a_0, a_1, \ldots, a_m) \). \( A = \text{diag}(a_0, a_1, \ldots, a_m) \).

Under the Bayesian framework, the weight posterior \( p(w|y, \alpha, \sigma^2) \) is calculated by maximizing the marginal likelihood
\[ p(y|\alpha, \sigma^2). \text{ The posterior of hyperparameter is expressed as follows:} \]
\[ p(\alpha, \sigma^2|y) \propto p(y|\alpha, \sigma^2)p(\alpha)p(\sigma^2) \]  
(18)

which means \( \alpha \) and \( \sigma^2 \) are calculated by seeking the approximation function \( p(\alpha, \sigma^2|y) \approx \delta(\alpha_{MP}, \sigma_{MP}^2) \). The logarithm of RVM marginal likelihood is expressed as follows:
\[ L = \log p(y|\alpha,\sigma^2) = -\frac{1}{2}[N \log 2\pi + y^T C^{-1} y + \log |C|] \]  
(19)

where \( C = \Phi A^{-1} \Phi^T + \sigma^2 I \).

The values of \( (\alpha_{MP}, \sigma_{MP}^2) \) can be obtained iteratively by using the updating formulae:
\[ a_{i}^{\text{new}} = \frac{y_i}{\mu_i^2} \]  
(20)

\[ (\sigma^2)_{\text{new}} = \frac{||y - \Phi \mu||}{N - \sum_{i=1}^{m} y_i} \]  
(21)

where \( y_i = 1 - \alpha_i \Sigma_i \) wherein \( \Sigma_i \) is the \( i \)th diagonal element of the posterior covariance matrix \( \Sigma \). The prediction distribution can be based on the posterior distribution of the weights, conditional on maximizing the values \( \alpha_{MP} \) and \( \sigma_{MP}^2 \). We then calculate the predictive distribution for new test data \( x_i^* \):
\[ p(y_i|\alpha_{MP}, \sigma_{MP}^2) = \int p(y_i|w, \alpha_{MP}, \sigma_{MP}^2) \]  
\[ p(w|y, \alpha_{MP}, \sigma_{MP}^2) dw = N(y_i|f_i, \sigma_i^2) \]  
(22)

where the predictive mean \( f_i \) and variance \( \sigma_i^2 \) are expressed as follows:
\[ f_i = \mu_i^T \phi(x_i) \]  
(23)

\[ \sigma_i^2 = \sigma_{MP}^2 + \phi(x_i)^T \Sigma \phi(x_i) \]  
(24)

where the mean \( f_i \) and the variance \( \sigma_i^2 \) denote the predicted value and the uncertainty of the predicted distribution at the test data \( x_i \), respectively. The 95% CI of the predicted results is \( [f_i - 1.96 \sigma_i, f_i + 1.96 \sigma_i] \).

In learning RVM modeling, choosing an appropriate kernel function with a suitable hyperparameter is complicated. Constructing kernel function with better learning ability and generalization requires lots of experimental and computing resources. However, the multi-kernel model utilizing several kernel functions exhibits higher performance than a single kernel function. The linear combination of kernel functions is the most common way to construct the multi-kernel RVM model and is expressed as follows:
\[ K_{mix}(x, x_i) = \rho K_1 + (1 - \rho) K_2 \]  
(25)

s.t. 0 ≤ \( \rho \) ≤ 1, \( \rho \in \mathbb{R} \) where \( K_1 \) and \( K_2 \) are different kernel functions (e.g., the radial basis function and polynomial kernel function). \( \rho \) is the weight parameter for kernel function in combination with kernel function. Multi-kernel function integrates the excellent characteristics of different kernel functions, showing better learning and generalization ability when dealing with more complex data features. This paper adopts the linear weighted combination kernel function, which can adaptively adjust the weight of each independent kernel function for different input samples to achieve the advantages of varying kinds of basic kernel functions.

\[ K(x, x_i) = \sum_{s=1}^{M} \rho_s K_s(x, x_i) \]  
(26)

where \( M \) is the number of the kernel function.

Multi-kernel function integrates the advantage of the different kernel functions. For example, the radial basis function kernel function captures the local nonlinear change trend in the tool wear process. The linear kernel function captures the global monotonic decreasing trend of tool wear. Since the kernel function of SVM does not have to satisfy Mercer’s theorem as SVM, \( K_s(x, x_i) \) can be any kernel function. Optimization of the weight parameter \( \rho_s \geq 0 (\sum_{s=1}^{M} \rho_s = 1) \) is required to ensure that the kernel function maps the data with high performance.

### 3.3 Sparrow search algorithm

The Sparrow search algorithm (SSA) is a swarm intelligence optimization algorithm inspired by the foraging and anti-predation behavior of the sparrow population proposed in 2020. As sparrow is an intelligent social creature with an excellent memory, the sparrow population has biological characteristics in foraging:

1. It divided sparrow populations into producers and predators. Producers have more extensive search space to find food sources, while others search for food based on producers.
2. Sparrows have active evaders; some can be chosen as Scouters in foraging.
3. If better food sources appeared, producers and scroungers would dynamically switch.
4. Scroungers can always find producers to provide better food sources; scroungers even monitor producers to get more food.

The flowchart of SSA is presented in Fig. 2.
The construction of a tool wear predicting model based on PKPCA-IRBF and WMKRVM

The reasonable selection of weight of kernel function determines the performance and sparsity of RVM. The commonly used parameter optimization method for MKRVM is to encode MKRVM’s parameters in the optimization algorithm’s solution. The process is time-consuming and laborious and requires high optimization performance from the optimization algorithm. To obtain suitable hyperparameters for the WMKRVM model and prevent over-fitting in the test dataset, the SSA optimized the hyperparameters of the WMKRVM model in stages.

SSA is a novel swarm optimization approach, mainly inspired by sparrows’ group wisdom, foraging, and anti-predation behaviors [18]. It offers high search accuracy and robustness compared to other algorithms. Experiment results of nineteen standard test functions [25–28], including unimodal test functions, multimodal test functions, and fixed-dimension test functions, verify the feasibility and effectiveness of SSA. Compared with grey wolf optimizer [29], gravitational search algorithm [25], and particle swarm optimization [30], SSA is superior in searching precision, convergence speed, stability, and robustness. Therefore, SSA is adopted as the parameter optimization method for MKRVM in this research.

The SSA optimizes the kernel parameter of the standard kernel function, then the optimized kernel function comprises the multi-kernel function of RVM. Finally, the SSA algorithm optimizes the weight parameters of the WMKRVM model.

The weight parameter $\rho$, of kernel function make up the search agent:

$$x_{ij} = (x_{i1}^{t}, x_{i2}^{t}, ..., x_{iM}^{t}) = (\rho_1, \rho_2, ..., \rho_M) \quad (27)$$

where $x_{ij}^{t}$ is the $j$th dimension of the $i$th agent. $t$ is the number of optimization calculation. The weight parameter optimized in WMKRVM is contained in the search agent $x_{ij}^{t} = (\rho_1, \rho_2, ..., \rho_M)$. The target function of SSA should be defined in a proper form.

The mean square error (MSE) between the target value and the predictive mean of the RVM is adopted as the target function of SSA, as shown in Table 5. The training dataset is used to train and optimize the parameters of the WMKRVM model. Besides, $k$-fold cross-validation is utilized to determine the target function values for each search agent of SSA, as shown in Fig. 3.

Subsequently, each subset was used once to validate the accuracy of the WMKRVM-based tool wear predictive model, which is trained by the remaining $k-1$ subsets. The training process is then executed $k$ times, and each subset is used only once as the validation dataset. The MSE between the actual value and the predicted value is used as the performance indicator of the model. The sum of validation results is used as the fitness value for the search agent. The iterative process continues until the maximum number of iterations is reached, and the optimized weight parameters of WMKRVM are found. After finishing the weight parameter optimization process of the WMKRVM model, the optimized weight parameters corresponding to the minimum fitness are generated and combined with the training set to construct the tool wear prediction model.

3.4 Construction of tool wear predicting model based on PKPCA-IRBF and WMKRVM

Fig. 2 The flowchart of Sparrow Search Algorithm
The training and test datasets provided by PKPCA_IRBF are used to optimize the hyperparameters of the proposed model and evaluate its performance. When the convergence value is satisfied, the iterative process of \((\alpha, \beta)\) will stop. The maximum of the iterative process shown in Fig. 4 is defined as 1000. The iterative process will stop when satisfying the convergence condition \(\max{|\alpha_{\text{new}}| - \max{|\alpha_{\text{old}}|} < \delta_{\text{Thresh}}}\). Once the iteration process is finished, the modeling parameter will be determined.

The initial hyperparameters \((\alpha, \beta)\) are determined by the Harris hawks optimization (HHO) [31] and \(k\)-fold cross-validation and have little impact on the performance of the WMKRVM model. Only the training dataset participates in searching the hyperparameters \((\alpha, \beta)\). The interval range of \((\alpha, \beta)\) are determined by trial and error: \(\alpha \in [10^{-4}, 10^{-1}], \beta \in [10^{-2}, 10^{1}]\).

### 4 Experimental results analysis

The tool wear predictive model based on the WMKRVM model and PKPCA_IRBF proposed in this paper aims at real-time and accurate monitoring tool wear during machining. The cutting experiment is carried out to validate the effectiveness of the proposed tool wear predictive model, as shown in Fig. 1. Data acquisition for cutting experiments

| Kernel function | Math equation | The initial interval range of hyperparameters |
|-----------------|---------------|---------------------------------------------|
| Gaussian kernel | \(K(x, y) = \exp\left(-\frac{||x-y||^2}{2\sigma_0^2}\right)\) | \(\sigma_0 \in (0.01, 50)\) |
| Laplace kernel  | \(K(x, y) = \exp\left(-\frac{||x-y||}{\sigma_0}\right)\) | \(\sigma_0 \in (0.01, 50)\) |
| Multiquadric kernel | \(K(x, y) = \left(||x-y||^2 + \epsilon^2\right)^{0.5}\) | \(\sigma_0 \in (0.01, 50)\) |
| Polynomial kernel | \(K(x, y) = (ax^T y + c)^d\) | \(a \in (1, 3), c \in (0.01, 50), d \in \{1, 2, 3\}\) |
| Exponential kernel | \(K(x, y) = \exp\left(-\frac{||x-y||}{2\sigma_0^2}\right)\) | \(\sigma_0 \in (0.01, 50)\) |
| Multi-kernel | \(K(x, x_i) = \sum_{s=1}^{5} \rho_s K_s(x, x_i)\) | \(\rho_s \in (0, 1)\) |
The purpose of multi-kernel is to integrate the advantages of different kernel functions to obtain better performance. When dealing with data with complex structural features, the multi-kernel performs better than the single kernel. Linear combinations of kernel functions are often used to construct the multi-kernel function. The linear weighted combination kernel function adopted in this paper is different from the standard linear combination of the local and global kernel functions. The proof that the prediction performance of the WMKRVM model outperforms single kernel RVM is:

Supposed that the set formed by the optimal solution of different single kernel RVM $K_i$ is expressed as follows: $\mathbb{K} = \{K_1, K_2, \ldots, K_N\}$.

Let $f = \min K_1 = \min \{K_1, K_2, \ldots, K_N\}$. Then the proposed WMKRVM model $g$ consisting of different single kernel function $g = \sum_{i=1}^{N} a_i K_i$, where $a_i \in [0, 1] \cup [0, 1]$, $\sum_{i=1}^{N} a_i = 1$, its optimum solution $g_{\min} \leq f$.

Counterproof:

Hypothesis $\forall a_i \in [0, 1]$, $i = 1, 2, \ldots, N$, $g_{\min} > f$. There exist $i = k \in \{1, 2, \ldots, N\}$ such that $f = K_k$, because $f = \min \{K_1, K_2, \ldots, K_N\}$.

Let $a_k = 1$ and $a_i = 0$, $i = 1, 2, \ldots, k−1, k+1, \ldots, N$, then.

$g = \sum_{i=1}^{N} a_i K_i = 0 \times K_1 + \cdots + 0 \times K_{k−1} + 1 \times K_k + \cdots + 0 \times K_N$

It’s means $g = K_k = f$.

The hypothesis does not hold. The optimum solution of WMKRVM $g_{\min} \leq f$ is proved.

In recent years, there has not been a unified theory and methodology for selecting kernel and its weight parameters. Commonly used methodology for optimizing kernel parameters include empirical selection, experimental comparison, large-scale search, and cross-validation. This research adopts SSA to optimize the hyperparameter of the WMKRVM model.
model. The convergence process curves for different single kernel RVM and the WMKRVM model optimized by SSA are shown in Fig. 5. SSA performs well for optimizing the hyperparameters of the RVM, as the fitness values of the RVM model remain stable after 100 iterations. The parameters of SSA are set to (100, 50, 1, 0.5, $10^{-9}$). The optimized kernel parameter and its weight parameter of the RVM model are listed in Table 6.

The performance of the tool wear predictive model is investigated by the evaluation indicator listed in Table 5.

![Fig. 6 Performance comparison of different RVM models based on training dataset ($\alpha = 10^{-4}, \beta = 10^{-2}$)](image-url)
In Table 5, \( y_i \) is the tool flank wear width and \( \bar{y} \) is the mean tool flank wear width measured by HD CCD video microscope. \( f_i \) is the predicted tool flank wear width value and \( \bar{f} \) is the mean predicted tool flank wear width value based on the tool wear predicting model at feature vector \( x_i \) of test dataset. MAE is the mean absolute error, i.e., the average absolute value of the deviation of all individual observations from the arithmetic mean. MAE avoids the errors from canceling each other out and reflects the actual forecast error. RMSE is the root mean square error, representing the sample

![Graphs showing performance comparison of different RVM models based on training dataset](image-url)

**Fig. 7** Performance comparison of different RVM models based on training dataset (\( \alpha = 10^{-4}, \beta = 10^{-2} \))
standard deviation of the difference between the predicted and observed values (residual) and explaining the sample’s dispersion. MAPE, the mean absolute percentage error, is a statistical indicator used to measure the accuracy of the prediction. PCC is the Pearson correlation coefficient and reflects the model’s performance.

The performance of different kernel RVM-based tool wear predicting models under different evaluation indicators based on training and test datasets is shown in Figs. 6 and 7, respectively. Compared to the Gaussian kernel, Laplace kernel, exponential kernel, polynomial kernel and multiquadric kernel, the weighted multi-kernel has the minimum MSE and RMSE values and the maximum PCC value based on the training dataset with the values 0.0026791, 0.050869, and 0.82959, respectively. The WMKRVM model performs better in the test dataset with the minimum MSE, RMSE, MAE, and MAPE values and the maximum PCC value with values 0.00094987, 0.03082, 0.021886, 0.11183, and 0.87582, respectively. The experiment results show that WMKRVM effectively combines the advantages of different kernel functions to capture the internal characteristics of complex data structures.

HHO and k-fold cross-validation are used to determine the initial hyperparameters \((\alpha, \beta)\) based on the predictive performance of RVM. Only the training dataset participates in the hyperparameter optimization process. The initial interval range of \((\alpha, \beta)\) is \((10^{-4}, 10^{-1}], [10^{-2}, 10^{3}]\). This research adopts the MSE between the target value and the predicted average of the RVM-based tool wear predictive model as the evaluation indicator. The optimization results of the hyperparameters \((\alpha, \beta)\) are listed in Table 7.

The test dataset evaluates the effectiveness of the proposed WMKRVM model. The optimized hyperparameters in Table 7 are used to construct the RVM model. The predicted results of different kernel RVM-based tool wear predictive models (Gaussian kernel, Laplace kernel, multiquadric kernel, polynomial kernel, exponential kernel, and multi-kernel) for the cutting process by using the original signal feature are shown in Fig. 8.

The real brown point represents the test point which is the actual tool wear value measured by HD CCD video microscope. The solid and dashed lines represent the predicted results and 95% CI obtained from the WMKRVM model, respectively. The deviation represents the absolute value of the difference between the predicted and the actual tool wear value.

The number above the bar in the figure represents the value of evaluation indicator. It is found that WMKRVM has the minimum MAE/RMSE/MAPE and the maximum PCC among the different kernel RVM models, as shown in Fig. 7. Therefore, it can be concluded that WMKRVM outperform traditional single kernel RVM in capturing the global structure of the nonlinear dataset.

However, the width of 95% CI obtained by WMKRVM does not capture the effect of process noise and redundancy on the model, affecting its practical application. Therefore, this paper adopts PKPCA_IRBF to fuse the original data features and sufficiently estimate the WMKRVM model’s data noise.

### 4.2 Model construction and evaluation based on PKPCA_IRBF fusion feature

The lower and upper bounds of the 95% CI are given by \(f_l = -2\sqrt{\sigma_i^2}\) and \(f_u = 2\sqrt{\sigma_i^2}\), respectively. However, the 95% CI of the predicted results obtained by the WMKRVM model did not accurately reflect the process noise. This means the model is lacking in the simulation of noise. Besides, the traditional PKPCA adopts RBF as its kernel function, referred to as the PKPCA_RBF technique, and is expressed as follows:

\[
K(x, y) = \exp\left(-\frac{||x - y||^2}{2\sigma_0^2}\right)
\]  

where \(\sigma_0\) is the kernel parameter of PKPCA_RBF. PKPCA_IRBF and PKPCA_RBF techniques are used to fuse the original force signal feature to ameliorate the 95% CI.

This research adopts PKPCA_IRBF and PKPCA_RBF to improve the model’s ability to simulate the noise and improve the 95% CI of the prediction results. The fused features obtained from PKPCA_IRBF and PKPCA_RBF

| Table 7 The hyperparameter of the RVM model |
|---------------------------------------------|
| RVM                      | \((\alpha, \beta)\)         |
|------------------------------------------------------------------|
| Gaussian kernel          | (0.36231, 351.68279)        |
| Laplace kernel           | (1.46194, 0.01)             |
| Multiquadric kernel      | (0.0001, 0.01)              |
| Polynomial kernel        | (0.0001, 401.17551)         |
| Exponential kernel       | (0.00011, 0.01316)          |
| Multi-kernel             | (0.14517, 1069.78859)       |

| Table 6 The optimized kernel parameter and the weight parameter of the RVM model |
|---------------------------------------------|
| RVM                      | Optimized hyperparameter values |
|------------------------------------------------------------------|
| Gaussian kernel          | (23.76904)                      |
| Laplace kernel           | (49.99982)                      |
| Multiquadric kernel      | (3.61620)                       |
| Polynomial kernel        | (1.05254, 1, 0.52895)           |
| Exponential kernel       | (6.81880)                       |
| Multi-kernel             | \((10^{-9}, 0.01282, 0.98689, 0.00029, 10^{-9})\) |
Fig. 8 Predicted results of the RVM model, \((\alpha, \beta)\) is as shown in Table 7
Fig. 9  a–f Original feature and the feature fused by PKPCA_RBF.

Fig.10  a–f The feature fused by PKPCA_IRBF.
Fig. 11 Predicted results of the WMKRVM-based tool wear predicting model by using the fused feature of PKPCA_IRBF: a $\sigma_0 = 1.5 \times 10^4$, $d = 69$; b $\sigma_0 = 1.5 \times 10^4$, $d = 85$; c $\sigma_0 = 1.5 \times 10^4$, $d = 99$; d $\sigma_0 = 1.5 \times 10^4$, $d = 146$; e $\sigma_0 = 1.5 \times 10^4$, $d = 164$; f $\sigma_0 = 1.5 \times 10^4$, $d = 197$
are used to comprise the training and the corresponding test datasets. Model construction and evaluation of WMKRVM are carried out using the training dataset and the corresponding test dataset. The evaluation indicator $CI_{\text{width}}$ is evaluated to quantify the reliability and stability of the 95% CI. The average width of the 95% CI is quantified by $CI_{\text{width}}$ given by:

$$CI_{\text{width}} = \frac{1}{N} \sum_{i=1}^{N} 4 \times \sqrt{\sigma_i^2}$$  (29)

This section adopts PKPCA_RBF and PKPCA_IRBF to fuse the original force signal features to improve the 95% CI of the WMKRVM model. Compared to the traditional KPCA_RBF technique, the proposed PKPCA_IRBF is a
new nonlinear dimension increment technique. The fused features constitute the feature vectors adopted as the input to the WMKRVM model.

The dense distribution and narrow discrete nature of the original force feature data, as shown in Fig. 9a, are not conducive to estimating the effects of noise and distinguishing changes between distinct data points. As shown in Fig. 9b–f, the distribution of the force features fused by PKPCA_RBF is uniform and discrete, indicating that PKPCA_RBF can effectively distinguish between distinct data points. However, the even data is not conducive to the model prediction. Although fusion features distinguish the differences between the overall data, uniformly distributed fusion features do not reflect the drastic changes in specific data points, which lowers the WMKRVM model’s prediction accuracy. Compared to the original force feature, the dispersion and volatility between the PKPCA_IRBF fusion features are more significant, which is beneficial in estimating the confidence interval of the WMKRVM model, as shown in Fig. 10a–f.

In this research, the kernel parameters in KPCA_IRBF are set to $\sigma_0=1.5 \times 10^4$ and $d=2\sim190$. Predicted results with corresponding 95% CI of the WMKRVM model using the fused feature of PKPCA_IRBF with the dimensions $d=69$, $d=85$, $d=99$, $d=146$, and $d=197$ are shown in Fig. 11, respectively. Performance evaluation of the WMKRVM-based tool wear predictive model under the fusion features of PKPCA_IRBF with different dimensions is shown in Fig. 12. The CI width obtained from the WMKRVM-based tool wear predictive model under the PKPCA_IRBF fused features is shown in Fig. 13. The cyan dashed line in Fig. 13 represents the CI width value obtained by the WMKRVM model. Compared to Fig. 8, the 95% CI of the proposed WMKRVM model has been improved under the PKPCA_IRBF fused features. It is found that the WMKRVM model under the PKPCA_IRBF fused features outperforms the WMKRVM model under the original features in terms of CI width. The selectable region for the kernel parameters in KPCA_IRBF is $\sigma_0=1.5 \times 10^4$ and $d=10\sim155$. 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.

The PKPCA RBF is also utilized to fuse the original feature for constructing the WMKRVM-based tool wear predictive model, revealing the superiority of PKPCA IRBF in this section. The best predictive performance of WMKRVM using the PKPCA_RBF-based fusion features is listed in Table 8. As can be shown, PKPCA IRBF outperforms PKPCA RBF. Therefore, it can be concluded that PKPCA IRBF can extract more effective signal features suitable for tool wear prediction.

### Table 8 Comparison between PKPCA_IRBF and PKPCA_RBF

| Evaluation indicator | Kernel parameters | MSE ($\times 10^{-3}$) | MAE ($\times 10^{-2}$) | RMSE ($\times 10^{-2}$) | MAPE ($\times 10^{-1}$) | SCC | PCC |
|----------------------|-------------------|------------------------|------------------------|------------------------|------------------------|-----|-----|
| PKPCA_IRBF           | $\sigma_0=1.5 \times 10^4, d=69$ | 1.3725 | 2.5850 | 3.7048 | 1.3968 | 0.7879 | 0.8139 |
|                       | $\sigma_0=1.5 \times 10^4, d=85$ | 1.3525 | 2.5617 | 3.6777 | 1.3720 | 0.7864 | 0.8174 |
|                       | $\sigma_0=1.5 \times 10^4, d=99$ | 1.3080 | 2.4965 | 3.6166 | 1.3495 | 0.8310 | 0.8246 |
|                       | $\sigma_0=1.5 \times 10^4, d=146$ | 1.2853 | 2.4937 | 3.5851 | 1.3704 | 0.7811 | 0.8285 |
|                       | $\sigma_0=1.5 \times 10^4, d=164$ | 1.2143 | 2.4196 | 3.4847 | 1.2930 | 0.8172 | 0.8391 |
|                       | $\sigma_0=1.5 \times 10^4, d=197$ | 1.3436 | 2.5544 | 3.6655 | 1.3521 | 0.7599 | 0.8193 |
| PKPCA_RBF             | $\sigma_0=5, d=150$ | 2.6854 | 5.1821 | 4.2039 | 11.0966 | 0.7825 | 0.7995 |
|                       | $\sigma_0=10, d=150$ | 2.1897 | 3.5209 | 4.6794 | 10.8052 | 0.7060 | 0.7552 |
|                       | $\sigma_0=50, d=150$ | 4.1779 | 4.3620 | 6.4637 | 10.7865 | −0.1896 | −0.3566 |
The predicted tool wear values have been squeezed into a more significant 95% CI, and most test points fall into the 95% CI. It intuitively reflects the effectiveness of PKPCA_IRBF in ameliorating the CI of the WMKRVM model.

5 Conclusion

This paper develops a novel tool wear predicting method based on the proposed PKPCA_IRBF technique and the WMKRVM model to monitor cutting tools’ in-process tool wear width. The major work is summarized accurately:

1. The proposed WMKRVM model outperforms the standard single kernel RVM, including Gaussian kernel, polynomial kernel, multiquadric kernel, exponential kernel, and Laplace kernel in predictive accuracy due to the superior performance of multi-kernel in capturing the global structure of the nonlinear dataset.

2. As a new optimization algorithm, SSA can effectively optimize the kernel parameters of the WMKRVM model, obtaining excellent predictive performance and robustness.

3. A new parameter optimization method, which uses a k-fold cross-validation method combined with HHO, is proposed to optimize the WMKRVM model’s initial hyper-parameters.

4. As a new nonlinear dimension-increment technique, the proposed PKPCA_IRBF is conducive to ameliorating the 95% CI of the WMKRVM model.

Cutting experiments are carried out to reveal the effectiveness of the tool wear predictive model based on the proposed WMKRVM model and the PKPCA technique. This research provides theoretical guidance for monitoring tool wear in the machining process.

6 Authorship contribution

All authors have contributed to the research conception and design. Experimental design, data collection, and algorithmic construction were carried out by Guohao Song, Jianhua Zhang, Yingshang Ge, Kangyi Zhu, Zhensheng Fu, and Luchuan Yu. The first draft was written by Guohao Song. All authors commented on previous versions of the manuscript. The final draft was read and approved by all authors.

Funding This research was supported by the Key R & D project of Shandong Province (No. 2019JZZY010445).

Declarations

Competing interests The authors declare no competing interests.

References

1. Dimla DE (2000) Sensor signals for tool-wear monitoring in metal cutting operations—a review of methods. Int J Mach Tools Manuf 40:1073–1098. https://doi.org/10.1016/S0890-6955(99)00122-4
2. Rehborn AG, Jiang J, Orban PE (2005) State-of-the-art methods and results in tool condition monitoring: a review. The International Journal of Advanced Manufacturing Technology 26:693–710. https://doi.org/10.1007/s00170-004-2038-2
3. Teti R, Jemielniak K, Donnell OG, Dornfeld D (2010) Advanced monitoring of machining operations. CIRP Ann 59:717–739. https://doi.org/10.1016/j.cirp.2010.05.010
4. Siddhpura A, Paurobally R (2013) A review of flank wear prediction methods for tool condition monitoring in a turning process. Int J Adv Manuf Technol 65:571–593
5. Liu T, Zhu K (2021) A switching hidden semi-Markov model for degradation process and its application to time-varying tool wear monitoring. IEEE Trans Industr Inf 17:2621–2631. https://doi.org/10.1109/TII.2020.3004445
6. Han D, Yu J, Tang D (2021) An HDP-HMM based approach for tool wear estimation and tool life prediction. Qual Eng 33:208–220. https://doi.org/10.1080/08982112.2020.1813760
7. Huang C, Yin X, Huang H, Li Y (2020) An enhanced deep learning-based fusion prognostic method for RUL prediction. IEEE Trans Reliab 69:1097–1109. https://doi.org/10.1109/TR.2019.2948705
8. Sun H, Zhang J, Mo R, Zhang X (2020) In-process tool condition forecasting based on a deep learning method. Robotics and Computer-Integrated Manufacturing 64:101924. https://doi.org/10.1016/j.rcim.2019.101924
9. Xu L, Huang C, Li C, Wang J, Liu H, Wang X (2021) Estimation of tool wear and optimization of cutting parameters based on novel ANFIS-PSO method toward intelligent machining. J Intell Manuf 32:77–90. https://doi.org/10.1007/s10845-020-01559-0
10. Shi C, Panoutsos G, Luo B, Liu H, Li B, Lin X (2019) Using multiple-feature-spaces-based deep learning for tool condition monitoring in ultraprecision manufacturing. IEEE Trans Industr Electron 66:3794–3803. https://doi.org/10.1109/TIE.2018.2856193
11. Wang G, Zhang Y, Liu C, Xie Q, Xu Y (2019) A new tool wear monitoring method based on multi-scale PCA. J Intell Manuf 30:113–122. https://doi.org/10.1007/s10845-016-1235-9
12. Huang Z, Zhu J, Lei J, Li X, Tian F (2020) Tool wear predicting based on multi-domain feature fusion by deep convolutional neural network in milling operations. J Intell Manuf 31:953–966. https://doi.org/10.1007/s10845-019-01488-7
13. Shen Y, Yang F, Habibullah MS, Ahmed J, Das AK, Zhou Y, Ho CL (2021) Predicting tool wear size across multi-cutting conditions using advanced machine learning techniques. J Intell Manuf 32:1753–1766. https://doi.org/10.1007/s10845-020-01625-7
14. Tipping ME (2001) Sparse Bayesian learning and the relevance vector machine. J Mach Learn Res 1:211–244. https://doi.org/10.11615/1324430152748236
15. Wang G, Yang Y, Xie Q, Zhang Y (2014) Force based tool wear monitoring system for milling process based on relevance vector machine. Adv Eng Softw 71:46–51. https://doi.org/10.1016/j.advengsoft.2014.02.002
16. Kong D, Chen Y, Li N, Duan C, Lu L, Chen D (2019) Relevance vector machine for tool wear prediction. Mech Syst Signal Process 127:573–594. https://doi.org/10.1016/j.ymssp.2019.03.023
17. Chen F, Cheng M, Tang B, Xiao W, Chen B, Shi X (2020) A novel optimized multi-kernel relevance vector machine with selected sensitive features and its application in early fault diagnosis for rolling bearings. Measurement 156:107583. https://doi.org/10.1016/j.measurement.2020.107583
18. Xue J, Shen B (2020) A novel swarm intelligence optimization approach: sparrow search algorithm. Systems Science & Control Engineering 8:22–34. https://doi.org/10.1080/21642583.2019.1708830
19. Wang J, Xie J, Zhao R, Zhang L, Duan L (2017) Multisensory fusion based virtual tool wear sensing for ubiquitous manufacturing. Robotics and Computer-Integrated Manufacturing 45:47–58. https://doi.org/10.1016/j.rcim.2016.05.010
20. Ge Z, Song Z (2010) Kernel generalization of PPCA for nonlinear probabilistic monitoring. Ind Eng Chem Res 49:11832–11836. https://doi.org/10.1021/ie100852s
21. Kim D, Lee I (2003) Process monitoring based on probabilistic PCA. Chemom Intell Lab Syst 67:109–123. https://doi.org/10.1016/S0169-7439(03)00063-7
22. Neil DL (2005) Probabilistic non-linear principal component analysis with Gaussian process latent variable models. J Mach Learn Res 6
23. Tanaka A, Imai H, Kudo M, Miyakoshi M (2007) Integrated kernels and their properties. Pattern Recogn 40:2930–2938. https://doi.org/10.1016/j.patcog.2007.02.014
24. Zhang Z, Wang G, Yeung DY, Kwok JT (2004) Probabilistic kernel principal component analysis
25. Rashedi E, Nezamabadi-pour H, Saryazdi S (2009) GSA: a gravitational search algorithm. Inf Sci 179:2232–2248. https://doi.org/10.1016/j.ins.2009.03.004
26. Yang XS (2010) Engineering optimization (an introduction with metaheuristic applications) || Engineering Optimization
27. Jamil M, Yang X (2013) A literature survey of benchmark functions for global optimization problems[Z]. Cornell University Library, arXiv.org, Ithaca
28. Yang XS (2014) [Studies in Computational Intelligence] Cuckoo Search and Firefly Algorithm Volume 516 || Intelligent Firefly Algorithm for Global Optimization. In:315–330
29. Sm A, Smm B, Al A (2014) Grey wolf optimizer. Advances in Engineering Software:46–61
30. Eberhart, SY (2002) Particle swarm optimization: developments, applications and resources: Congress on Evolutionary Computation
31. Heidari AA, Mirjalili S, Faris H, Aljarah I, Mafarja M, Chen H (2019) Harris hawks optimization: algorithm and applications. Futur Gener Comput Syst 97:849–872. https://doi.org/10.1016/j.future.2019.02.028

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

Springer Nature or its licensor holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.