



# A new intelligent fault diagnosis method for bearing in different speeds based on the FDAF-score algorithm, binary particle swarm optimization, and support vector machine

Saeed Nezamivand Chegini<sup>1</sup> · Ahmad Bagheri<sup>1</sup> · Farid Najafi<sup>1</sup>

Published online: 11 November 2019  
© Springer-Verlag GmbH Germany, part of Springer Nature 2019

## Abstract

In this paper, a new hybrid intelligent technique is presented based on the improvement in the feature selection method for multi-fault classification. The bearing conditions used in this study include healthy condition, defective inner ring, defective outer ring, and the faulty rolling element at different rotating motor speeds. To form the feature matrix, at first, the vibration signals are decomposed using empirical mode decomposition and wavelet packet decomposition. Then, the time and frequency domain features are extracted from the raw signals and the components are obtained from the signal decomposition. The high-dimensional feature matrix leads to increasing the computational complexity and reducing the efficiency in the classification accuracy of faults. Therefore, in the first stage of the feature selection process, the redundant and unnecessary features are eliminated by the FDAF-score feature selection method and the preselected feature set is formed. The FDAF-score technique is a combination of both F-score and Fisher discriminate analysis (FDA) algorithms. Since there may exist the features that are not susceptible to the presence of faults, the binary particle swarm optimization (BPSO) algorithm and the support vector machine (SVM) are used to select the optimal features from the preselected features. The BPSO algorithm is used to determine the optimal feature set and SVM classifier parameters so that the predictive error of the bearing conditions and the number of selected features are minimized. The results obtained in this paper demonstrate that the selected features are able to differentiate the different bearing conditions at various speeds. Comparing the results of this article with other fault detection methods indicates the ability of the proposed method.

**Keywords** Bearing fault detection · Feature extraction · FDAF-score method · Binary particle swarm optimization · Support vector machine

## 1 Introduction

The fault detection in rotating machinery is one of the most important topics in the field of fault detection. On the other hand, bearings are one of the vital components of these machines, so that the fault of this component may result in full machine breakdown. Therefore, detecting the presence of defects and their characteristics in bearings plays an important role in the maintenance of machinery.

The analysis of vibration signals is one of the common points of view for the bearing fault diagnosis which has been particularly interested in researches. The vibration signals acquired from rotating machinery are nonlinear and non-stationary. The conventional methods in time and frequency domain are not able to extract useful information from these data. In recent years, the advanced intelligent methods have been developed in the time–frequency domain for extracting the information of defects.

The vibration data acquired from the rotating machineries have the high complexity as usual, and many researchers and engineers have used the concepts such as artificial intelligence, machine learning, and signal processing techniques to detect the occurrence of the fault of machine. Intelligent fault diagnosis methods for the rotating machinery include four stages: preprocessing, signal processing and feature extraction, post-processing, and

---

Communicated by V. Loia.

✉ Saeed Nezamivand Chegini  
saeed.nezamivand@gmail.com

<sup>1</sup> Faculty of Mechanical Engineering, University of Guilan, Rasht, Iran

pattern recognition. At the feature extraction stage, the statistical time and frequency features and the signal processing techniques are used to extract the fault-sensitive features in rotary machines. Among the signal processing methods, empirical mode decomposition (EMD) (Huang et al. 1998) and its new versions, wavelet transform (WT) and wavelet packet decomposition (WPD) (Mallat 1989), empirical wavelet transform (EWT) method (Gilles 2013), and intrinsic time-scale decomposition (ITD) (Frei and Osorio 2007) are widely used in fault detection. The EMD and ITD methods are adaptive techniques that are able to decompose each complex and multi-component signal into a collection of complete or almost mono-components with physical meanings by considering their local characteristics such as local maximum and minimum points. The wavelet transform can provide the excellence local analysis in the time–frequency/scale domain for the non-stationary signals. Since defects often appear as impulses that cover a wide range of frequencies, the WT assigns the large coefficients to such impulses. WPD is the extended version of WT that decomposes both the approximation and detail coefficients, and this fact causes to extract and retain very useful information contained in high-frequency components. In other words, WPD is a very useful tool that has been widely applied for processing the non-stationary vibration signals. This technique exploits the meaningful properties and has a good performance for analyzing both high and low frequencies.

In the smart fault detection methods, an intelligent classification system is required to automatically diagnose and classify the actual state of the rotating machine components. To this end, many modern machine learning tools such as artificial neural network (ANN) (Sperduti and Starita 1997), support vector machine (SVM) (Vapnik 1995) and adaptive neural fuzzy inference system (ANFIS) (Jang 1993) have been used. Since the theoretical basis of the neural network and other conventional artificial intelligent techniques is based on empirical risk minimization (ERM) principle, these techniques have limitation on generalization of results in models that can over-fit the samples. The SVM is a computational supervised learning method based on statistical learning theory, and to be implemented is based on the structural risk minimization (SRM) principle so that the training accuracy and the capacity of the classifier model are considered. The SVM has properties such as high accuracy and good generalization for smaller number of fault samples. This method has been extensively studied by researches for recognizing the rotary machine faults. The development of each of the intelligent fault diagnosis steps is the subject of many researches. Some of these researches will be described in the following.

Lei and Zuo (2009) decomposed the vibration signals by ensemble EMD (EEMD) method, and then, they improved the Hilbert–Huang transform based on selecting the sensitive IMFs for extracting the fault features. In papers such as Jiang et al. (2013), Lei et al. (2017), Li et al. (2015), Xue et al. (2015), and Guo and Deng (2017), researchers developed the bearings and gears fault detection method based on improving the implementation steps of the EMD and EEMD techniques. Nguyen et al. (2015) presented a new robust fault detection technique for rolling bearings. They suggested a new approach based on de-noising using EMD method, Naive Bayes classifier, and thresholding of the noisy components. Then, the authors presented a technique for identifying the bearing condition by utilizing WPD, Hilbert transform and the envelope spectra of the signal. Nezamivand Chegini et al. (2019) introduced a new vibration signal de-noising strategy based on EWT, the kurtosis factor, and the envelope spectrum analysis for the bearing fault diagnosis. Tabrizi et al. (2015) introduced a method for early fault detection in bearings. They de-noised the signals using WPD and then extracted the meaningful feature vectors by applying the EEMD method. Finally, they used the SVM for the bearing fault diagnosis. In Zhao et al. (2016), in order to improve the de-noising methods by means of EMD, authors inspired by the noise cleaning methods in wavelet transform. In this work, two concepts of the kurtosis and the cross-correlation are used to select the noisy and meaningful IMFs. Wang et al. (2017) proposed a self-adaptive filter using the EEMD method in order to eliminate the noise from vibration signals acquired from a locomotive damaged bearing. For this purpose, an adaptive relationship is suggested for computing the number of the sifting process based on the number of the signal IMFs. The results of their work demonstrate that the fault characteristics can easily be seen in the frequency spectrum of the de-noised signal. Bordoloi and Tiwari (2014a, b, 2015) introduced a multi-classifier method to classify the multiple faults in gears and bearings. They optimized the supporting vector machine parameters by genetic algorithm (GA) and artificial bee colony (ABC) algorithm before the learning and final testing SVM. In this work, statistical functions such as kurtosis, standard deviation, and skewness are used as the SVM inputs. Ben Ali et al. (2015) extracted some features from nonlinear signals using the EMD method. Then, they used these features for training the ANN and classifying the bearing faults. Jedlinski and Jonak (2015) used a method for early fault detection in the gearbox. They decomposed vibration signals into wavelet coefficients by means of continuous wavelet transform (CWT). Then, scales and coefficients of large values are selected as inputs of the ANN and SVM. Fu et al. (2019) introduced a novel bearing fault diagnosis technique based on blind parameter identification of MAR

model and mutation hybrid gray wolf optimization (GWO)–sine cosine algorithm (SCA) optimized SVM. The results of experimental data set indicate the superiority of their proposed approach.

In other studies, in order to increase the accuracy of the fault diagnosis, various features such as different levels of wavelet packet transform, time and frequency domain features of the components derived from EMD and EEMD methods were used. However, with increasing the number of attributes, the feature vector not only includes the useful features but also contains the insensitive and redundant features. Increasing the dimension of the feature matrix leads to the computational complexity and the reduction in the classification accuracy. In such cases, in order to solve the dimensionality feature space problem, the feature selection methods were used in different researches. Feature selection is one of the matters that is applied in various fields such as fault detection (Chen and Chen 2015), machine learning (Banka and Dara 2015), and data mining (Bhuyan and Kamila 2015). The feature selection methods are grouped into three categories: the filter method, the wrapper method, and the hybrid method. In the filter methods, the weight of each attribute regardless of the classifier is calculated using a criterion function and the features with the highest weight are chosen as effective features, and the rest are eliminated. The wrapper methods select the features as the optimal feature set that can obtain the most accuracy for the classifiers. The wrapper method is superior to the filter method, but its computational cost is higher. The hybrid method is a combination of both the filter and the wrapper methods and has the advantages of both methods (Zeng et al. 2015). Fatima et al. (2015) used the SVM to study the fault classification in bearings at five different speeds. Firstly, they extracted twelve statistical features in the time domain and then used the compensation distance evaluation technique (CDET) method for identifying the most informative features as inputs of the multi-class SVM. Wei et al. (2017) proposed a new signal processing for bearing fault detection. They extracted the time and frequency statistical features from the vibration signal using the WPD and EEMD methods. Then, the authors presented a novel optimal feature selection method based on the adaptive feature selection technique and affinity propagation clustering method. Yan and Jia (2018) suggested a new multi-fault classification technique for bearing fault diagnosis in different working conditions. In this study, the multi-domain feature is extracting using three strategies: statistical characteristics, fast Fourier transform (FFT) and variational mode decomposition (VMD). Then, informative and sensitive features are selected using the Laplacian feature selection method. Finally, the PSO-SVM method is implemented for the identification of rolling bearing conditions. Vakharia et al.

(2016) designed a technique for bearing fault detection in four conditions: healthy bearing, defected ball, defected inner race, and defected outer race. In that work, the feature vector of each signal includes features such as kurtosis, skewness, mean, root mean square, and Shannon entropy. Then, most sensitive features are extracted using filter methods such as chi-square and Relief-F methods. In Ziani et al. (2017a), a new optimal feature selection method was presented for bearing fault detection with various types of faults. In this reference, most effectiveness features were chosen using the BPSO algorithm. In this study, regularized Fisher's criterion (RFC) was used as a fitness function in order to increase the performance of the classification accuracy. Ziani et al. (2017b) investigated a new feature selection scheme for the gearbox fault diagnosis. In this work, statistical characteristics, spectral features, and coefficients of the WPD and EMD are used for extracting the feature vector of each signal. In the next step, the most important features are selected according to three different algorithms such as the Fisher score, correlation criterion, and signal-to-noise criterion. Then, the Pareto method was used to determine the sensitive features. The results of this study demonstrated that Pareto–Fisher with SVM classifier leads to high-performance accuracy. Attoui et al. (2017) introduced a new procedure for identifying bearing fault conditions such as the damaged ball, inner race, and outer race at different speeds. They extracted fault features using WPD and short-time Fourier transform (STFT) methods. In this work, linear discriminant analysis (LDA) and locality sensitive discriminant analysis (LSDA) were used as feature dimensionality reduction techniques and ANFIS was utilized as a classification system. Zhang et al. (2015) suggested a pattern recognition technique based on synchronous feature selection and SVM parameters optimization using ant colony algorithm (ACA). This approach was utilized for identifying the rotating machinery conditions. Lu et al. (2015) presented a new strategy based on the GA, EMD, and SVM in rotary mechanical fault diagnosis. They employed a modified GA with a dynamic searching strategy for selecting the most representative features. Shan et al. (2019) suggested a rotating machinery fault diagnosis method based on improved VMD (IVMD) and hybrid artificial sheep algorithm (HASA). The authors utilized IVMD to decompose the vibration data and extract the feature set. Finally, the HASA is applied to select the optimal features and optimize the SVM parameters.

In this paper, a new method is presented for the bearing fault diagnosis with different faults at various motor speeds based on selecting the optimal feature set and classifying the faults. Each signal is decomposed into simpler oscillating components by means of the WPD and EMD methods. To form the feature vector of each raw signal, time and frequency statistical features related to the raw

signal, IMFs derived from EMD and the wavelet coefficients derived from WPD have been extracted. Increasing the number of features in the feature matrix leads to produce the unrelated information to bearing conditions. The optimal feature selection procedure and fault detection method presented in this paper contain two stages: The first stage involves eliminating the redundant and inefficient features that are carried out using a new method called FDAF-score (Song et al. 2017). In fact, in this step the FDAF-score method is applied for selecting the useful features from the original feature set with high dimension. In the second step, the BPSO algorithm (Kennedy and Eberhart 1997) is used to determine the most appropriate features of the preselected feature set and identify the optimal SVM parameters, simultaneously. In fact, the proposed feature selection method is a combination of the wrapper and filter methods. In the optimization process with BPSO, the optimal feature sets and the SVM parameters have been obtained so that the prediction error of the bearing conditions and the number of optimal features are minimized. Finally, the proposed method is compared with other techniques presented in recent years. The results show that the technique presented in this paper has a good capability for detecting the bearing defects.

The rest of this paper is organized as follows: EMD method, feature extraction, FDAF-score method, SVM classifier, and BPSO algorithm are discussed in Sect. 2. In Sect. 3, the feature selection scheme and the proposed fault detection method are presented. In Sect. 4, case studies are introduced. The results of the proposed method are presented in Sect. 5. Finally, the paper is concluded in Sect. 6.

## 2 Methods

### 2.1 Empirical mode decomposition

The acquired vibration signals from the rotating machinery are always non-stationary, complex, and nonlinear. Therefore, in order to extract the useful information related to defects, it is necessary to utilize an appropriate signal processing method. The EMD method decomposes a complex signal based on its local behavior into simple oscillating modes that are called intrinsic mode functions (IMFs). An IMF has the two following features (Huang et al. 1998):

1. In total data sets, the number of maximum and minimum points and the number of zero-crossings are equal or differ by one.
2. In any data location, the mean value of the upper envelope determined with the local maximum and

lower envelope determined by the local minimum is zero.

For given signal  $x(t)$ , the signal decomposition process in the EMD method contains the following steps (Huang et al. 1998):

1. Determine all local maxima and local minima of signal  $x(t)$ .
2. Calculate the upper envelope curve by connecting the local maxima by cubic spline lines. Repeat the same process for the local minima to obtain the lower envelope curve.
3. Compute the mean value of the upper envelope and the lower envelope. This parameter is denoted as  $m_1(t)$ .
4. Identify the difference between signals  $x(t)$  and  $m_1(t)$ , denoted as  $h_1(t)$ :

$$x(t) - m_1(t) = h_1(t) \quad (1)$$

5. If  $h_1(t)$  has two conditions of an IMF, it is considered as first IMF. Otherwise,  $h_1(t)$  is considered instead of the original signal and steps 1–4, called sifting process, are repeated. After  $k$  times repetition of the sifting process, the first IMF  $c_1(t) = h_{1k}(t)$  is obtained. In this study, Cauchy-type convergence is considered as the stopping criterion in the sifting process:

$$D_k = \frac{\sum_{t=0}^T |h_1^{k-1}(t) - h_1^k(t)|^2}{\sum_{t=0}^T |h_1^{k-1}(t)|^2} \leq SD \quad (2)$$

where  $SD$  is considered in interval  $[0.2, 0.3]$ .

6. By subtracting  $c_1(t)$  from the original signal  $x(t)$ , the residue  $r_1(t)$  is obtained:

$$r_1(t) = x(t) - c_1(t) \quad (3)$$

7.  $r_1(t)$  is considered as new time series and by repeating steps 1–6, it can be calculated the intrinsic mode functions  $c_1, c_2, \dots, c_n$ .
8. When residue  $r_n(t)$  becomes a monotonic function, in this case, no more IMF can be extracted and decomposition procedure is stopped.

Finally, signal  $x(t)$  can be expressed as follows:

$$x(t) = \sum_{i=1}^n c_i + r_n. \quad (4)$$

### 2.2 Feature extraction

The bearings have nonlinear dynamics behavior, and the acquired vibrational signals are always non-stationary. If

**Table 1** Time-domain and frequency-domain features (Ben Ali et al. 2015)

| <i>Time-domain features</i>   |  |
|---|--|
| Mean = $\frac{1}{N} \sum_{i=1}^N x(i)$  | Standard deviation = $\sqrt{\frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2}$   |
| Maximum = $\max(x(i))$  | Skewness = $\frac{\frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^3}{(\frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2)^{\frac{3}{2}}}$ |
| Minimum = $\min(x(i))$  | Kurtosis = $\frac{\frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^4}{(\frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2)^2}$             |
| Median = $x(\frac{N+1}{2})$   | Crest factor = $\frac{\max(abs(x))}{RMS}$  |
| Energy = $\int  x(t) ^2 dt$   | Shape factor = $\frac{x_{RMS}}{(\frac{1}{N}) \sum_{i=1}^N  x_i }$  |
| Entropy = $-\sum_{i=1}^N x_i^2 \log(x_i^2)$   | Impulse factor = $\frac{Peak}{(\frac{1}{N}) \sum_{i=1}^N  x_i }$   |
| Range = $\max(x_i) - \min(x_i)$   | Clearance factor = $\frac{Peak}{((\frac{1}{N}) \sum_{i=1}^N \sqrt{ x_i })^2}$  |
| Root mean square = $\sqrt{\frac{1}{N} \sum_{i=1}^N x_i^2}$  | Hybrid feature 1 = $\log(Kurtosis + \frac{RMS}{0.078})$  |
| Peak value = $\max( x_i )$  | Hybrid feature 2 = $\log(Kurtosis^{Crestfactor} + (\frac{RMS}{0.078})^{Range})$  |
| Mean peak = $\text{mean}(\text{total peak})$  | Tigear – Kaiser energy = $\sum_{i=1}^N ([x(i)]^2 - x(i-1)x(i+1))$  |
| In these features, $x(n)$ denotes a vibrational signal and $n = 1, 2, 3, \dots, N$                            |  |
| <i>Frequency domain features</i>  |  |
| Mean frequency = $\frac{1}{K} \sum_{k=1}^K s(k)$  | Standard deviation frequency = $\sqrt{\frac{\sum_{k=1}^K f_k^2 s(k)}{\sum_{k=1}^K s(k)}}$                                  |
| Frequency center = $\frac{\sum_{k=1}^K f_k s(k)}{\sum_{k=1}^K s(k)}$  | Root - mean - square frequency = $\sqrt{\frac{\sum_{k=1}^K (f_k - F_{15})^2 s(k)}{\sum_{k=1}^K s(k)}}$                     |
| $s(k)$ : spectrum, $K$ : number of spectrum lines, $f_k$ : frequency value of the of the $k$ th spectrum line |  |

only vibration signals are investigated, it cannot be obtained much useful information about the type and size of the defects. As a result, the different feature extraction methods will be used here. In this paper, the following procedure is employed for forming the feature vector corresponding to each signal:

1. From each raw signal, the time-domain statistical features presented in Table 1 are extracted (features  $FV_1$ – $FV_{21}$ ).
2. Each vibration signal is decomposed into different modes using the EMD algorithm. Since the first few components include more information about defects than the rest of the IMFs (Tabrizi et al. 2015), in this study, the first five IMFs are chosen to extract the time-domain features (features  $FV_{22}$ – $FV_{126}$ ).
3. From each raw signal, the frequency-domain statistical parameters presented in Table 1 are extracted (features  $FV_{127}$ – $FV_{130}$ ). Also, these features are calculated for each IMF (features  $FV_{131}$ – $FV_{150}$ ).
4. The WPD is one of the suitable signal processing methods and also the feature extraction techniques in

the rotating machinery fault diagnosis. This method extracts the useful information from non-stationary signals in the high and low frequencies (Li et al. 2013). One of the most effective and important parameters in the WPD is the mother wavelet function. In this study, similar to (Ziani et al. 2017a), the db44 wavelet is used. The maximum depth for the decomposition tree is equal to 3. Therefore, for each signal, fourteen wavelet coefficients will be obtained. Finally, for each coefficient, the time-domain statistical parameters according to Table 1 are calculated (features  $FV_{151}$ – $FV_{444}$ ).

After extracting the above features, for each signal sample, a feature vector will be obtained. The elements of the feature vector are shown in Fig. 1.

The above feature vector is calculated for all signals, and the new data set is obtained in the general form as follows:

$$FM = \{f_{m,j,i}\}, m = 1, 2, \dots, M; i = 1, 2, \dots, C; j = 1, 2, \dots, J. \tag{5}$$

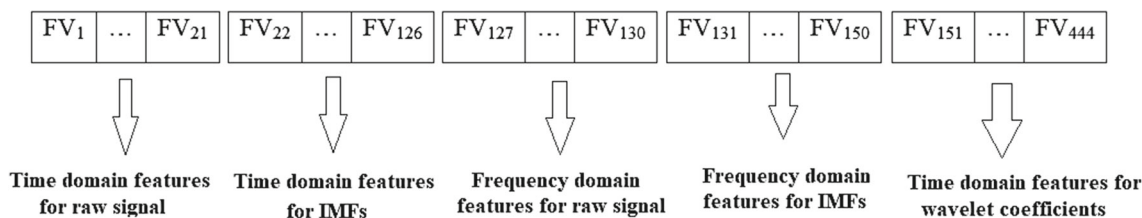


Fig. 1 Schematic of feature vector elements

In this matrix,  $f_{m,j,i}$  is the  $j$ th feature corresponding to the  $m$ th signal under the  $i$ th condition.  $M_i$  is the number of signals under condition  $i$ ,  $J$  is the number of extracted features for each signal, and  $C$  is the number of conditions. In this paper,  $J = 444$  and  $C$  is equal to number of fault classes.

### 2.3 Feature selection using FDAF-score method

Consider the data set  $\{(x_n, y_n)\}_{n=1}^N \in X \times Y$  where  $X \in R^M$  is the feature space with  $M$  dimensionality and  $Y = \{1, 2, \dots, C\}$  is the labels of these data. If the space dimensionality of the features extracted from the training data is very large, then feature matrix not only includes the fault-sensitive features, but also the non-sensitive features will appear in this matrix. In 2017, a new feature selection approach has been proposed that is called as FDAF-score (Song et al. 2017). This technique is the combination of two methods: F-score and Fisher discriminant analysis. The F-score feature selection method is applicable only for the data samples with two classes ( $c = 1, 2$ ). In this technique, the following criteria are used to select the effective features (Guyon et al. 2002):

$$s_i = \frac{(\bar{X}_i^1 - \bar{X}_i)^2 + (\bar{X}_i^2 - \bar{X}_i)^2}{\frac{1}{n_1-1} \sum_{j=1}^{n_1} (x_{i,j}^1 - \bar{X}_i^1)^2 + \frac{1}{n_2-1} \sum_{j=1}^{n_2} (x_{i,j}^2 - \bar{X}_i^2)^2}, \quad (6)$$

where  $s_i$  is the  $i$ th feature,  $\bar{X}_i$ ,  $\bar{X}_i^1$ , and  $\bar{X}_i^2$  are the averages of the  $i$ th feature for all classes, class 1, and class 2, respectively.  $X_{j,i}^1$ , and  $X_{j,i}^2$  are the  $i$ th feature of the  $j$ th sample of class 1 and class 2, respectively.

In Song et al. (2017), researchers have used Fisher discriminant analysis to develop the F-score method from two classes to multi-class. In FDAF-score, two concepts « average between-class distance » and « within-class scatter » are used for evaluating each feature. The average between-class distance is calculated for the  $k$ th feature, i.e.,  $x_k$  as follows:

$$D(x_k) = \sum_{1 \leq j < i \leq C} \left( \frac{n_j + n_i}{N} \right) (x_j^k - x_i^k)^2, \quad (7)$$

where  $N$  is the number of samples, and  $i$  and  $j$  are class labels.  $n_i$  and  $n_j$  are the number of observations in  $i$ th and

$j$ th classes, respectively.  $\bar{X}_i^k$  and  $\bar{X}_j^k$  are the averages of classes  $i$  and  $j$  for the  $k$ th feature.

The within-class scatter is defined as follows:

$$S(x_k)_j = \frac{\frac{1}{n_j} \sum_{l=1}^{n_j} \left( (x_j^k)_l - \bar{x}_j^k \right)^2 - \min_{1 \leq l \leq n_j} \left( (x_j^k)_l - \bar{x}_j^k \right)^2}{\max_{1 \leq l \leq n_j} \left( (x_j^k)_l - \bar{x}_j^k \right)^2 - \min_{1 \leq l \leq n_j} \left( (x_j^k)_l - \bar{x}_j^k \right)^2}. \quad (8)$$

In this equation,  $j$  is the class type and  $n_j$  indicate the number of samples in the  $j$ th class.  $(x_j^k)_l$  is the  $l$ th sample of the  $j$ th class corresponding to the  $k$ th feature.

Similar to FDA technique, in FDAF-score approach, the following criterion is used to evaluate the  $k$ th feature:

$$J(x_k)_j = \frac{\sum_{j=1}^C D(x_k)_j}{\sum_{j=1}^C S(x_k)_j}. \quad (9)$$

The  $J(x_k)$  value represents the degree of the correlation coefficient of the  $k$ th feature with classes 1 to  $C$ . The large  $J(x_k)$  values for feature  $x_k$  indicate that this feature has good capability to separate the different classes.

The FDAF-score feature selection process is summarized as follows:

1. Preset the value of the threshold  $\lambda$  and the maximum classification accuracy  $p$ .
2. Compute the expressions  $D(x_k)$  and  $s(x_k)_j$  using Eqs. (7) and (8) for all features.
3. Compute the ranking index of the features, i.e.,  $J(x_k)$ , by means of Eq. (9) for all the attributes.
4. If  $J(x_k) > \lambda$ , then the  $k$ th feature is appropriate, otherwise eliminate it.
5. Construct the SVM classifier using the selected feature subset in step 4. If the classification accuracy reaches the desired value of  $P$ , then pick out the chosen feature as the optimal feature set. Otherwise, change the value of  $\lambda$  and return to step 4.
6. Repeat the above steps to obtain the prediction accuracy of  $P$  and optimal feature set.

The flowchart of the FDAF-score feature selection method is presented in Fig. 2.

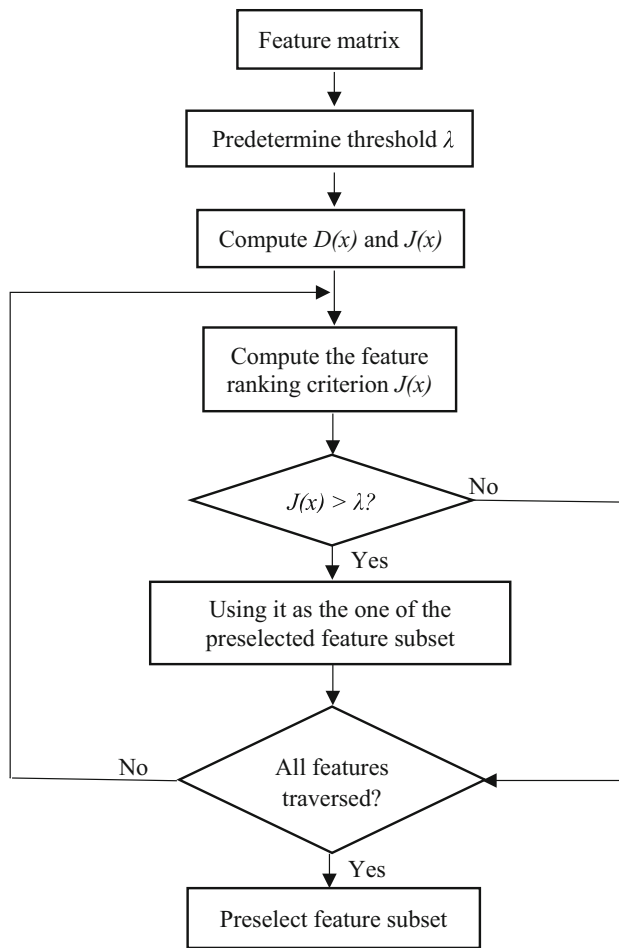


Fig. 2 Flowchart of the FDAF-score feature selection method (Song et al. 2017)

Since the classification accuracy is considered as a decisive parameter during their feature selection process using FDAF-score, the capability of this method is understandable. The adjusting screw of this method is the threshold parameter  $\lambda$ , which can be changed according to the predetermined accuracy  $P$ .

### 2.4 Support vector machine

Consider data set  $\{x_i, y_i\}_{i=1}^N$  labeled as two classes: positive ( $y_i = +1$ ) and negative ( $y_i = -1$ ). Suppose that these data are separable using hyperplane  $w\mathbf{x} + b = 0$ . The parameters  $w$  and  $b$  are normal vector and scalar parameter, respectively. The training data set satisfies the following constraints (Vapnik 1995):

$$w\mathbf{x} + b \geq +1 \quad \text{if } y_i = +1 \tag{10}$$

$$w\mathbf{x} + b \leq -1 \quad \text{if } y_i = -1. \tag{11}$$

Here, the goal is to find the parameters  $b$  and  $w$  so that the hyperplane maximizes the margin between two planes

$w\mathbf{x} + b = +1$  and  $w\mathbf{x} + b = -1$ . These parameters can be achieved by minimizing the expression  $\frac{1}{2}w^2$ . In actual cases, the data are not linearly segregated. In these situations, the optimal hyperplane is determined by solving the following problem:

$$\min \left( \frac{1}{2}w^2 + C \sum_{i=1}^N \xi_i \right) \text{ subject to } \begin{cases} y_i(w \cdot \mathbf{x} + b) \geq 1 - \xi_i & i = 1, \dots, N, \\ \xi_i \geq 0 \end{cases} \tag{12}$$

where  $\xi_i \geq 0$  and  $C$  are slack variable and penalty parameter. The above problem is converted into the Lagrangian dual problem using the Kuhn–Tucker condition. By introducing the Lagrangian multipliers  $\alpha_i$  and  $\beta_i$  for the problem constraints (12), the following quadratic optimization problem is obtained:

$$\min L(\alpha) = \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i=1}^N \alpha_i \alpha_j y_i y_j \mathbf{x}_i \cdot \mathbf{x}_j \tag{13}$$

subject to :  $\begin{cases} \sum_{i=1}^N \alpha_i y_i & i = 1, \dots, N. \\ 0 \leq \alpha_i \leq C \end{cases}$

The coefficients  $\alpha_i$  are obtained by solving the optimization problem (13). Therefore, the nonlinear decision function for classifying new data can be written as follows:

$$f(\mathbf{x}) = \text{sign} \left( \sum_{i=1}^N \alpha_i y_i (\mathbf{x}_i \cdot \mathbf{x}_j) + b \right). \tag{14}$$

The input space with nonlinear classification can be mapped using a nonlinear vector function ( $\phi(x)$ ) into a high-dimensional feature space, in case the linear classification to be possible. Finally, the decision function is converted to the following form:

$$f(\mathbf{x}) = \text{sign} \left( \sum_{i=1}^N \alpha_i y_i K(\mathbf{x}_i \cdot \mathbf{x}_j) + b \right), \tag{15}$$

where  $K(\mathbf{x}_i \cdot \mathbf{x}_j)$  is the kernel function and defined as follows:

$$K(\mathbf{x}_i \cdot \mathbf{x}_j) = \phi(\mathbf{x}_i) \cdot \phi(\mathbf{x}_j). \tag{16}$$

Since the radial basis kernel function (RBF) is widely employed in the machinery fault detection, so this function is used in this study (Huang and Wang 2006). The mathematical formulation of RBF function is as follows:

$$K(\mathbf{x} \cdot \mathbf{x}_j) = \exp \left( - \frac{\|\mathbf{x} - \mathbf{x}_j\|^2}{\sigma^2} \right). \tag{17}$$

The choice of parameters  $C$  and  $\sigma$  affects the efficiency of the SVM method. It can be seen in this study that the

PSO algorithm will be used to find the optimal values for these two parameters.

Two strategies « one-against-all (OAA) » and « one-against-one (OAO) » are commonly used to construct the multi-class SVM. Hsu and Lin (2002) demonstrated that the SVM-OAO method is more appropriate for practical applications. Therefore, this strategy is used in this work.

### 2.5 Binary particle swarm optimization

The particle swarm optimization algorithm is one of the population-based and the swarm intelligence algorithms that is used in many practical applications. In this algorithm, the particle concept is used to represent the responses of a problem. For a  $d$ -dimensional problem, each particle is to be introduced with two numbers: the velocity and the position vectors. The PSO algorithm consists of two stages: initialization and calculation. In the initialization phase, an initial position and an initial velocity are allocated randomly to each particle. In the calculation phase, each particle uses its personal best experience ( $\vec{X}_{pBest}$ ) and the best solution obtained by all particles ( $\vec{X}_{gBest}$ ) to find its next position and move in the search space as follows (Shi and Eberhart 1998):

$$\vec{V}_i(t + 1) = w\vec{V}_i(t) + c_1r_1(\vec{X}_{pBest_i} - \vec{X}_i(t)) + c_2r_2(\vec{X}_{gBest} - \vec{X}_i(t)) \tag{18}$$

$$\vec{X}_i(t + 1) = \vec{X}_i(t) + \vec{V}_i(t + 1), \tag{19}$$

where  $\vec{V}_i(t + 1)$  and  $\vec{X}_i(t + 1)$  are the velocity and position vectors of  $i$ th particle in  $t + 1$  iteration, respectively.  $r_1$  and  $r_2$  are random variables in the interval  $[0,1]$ . In this paper, the learning factors  $C_1$  and  $C_2$  and inertia weight  $w$  according to Nezamivand Chegini et al. (2018) are calculated as follows:

$$c_1(t) = c_{1min} + \left(\frac{t_{max} - t}{t_{max}}\right)(c_{1max} - c_{1min}) \tag{20}$$

$$c_2(t) = c_{2max} + \left(\frac{t_{max} - t}{t_{max}}\right)(c_{2min} - c_{2max}) \tag{21}$$

$$w(t) = w_f + \left(\frac{1 + \cos\left(\frac{\pi t}{t_{max}}\right)}{2}\right)^k (w_i - w_f), \tag{22}$$

where  $C_{1min}$ ,  $C_{2min}$ , and  $W_{min}$  are the minimum values of  $C_1$ ,  $C_2$ , and  $W$  parameters, respectively.  $C_{1max}$ ,  $C_{2max}$ , and  $W_{max}$  are the maximum values of  $C_1$ ,  $C_2$ , and  $W$  parameters, respectively. These parameters are set according to Nezamivand Chegini et al. (2018).

The BPSO algorithm was introduced by Kennedy and Eberhart (1997) for solving the optimization problems in the discrete binary space. Recently, this technique has been

applied in many fault detection researches. In the BPSO algorithm, each particle is constructed by bits that only consist of ‘0’ or ‘1’ values. In BPSO similar to PSO, the velocity of each particle is calculated by Eq. (18). In this case,  $\vec{X}_{pBest}$  and  $\vec{X}_{gBest}$  are vectors that their elements become ‘0’ or ‘1.’ In this algorithm, the position of each particle is updated in direction  $d$  using the following equation (Ziani et al. 2017a):

$$x_{id}(t + 1) = \begin{cases} 1, & \text{if rand() < } s(v_{id}(t + 1)) \\ 0, & \text{otherwise} \end{cases}, \tag{23}$$

where  $\text{rand}()$  is a random variable in interval  $[0,1]$  and  $s(x)$  is a sigmoid function that is defined as follows:

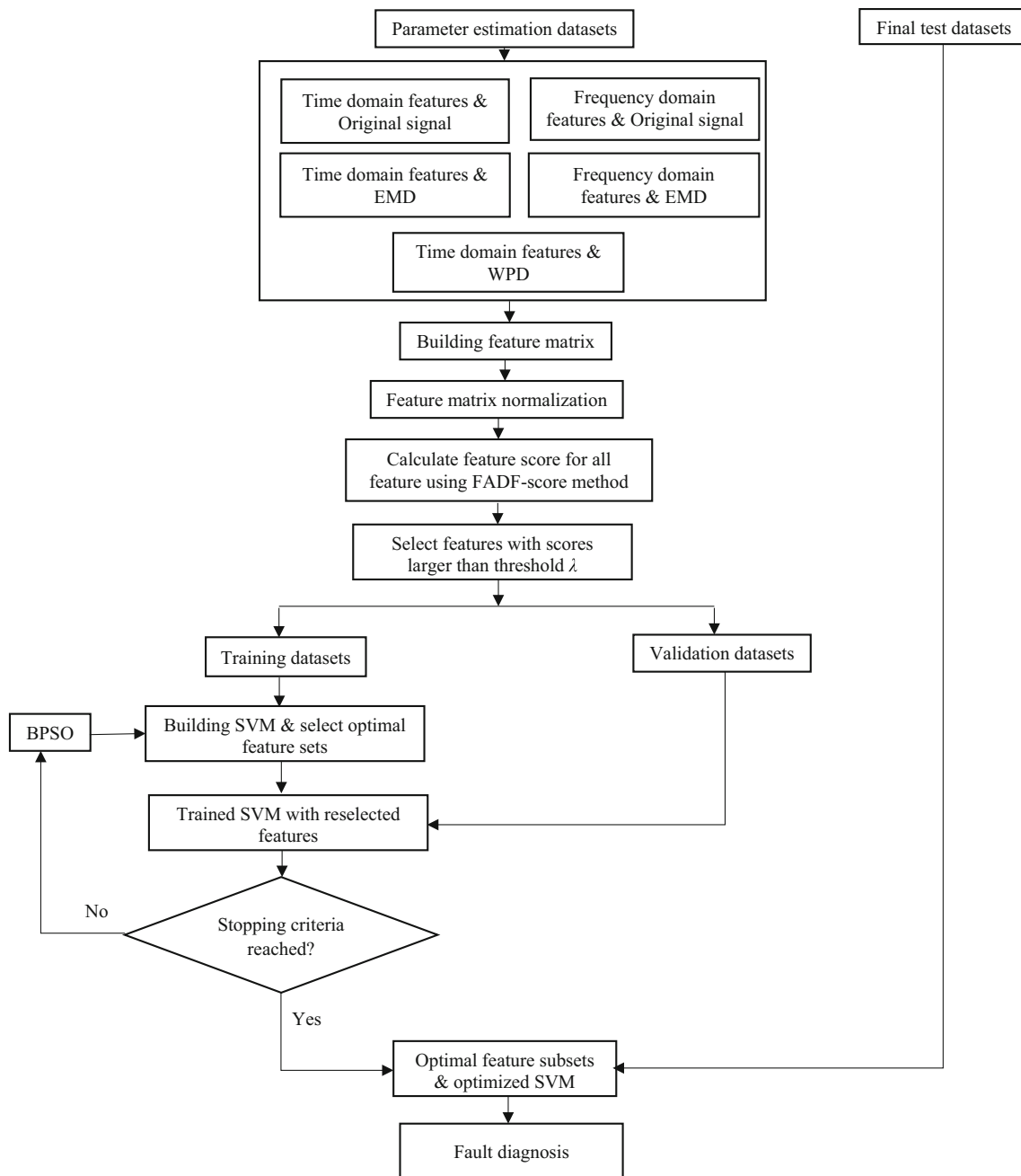
$$s(x) = \frac{1}{1 + e^{-x}}. \tag{24}$$

### 3 Proposed intelligent method

The main parts of the new hybrid intelligent method presented in this paper are: the feature extraction, the feature preselection using the FDAF-score method, and determination of the optimal feature set and the optimal SVM parameters using the BPSO algorithm. The flowchart of this method is shown in Fig. 3. The steps of the proposed algorithm are described below:

1. 60 percent of the signals are assigned to « training data set for parameter estimation », and the rest of the signals are used for « final test data set ».
2. Feature extraction: According to Subsection 2.2, some of the features extracted from each signal sample belong to « training data set for parameter estimation ». Each feature vector corresponding to each signal consists of the time and frequency domain features of the original signal, time and frequency features of the first five IMFs, and time-domain features of the wavelet coefficients. In this step, 444 features are obtained for each vibration data. The feature matrix is formed so that its rows correspond to the signals and columns correspond to the features. Then, the data for each column are normalized so that their values place in the interval  $[0, 1]$ .
3. Feature preselection using the FDAF-score method: The original feature matrix obtained in step 2 has high dimensions. If this matrix is used as the SVM classifier input, the computational cost is increased. In this study, the FDAF-score feature selection method is applied to select the primary effective features. Features are chosen in a way that their scores are larger than the threshold  $\lambda = 0.5$ . The value of  $\lambda$  is obtained experimentally.





**Fig. 3** Flowchart of the proposed method

4. The feature matrix with the preselected features is separated into two new data sets: « training data » and « validation data ».
5. Selecting the optimal features and improving SVM: In the third step, some of the redundant features are eliminated and the rest of the features create the preselected feature set. In this feature set, there may exist features that are not sensitive to the presence of defects. In this stage, the SVM classifier and the BPSO algorithm are used to select the optimal feature set. On

the other hand, the SVM algorithm has two parameters such as  $C$  and  $\sigma$  that can influence the prediction accuracy of this classifier. In this study, selecting the optimal feature set and determining the optimal SVM parameters are done simultaneously. The process of finding the optimal features and the parameters  $C$  and  $\sigma$  using the BPSO algorithm is indicated in Fig. 4.

As shown in Fig. 4, in the first step of the optimization process, the parameters of the BPSO algorithm and the interval of variations of the

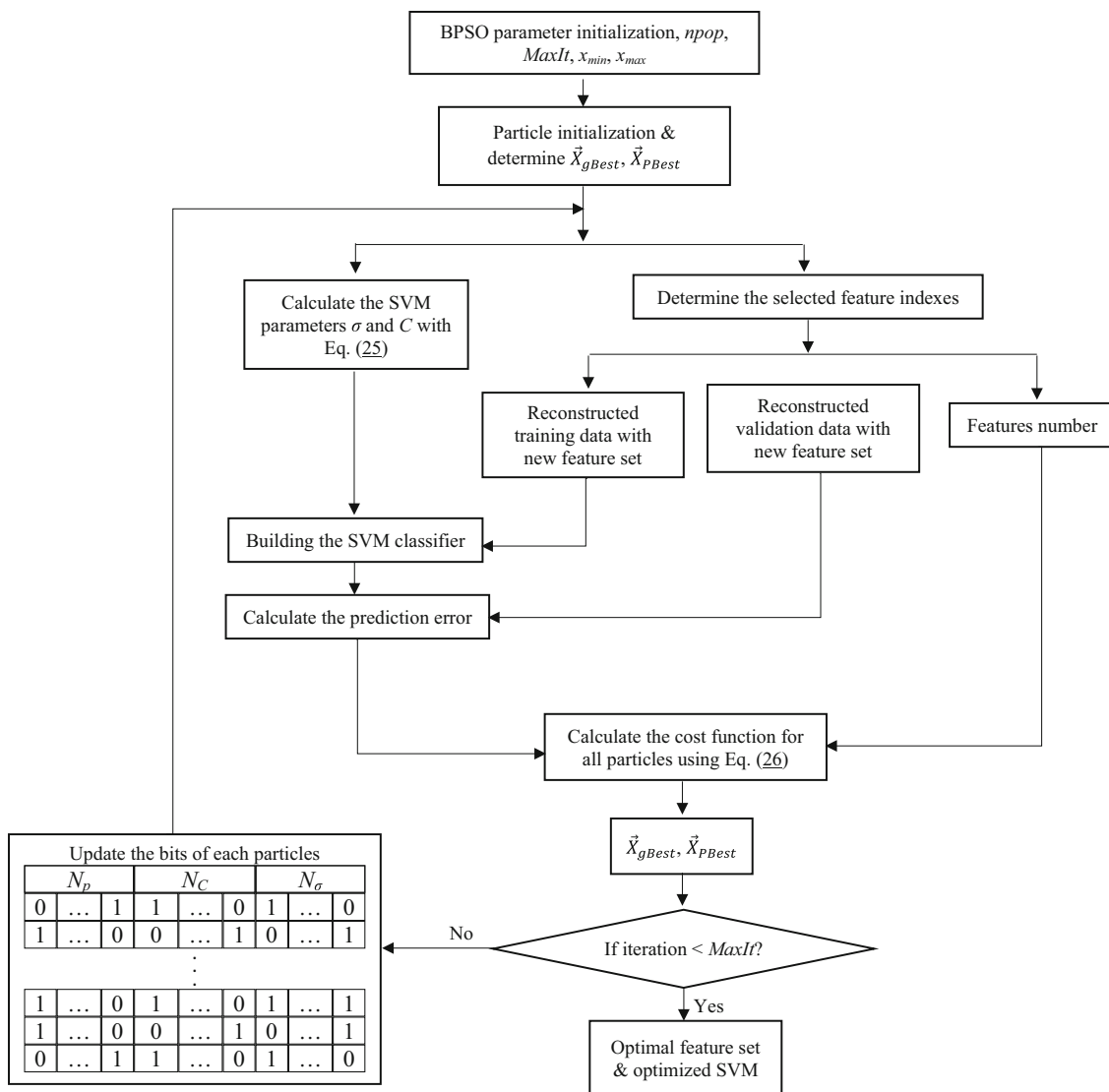


Fig. 4 Flowchart for determining the optimal feature selection and the SVM parameters

parameters  $C$  and  $\sigma$  are determined. In this paper, the maximum iteration and the number of particles are set as  $MaxIt = 100$  and  $npop = 20$ , respectively. The parameters  $C$  and  $\sigma$  are limited in interval  $[0.001, 100]$  and  $[0.01, 10]$ , respectively. Then, the initial population is randomly produced and the optimization process begins.

Each particle in the BPSO is composed of  $N = N_p + N_c + N_\sigma$  bits for displaying the selection state of the features and values of  $C$  and  $\sigma$  parameters. The value of each bit can be only 0 or 1. A schematic of a particle is shown in Fig. 5.  $N_p$  is equal to the number of

the features preselected by FDAF-score. If the value of bit is '1,' the feature is selected, and otherwise, the feature is discarded. The  $N_c$  and  $N_\sigma$  are corresponding to the number of bits of the  $C$  and  $\sigma$  parameters, respectively. The decimal value of  $C$  and  $\sigma$  can be obtained via the following equation (Zhang et al. 2018):

$$x_d = \frac{\sum_{i=1}^N (\text{bit}(i) \cdot 2^i)}{2^N - 1} (x_{dmax} - x_{dmin}) + x_{dmin}, \quad (25)$$

Fig. 5 Particle encoding schematic in BPSO

| $N_p$<br>Feature |   |     |   | $N_c$<br>Parameter C |   |     |   | $N_\sigma$<br>Parameter $\sigma$ |   |     |   |
|------------------|---|-----|---|----------------------|---|-----|---|----------------------------------|---|-----|---|
| 0                | 1 | ... | 0 | 1                    | 0 | ... | 1 | 0                                | 1 | ... | 0 |

where  $X_d$  is the decimal representation of the parameters  $C$  and  $\sigma$ .  $N$  is the number of bits.  $N$  for  $C$  and  $\sigma$  is equal to  $N_c$  and  $N_\sigma$ , respectively. Interval  $[x_{\min}, x_{\max}]$  is the search space for  $C$  and  $\sigma$ . In this work similar to Zhang et al. (2018),  $N_c$  and  $N_\sigma$  are considered 8 and 16, respectively.

During the optimization process,  $N_p$  bits in each particle are assigned to determine the reselected feature subset from the preselected feature subset. The value of parameters  $C$  and  $\sigma$  is computed using Eq. 25. The training data set and the validation data set are reconstructed using the reselected features. The SVM classifier is constructed by the reconstructed training data and the  $C$  and  $\sigma$  parameters. Then, the prediction error of the reconstructed SVM is calculated by the reconstructed validation data set. To evaluate each particle during the optimization process, the following objective function is used. It includes two terms: the prediction error and the number of selected features (Zhang et al. 2018):

$$\text{Objective function} = (1 - \alpha) \cdot \text{Error} + \alpha \cdot \frac{R}{N_p} \quad (26)$$

where Error is the prediction error of the validation data set,  $R$  is the number of the optimal features, and  $N_p$  is the number of the preselected features. Here, similar to Zhang et al. (2018), the weight coefficient  $\alpha$  is set 0.01. At the end of each repetition of the optimization process,  $\vec{X}_{g\text{Best}}$  indicates the current best result.

The termination condition of the optimization process with BPSO is the maximum iteration. The above process is repeated until the iteration number is equal to MaxIt and the bits of all particles are updated using the position and velocity updating equations. In the end, the most appropriate feature set and the optimal SVM parameters are achieved with the maximum prediction accuracy in the training step.

6. Fault diagnostics: According to Fig. 2, the final test is used for evaluating the proposed method. The optimal features obtained at previous step are utilized to form the feature vectors for the final test data. Finally, the data class and the bearing conditions are recognized by the improved SVM classifier.

## 4 Vibrational data sets

In this study, the bearing vibration signals acquired from Bearing Data Center of Case Western Reserve University (Bearing Data Center 2016) have been used to evaluate the

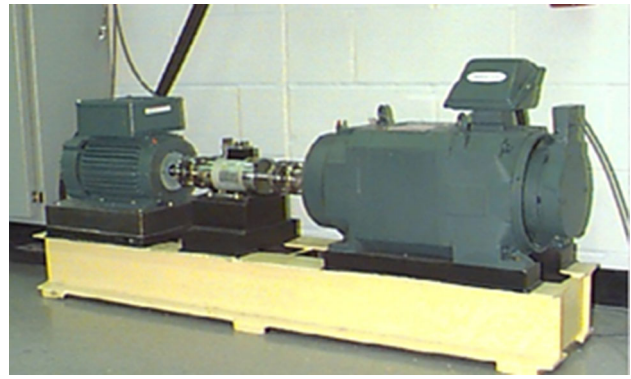


Fig. 6 Experimental set (Bearing Data Center 2016)

Table 2 Details of ball bearing 6205-2RS JEM SKF (Bearing Data Center 2016)

| Parameter        | Value (in) |
|------------------|------------|
| Inside diameter  | 0.9843     |
| Outside diameter | 2.0472     |
| Thickness        | 0.5906     |
| Ball diameter    | 0.3126     |
| Pitch diameter   | 1.5370     |

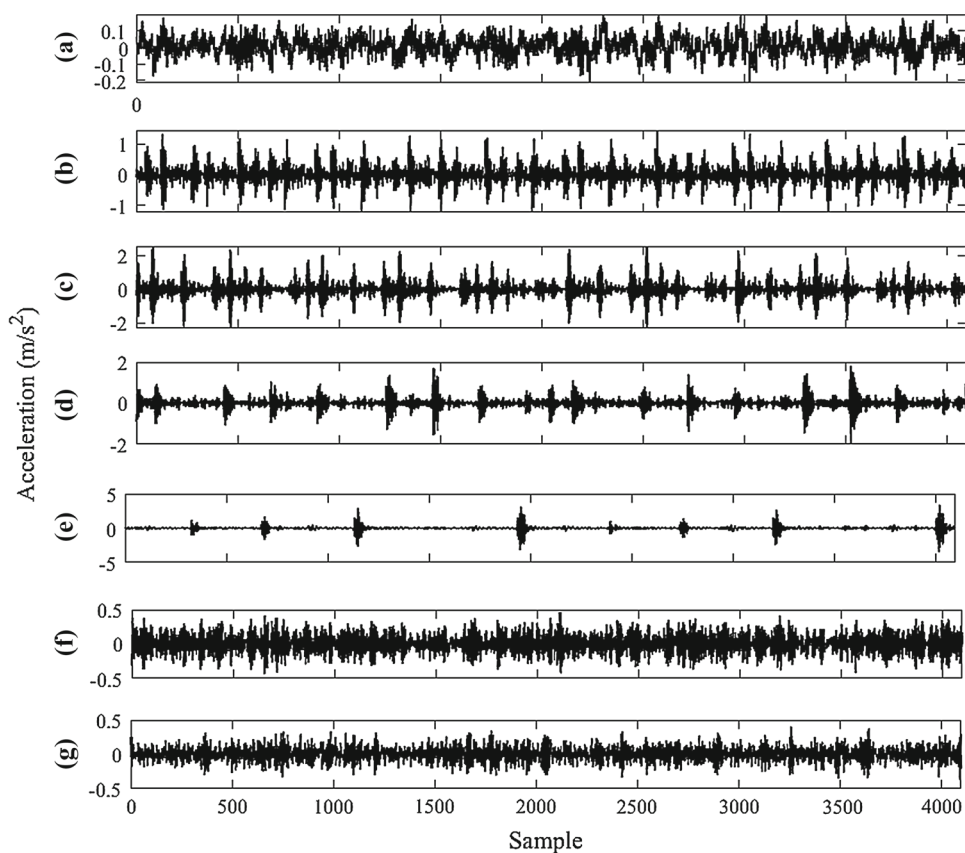
ability of the proposed method and compare its results with other methods. As shown in Fig. 6, the experimental setup includes a motor (left), coupling (middle), and dynamotor (right). The vibration signals used in this article are related to a bearing of type SKF 6205-2RS JEM. The geometric characteristics of this bearing are presented in Table 2. Single-point defects created in the bearings were produced using electro-discharge machining. These faults were produced at the inner race, outer race, and rolling element with the diameters of 0.007 in, 0.014 in, 0.021 in, and 0.028 in and the depth of 0.011 in. The vibration data were measured by accelerometers placed at the 12 o'clock position at the drive end with a sampling frequency of 12 kHz. These data were recorded for four different loads 0, 1, 2, and 3 in HP at the rotating speeds 1797 rpm, 1772 rpm, 1750 rpm, and 1730 rpm, respectively.

Table 3 describes the characteristics of the vibration data and the case studies used in this work. Case 1 is used to evaluate the capability of the proposed method in the early detection of the fault type with the smallest fault diameter of 0.007 in at different speeds. The aim of applying case studies 2–4 is to identify the different fault sizes at the different rotational speeds. The vibration data sets of the cases 2 to 4 are corresponding to the fault sizes 0.007 in, 0.014 in, 0.021 in, and 0.028 in. For example, the vibration signals corresponding to different conditions with two defect sizes 0.007 in and 0.021 in at speed 1750 rpm are illustrated in Fig. 7. As shown in Fig. 7, the apparent characteristics of vibration signals, such as shocks and amplitudes, are dependent on the type and size of the

**Table 3** Description of case studies used in the proposed method

| Case studies   | Number of classes | Number of training samples | Number of testing samples | Working condition                                    | Fault size (in)  |
|--|-------------------|----------------------------|---------------------------|--|------------------|
| Case 1: fault identification                           | 4                 | 48                         | 32                        | Normal<br>Inner race, outer race and rolling element | –<br>0.007       |
| Case 2: fault level identification for inner race      | 5                 | 60                         | 40                        | Normal<br>Inner race                                 | –<br>0.007–0.028 |
| Case 3: fault level identification for outer race      | 4                 | 48                         | 32                        | Normal<br>Outer race                                 | –<br>0.007–0.021 |
| Case 4: fault level identification for rolling element | 5                 | 60                         | 40                        | Normal<br>Rolling element                            | –<br>0.07–0.028  |

**Fig. 7** Vibrational signals under speed 1750 rpm for four different cases: **a** normal, **b** inner race (0.007 in), **c** inner race (0.021 in), **d** outer race (0.007 in), **e** outer race (0.021 in), **f** rolling element (0.007 in), and **g** rolling element (0.021 in)



defect. On the other hand, the exact diagnosis of the defect type and its severity only by examining these characteristics is a difficult problem. Therefore, it is necessary to develop a powerful feature extraction method and introduce an intelligent fault detection technique that can diagnose the fault characteristics.

The vibration signals of all classes of case studies used in this work (see Table 3) were acquired at four speeds of 1730, 1750, 1772, and 1797 in rpm. For example, in Case

1, twenty vibration signals are provided for each operating condition, *i.e.*, normal state, faulty inner race, faulty outer race, and faulty rolling element. Also, in each working condition, five signals corresponding to each rotational speed (*i.e.*, 1730, 1750, 1772, and 1797 in rpm) are considered. Finally, with considering all operating conditions, twenty signals are considered for each rotational speed. This pattern has been applied to other case studies. The fault identification results presented in this paper have been

obtained by taking into account two important factors: « various defects with different sizes » and « different rotational shaft speeds ».

## 5 Results and discussion

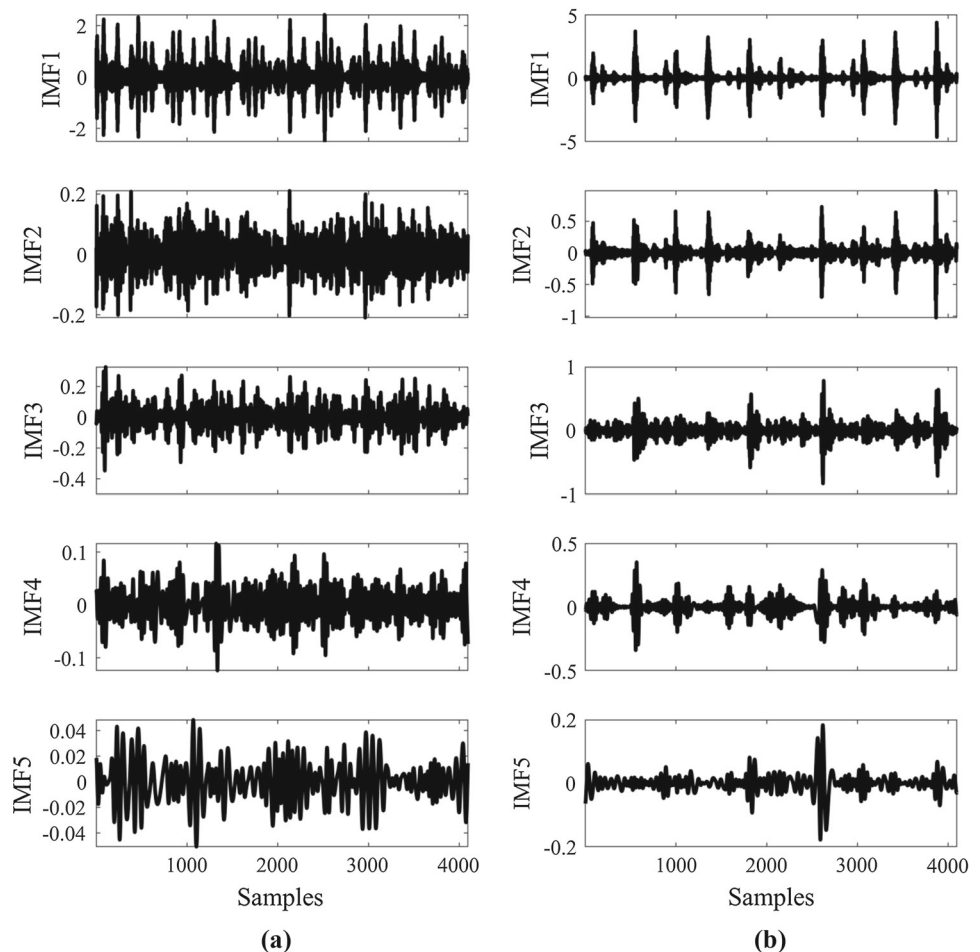
### 5.1 Optimal feature selection using the proposed method

According to the flowchart of the proposed method shown in Fig. 3, some of the features extracted from each signal are related to IMFs obtained by the EMD method. For instance, the first five IMFs of the vibration signals shown in Fig. 7c, d, which correspond to the faulty inner race and faulty outer race, are illustrated in Fig. 8.

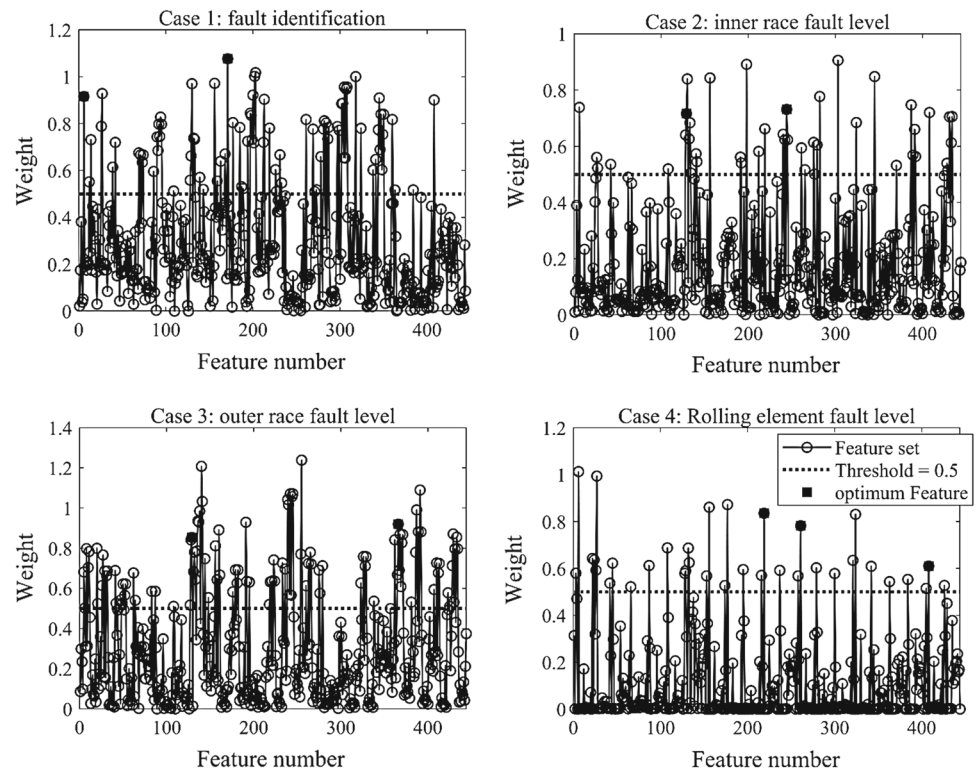
In the next step, the time-domain and frequency-domain features have been extracted from the raw signals, the first five IMFs are obtained by the EMD method, and the different levels are derived by the WPD technique. If the original feature set with high dimension is used as the SVM classifier input, then it leads to an increase in the

computational time and a decrease in the efficiency of the SVM classifier in recognizing the bearing conditions (Zhang et al. 2018). Therefore, in this work, the FDAF-score feature selection method is used to identify the weight of each feature. Then, the features have been selected which have the weight larger than the threshold value  $\lambda = 0.5$ . In the next step, the preselected features set has been used as candidates for selecting the optimal feature set using the BPSO algorithm. Figure 9 illustrates the weight scores of all features calculated by the FDAF-score method and the best features determined by the BPSO algorithm for all case studies described in Table 3. As shown in this figure, the features  $FVn_6$  (entropy of the raw signal) and  $FVn_{171}$  (Teager–Kaiser energy of the packet (1, 0) coefficients) for case 1, features  $FVn_{129}$  (standard deviation of the raw signal) and  $FVn_{244}$  (the mean of peaks of the packet (2, 2) coefficients) for case 2, the features  $FVn_{129}$  (standard deviation of the raw signal) and  $FVn_{366}$  (entropy of the packet (3, 4) coefficients) for case 3, and features  $FVn_{219}$  (entropy of the packet (2,1) coefficients),  $FVn_{261}$  (entropy of the packet (2,3) coefficients), and  $FVn_{408}$  (entropy of the packet (3,6) coefficients) for the

**Fig. 8** IMFs of the bearing signals: **a** defected inner race with fault depth 0.021 in and **b** defected outer race with fault depth 0.021 in



**Fig. 9** Weight scores of all features for all case studies



case 4 have been selected as the optimal feature set. It can be seen that IMFs of the vibration signals do not appear in any of the optimal features set. It can be concluded that the wavelet packet decomposition method is superior to the empirical mode decomposition in extracting the most effective features.

The distribution of optimal feature sets for the case studies used in this paper is plotted and shown in Fig. 10. As shown in the distribution of the feature vector of the first case study, by selecting the optimal feature set obtained by the proposed method, the different fault types can be well separated. With regard to the distribution of the optimal feature for all cases, it can be concluded that distance between the normal condition and the faulty bearing is high. Another important point in the distribution of the optimal features is the excellent separation of both damaged and healthy bearings from each other. By investigating the feature vectors of the second and third cases of study, it is seen that different fault sizes and healthy state are completely separated from each other for the defected inner ring and the defected outer ring. But, in the fourth case study, i.e., defected rolling element, there is an overlap between two fault sizes 0.007 in and 0.014 in.

## 5.2 Comparison with other methods

In this section, in order to evaluate the performance of the proposed method, its results have been compared with the

methods that are described in Table 4. Some of these methods have been studied by researchers. In method 1 (Yin et al. 2014), the BPSO algorithm has been used to select the optimal feature set from the high-dimensional feature set and to optimize the SVM parameters, simultaneously. Method 1 has been selected to appraise the effect of applying the FDAF-score feature selection technique in the proposed approach in this article. In method 2 (Zhang et al. 2018), the vibration signals are decomposed using the ITD decomposition method and the features are ranked by the Relief-F algorithm, and then, the preselected features are identified. By comparing the present study with method 2, it can be evaluated the capability of the hybrid methods FDAF-score + EMD and Relief-F + ITD. The method 3 is similar to the proposed technique. But, in method 3, GA is used as substitutions of the BPSO algorithm. In method 4, similar to the proposed method, the feature preselection is performed by the FDAF-score algorithm and the optimal feature selection is carried out by the BPSO algorithm. In this method, the SVM classifier with default parameters is employed for identifying faults. By investigating the results of this method, we can study the effect of the optimization of the SVM parameters on the fault detection accuracy.

In all of the methods described in Table 4, the meta-heuristics optimization algorithms have been used. Since meta-heuristic algorithms have a stochastic nature and their solutions might be distinct for different runs, the proposed method and other methods explained in Table 4 are run 30

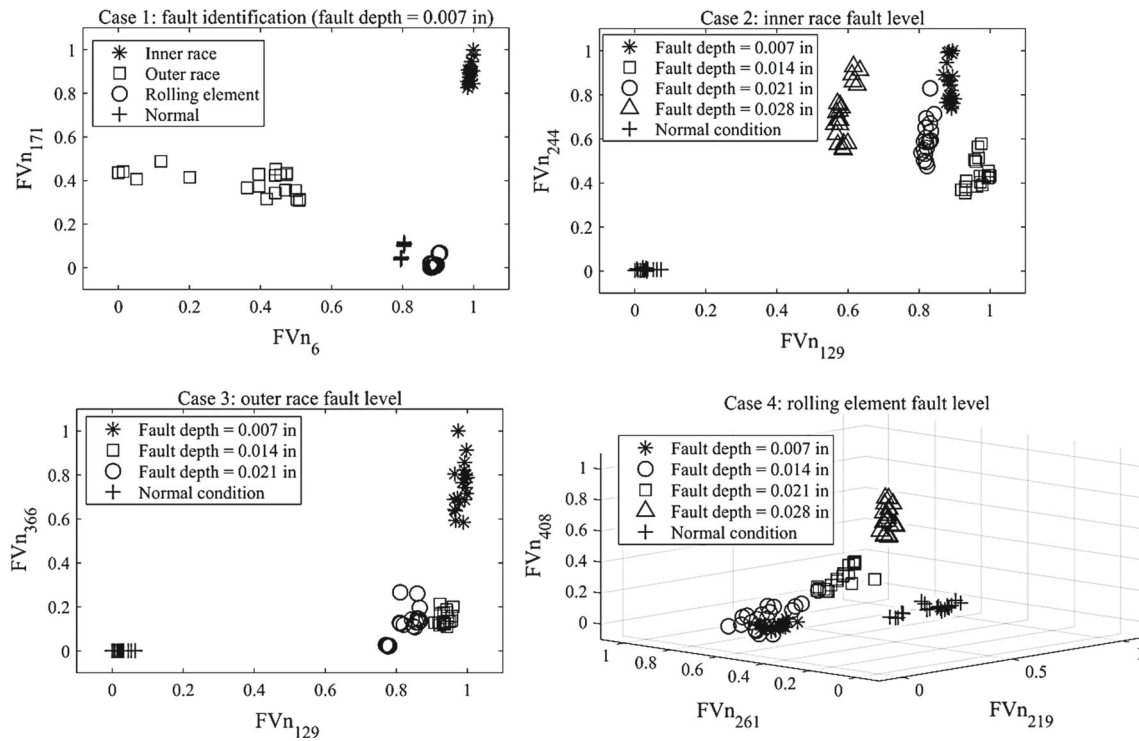


Fig. 10 Distributions of the optimal feature sets for all cases

Table 4 Description of the other fault diagnosis method for comparison with the proposed method

| Method                       | Description of method   |
|------------------------------|---|
| The proposed method          | Construct the preselected features using FDAF-score method + determine the optimal features and optimize the parameters of the SVM classifier using BPSO algorithm + fault diagnosis by the optimal feature subsets and the optimized SVM |
| Method 1 (Yin et al. 2014)   | Select the optimal feature subsets from the original feature set with high dimension and optimize the SVM classifier simultaneously using the BPSO algorithm  |
| Method 2 (Zhang et al. 2018) | Preselect the primary features using Relief-F method + find the optimal features and optimize the SVM using BPSO algorithm + detect the condition of the system by selected features and optimized SVM                                    |
| Method 3                     | FDAF-score + GA + SVM: In this method, the GA replaces the BPSO algorithm in the proposed method  |
| Method 4                     | Preselect features using FDAF-score + select the optimal feature subsets using BPSO + fault detection using the final selected features and SVM classifier with the default parameters  |

times independently. Finally, the results of the fault diagnosis accuracy are reported in Table 5 as the average training accuracy and the average testing accuracy. Also, the optimal parameters of the SVM method and the most appropriate features are presented in Table 5.

The interpretation of the results reported in Table 5 is as follows:

1. Optimal features: In method 1, there is no filter for removing the unrelated and redundant features. Therefore, the number of optimal features obtained by method 1 is higher than that of the other methods. Since in other methods, there are the preselection

2. The fault prediction accuracy: As can be seen in Table 5, among the methods discussed in this paper, the fourth method has the least accuracy in predicting the bearing conditions in both the training and the testing stages for cases of studies 1–3. In the method 4, the trained model has been constructed using the optimal feature set and the SVM classifier with default

**Table 5** Results of the proposed method and the other fault detection techniques

| Case study | Method     | Optimal features or number of features                  | Optimal parameters               | Average rate accuracy in train (%) | Average rate accuracy in test (%) |
|------------|------------|---|----------------------------------|------------------------------------|-----------------------------------|
| Case 1     | This study | $FVn_6, FVn_{171}$                                      | $\sigma = 6.7459, C = 94.5099$   | 100                                | 99.6875                           |
|            | Method 1   | 171 features  | $\sigma = 8.9005, C = 92.1569$   | 99.72                              | 94.47                             |
|            | Method 2   | $FVn_{18}, FVn_{30}$                                    | $\sigma = 9.4753, C = 23.9223$   | 100                                | 97.6042                           |
|            | Method 3   | $FVn_{177}$   | $\sigma = 7.5172, C = 7.0598$    | 100                                | 99.3750                           |
|            | Method 4   | 20 features   | Default parameter                | 70.4167                            | 70.1042                           |
| Case 2     | This study | $FVn_{129}, FVn_{244}$                                  | $\sigma = 9.1396, C = 60.3926$   | 100                                | 98.9167                           |
|            | Method 1   | 173 features  | $\sigma = 9.9599, C = 56.0789$   | 96.77                              | 91.58                             |
|            | Method 2   | $FVn_{130}, FVn_{433}$                                  | $\sigma = 1.7478, C = 88.6276$   | 100                                | 90.6667                           |
|            | Method 3   | $FVn_{219}, FVn_{279}$                                  | $\sigma = 6.7319, C = 85.4903$   | 100                                | 96.6667                           |
|            | Method 4   | 10 features   | Default parameter                | 81.2222                            | 64.5                              |
| Case 3     | This study | $FVn_{129}, FVn_{366}$                                  | $\sigma = 7.3784, C = 23.1380$   | 100                                | 98.6458                           |
|            | Method 1   | 173 features  | $\sigma = 9.7907, C = 52.1537$   | 99.86                              | 98.02                             |
|            | Method 2   | $FVn_{103}, FVn_{255}$                                  | $\sigma = 9.8184, C = 676.8630$  | 99.8611                            | 98.3333                           |
|            | Method 3   | $FVn_{10}$  | $\sigma = 5.4232, C = 10.9812$   | 100                                | 97.6042                           |
|            | Method 4   | 25 features   | Default parameter                | 99.5833                            | 95.1042                           |
| Case 4     | This study | $FVn_{219}, FVn_{261}, FVn_{408}$                       | $\sigma = 281.5803, C = 49.4123$ | 100                                | 97.25                             |
|            | Method 1   | 175 features  | $\sigma = 9.7276, C = 72.9414$   | 98.88                              | 95.66                             |
|            | Method 2   | $FVn_{130}, FVn_{134}, FVn_{244}$                       | $\sigma = 3.2577, C = 36.4712$   | 95.6667                            | 91.1667                           |
|            | Method 3   | $FVn_{261}, FVn_{279}, FVn_{263}$                       | $\sigma = 5.9764, C = 50.1966$   | 100                                | 96.5                              |
|            | Method 4   | $FVn_{129}, FVn_{258}, FVn_{261}, FVn_{279}, FVn_{408}$ | Default parameter                | 94.3333                            | 94.6667                           |

parameters. Therefore, if this model is exerted on the final testing data, it cannot obtain high accuracy in the fault diagnosis. Consequently, the compound of features and the parameters of SVM influence on the performance of fault detection.

The proposed method identifies exactly the conditions of the validation data set in the training step. Another point that can be found by observing the results of all case studies is the superiority of the proposed method to the other methods for detecting the bearing conditions for the final testing data. This fact indicates the ability of the FDAF-score and BPSO-SVM methods. According to Table 5, the method 3 is ranked second among all methods. In the third method, the GA is used to select the optimal features and optimize the SVM model. This is while the BPSO algorithm plays this role in the proposed technique. The comparison of the proposed technique with method 3 shows that BPSO is superior to GA in the fault detection process. Of course, it should be pointed out that the results of the third method are close to the results of the proposed method.

Also, in Table 5, it can be seen that the proposed method is more effective and more accurate than the second method. In other words, when the FDAF-score method is used instead of the Relief-F method, the fault detection

accuracy in cases such as case 2 and case 4 increases significantly. This improvement is insignificant in cases 1 and 3.

When the optimization algorithms with random behavior are used in the intelligent fault detection methods, it is important to consider the issue of stability of these methods and the non-scattering of solutions in different implementations. Here, the standard deviation has been used in order to investigate the stability of the method presented in this paper and the methods described in Table 4. For this purpose, these methods are run 100 times independently, and the standard deviation of the fault prediction results has been calculated at the final test stage. The results of the current work and other techniques are presented in Table 6 for all case studies. It should be noted that the low value of standard deviation for a particular method reflects its robustness and the non-scattering of the responses obtained in different runs. As can be seen in Table 6, the data prediction by the proposed approach for all case studies has the lowest standard deviation than other methods. This result shows that the method presented in this paper has very good stability in predicting the status of a signal sample in different implementations. In other words, the dispersion of solutions predicted by the present work is less than that of the other methods.



**Table 6** Results of the standard deviation of the proposed approach and other methods

| Case studies | Methods  |          |          |          |            |
|--------------|----------|----------|----------|----------|------------|
|              | Method 1 | Method 2 | Method 3 | Method 4 | This study |
| Case 1       | 2.8570   | 3.1974   | 4.1021   | 4.4700   | 0.8720     |
| Case 2       | 4.6037   | 2.2310   | 3.9385   | 7.5952   | 1.5364     |
| Case 3       | 2.5844   | 3.4029   | 2.8529   | 5.0145   | 1.4060     |
| Case 4       | 4.0735   | 4.9322   | 4.7130   | 4.2262   | 3.6397     |

## 6 Conclusion

The main subject of this paper is the bearing condition monitoring with various defects under the different rotational speeds. The analyzed signals are corresponding to four cases: normal condition, damaged inner race, damaged outer race, and damaged rolling element. The intelligent approaches presented in this study, the empirical mode decomposition (EMD) and the wavelet packet decomposition (WPD), are applied for decomposing and processing the vibration signals. In the next stage, the time- and frequency-domain features are utilized to construct the feature matrix. These features are extracted from the raw signal, the first five IMFs obtained by EMD, and the wavelet coefficients. The increase in the number of features leads to producing a high-dimensional feature matrix that may include meaningless and redundant features. Consequently, the FDAF-score technique is applied for removing some insensitive features. Finally, the most informative features and the optimal parameters of the support vector machine (SVM) are determined using the binary particle swarm optimization (BPSO) algorithm, simultaneously.

The results of the proposed method are briefly listed below:

1. The proposed method is able to select the features that are sufficiently sensitive in case of the presence of defects in bearings. These features separate the normal and faulty conditions very well.
2. The proposed method is able to identify the different fault sizes for the three faulty states in bearings. In fact, the optimal features obtained by the hybrid method FDAF-score + BPSO are sensitive to the different fault sizes in each bearing component. Also, the results demonstrate that the proposed method for detecting the different fault types with the smallest size or the early fault detection is superior to the other methods considered in the literature.
3. The dimensionality problem of the feature space is solved by utilizing the FDAF-score method for pre-selecting the useful features and the BPSO algorithm for reselecting the optimal features and improving the SVM classifier. The results show that the proposed method is efficient and suitable in selecting the most

appropriate features and diagnosing the bearing conditions. Also, these results indicate the high performance of the proposed technique compared to other fault diagnosis methods. This superiority implies the capability of the SVM, BPSO, FDAF-score, EMD, and WPD methods in the proposed approach in this article.

4. The fault prediction results show that the proposed technique has very good stability in different implementations.

**Author contributions** S.N.C designed and coordinated the study. S.N.C wrote the manuscript. S.N.C, A.B, and F.N reviewed the manuscript and contributed to its revision. All the authors discussed the results and gave their final approval for publication.

**Acknowledgements** This study was financially supported by the University of Guilan to S.N.C. The funder had no role in study design, data collection, and analysis, decision to publish, or preparation of the manuscript.

## Compliance with ethical standards

**Conflict of interest** We declare we have no competing interests.

## References

- Ali JB, Fnaiech N, Saidi L, Chebel-Morello B, Fnaiech F (2015) Application of empirical mode decomposition and artificial neural network for automatic bearing fault diagnosis based on vibration signals. *Appl Acoust* 89:16–27. <https://doi.org/10.1016/j.apacoust.2014.08.016>
- Attoui I, Fergani N, Boutasseta N, Oudjani B, Deliou A (2017) A new time–frequency method for identification and classification of ball bearing faults. *J Sound Vib* 397:241–265. <https://doi.org/10.1016/j.jsv.2017.02.041>
- Banka H, Dara S (2015) A Hamming distance based binary particle swarm optimization (HDBPSO) algorithm for high dimensional feature selection, classification and validation. *Pattern Recognit Lett* 52:94–100. <https://doi.org/10.1016/j.patrec.2014.10.007>
- Bearing Data Center (2016) Case Western Reserve University. <http://csegroups.case.edu/bearingdatacenter/home>
- Bhuyan HK, Kamila NK (2015) Privacy preserving sub-feature selection in distributed data mining. *Appl Soft Comput* 36:552–569. <https://doi.org/10.1016/j.asoc.2015.06.060>
- Bordoloi DJ, Tiwari R (2014a) Support vector machine based optimization of multi-fault classification of gears with evolutionary algorithms from time–frequency vibration data.

- Measurement 55:1–14. <https://doi.org/10.1016/j.measurement.2014.04.024>
- Bordoloi DJ, Tiwari R (2014b) Optimum multi-fault classification of gears with integration of evolutionary and SVM algorithms. *Mech Mach Theory* 73:49–60. <https://doi.org/10.1016/j.mechmachtheory.2013.10.006>
- Bordoloi DJ, Tiwari R (2015) Optimisation of SVM methodology for multiple fault taxonomy of rolling bearings from acceleration records. In: Proceedings of the 9th IFToMM international conference on rotor dynamics, pp 533–542. [https://doi.org/10.1007/978-3-319-06590-8\\_43](https://doi.org/10.1007/978-3-319-06590-8_43)
- Chen G, Chen J (2015) A novel wrapper method for feature selection and its applications. *Neurocomputing* 159:219–226. <https://doi.org/10.1016/j.neucom.2015.01.070>
- Fatima S, Mohanty AR, Naik VNA (2015) Multiple fault classification using support vector machine in a machinery fault simulator. *Vib Eng Technol Mach*. [https://doi.org/10.1007/978-3-319-09918-7\\_90](https://doi.org/10.1007/978-3-319-09918-7_90)
- Frei MG, Osorio I (2007) Intrinsic time-scale decomposition: time-frequency–energy analysis and real-time filtering of non-stationary signals. *Proc R Soc Lond A Math Phys Sci* 463(2078):321–342. <https://doi.org/10.1098/rspa.2006.1761>
- Fu W, Tan J, Zhang X, Chen T, Wang K (2019) Blind parameter identification of MAR model and mutation hybrid GWO-SCA optimized SVM for fault diagnosis of rotating machinery. *Complexity*. <https://doi.org/10.1155/2019/3264969>
- Gilles J (2013) Empirical wavelet transform. *IEEE T Signal Process* 61(16):3999–4010. <https://doi.org/10.1109/TSP.2013.2265222>
- Guo T, Deng Z (2017) An improved EMD method based on the multi-objective optimization and its application to fault feature extraction of rolling bearing. *Appl Acoust* 127:46–62. <https://doi.org/10.1016/j.apacoust.2017.05.018>
- Guyon I, Weston J, Barnhill S, Vapnik V (2002) Gene selection for cancer classification using support vector machines. *Mach Learn* 46(1–3):389–422. <https://doi.org/10.1023/A:1012487302797>
- Hsu CW, Lin CJ (2002) A comparison of methods for multiclass support vector machines. *IEEE Trans Neural Netw* 13(2):415–425. <https://doi.org/10.1109/72.991427>
- Huang CL, Wang CJ (2006) A GA-based feature selection and parameters optimization for support vector machines. *Expert Syst Appl* 31(2):231–240. <https://doi.org/10.1016/j.eswa.2005.09.024>
- Huang NE, Shen Z, Long SR, Wu MC, Shih HH, Zheng Q, Yen NC, Tung CC, Liu HH (1998) The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proc R Soc Lond A Math Phys Sci* 454(1971):903–995. <https://doi.org/10.1098/rspa.1998.0193>
- Jang JS (1993) ANFIS: adaptive-network-based fuzzy inference system. *IEEE Trans Syst Man Cybern* 23:665–685. <https://doi.org/10.1109/21.256541>
- Jedliński L, Jonak J (2015) Early fault detection in gearboxes based on support vector machines and multilayer perceptron with a continuous wavelet transform. *Appl Soft Comput* 30:636–641. <https://doi.org/10.1016/j.asoc.2015.02.015>
- Jiang F, Zhu Z, Li W, Chen G, Zhou G (2013) Robust condition monitoring and fault diagnosis of rolling element bearings using improved EEMD and statistical features. *Meas Sci Technol* 25(2):025003. <https://doi.org/10.1088/0957-0233/25/2/025003>
- Kennedy J, Eberhart RC (1997) A discrete binary version of the particle swarm algorithm. In: 1997 IEEE international conference on systems, man, and cybernetics. Computational cybernetics and simulation, vol 5, pp 4104–4108. IEEE. <https://doi.org/10.1109/icsmc.1997.637339>
- Lei Y, Zuo MJ (2009) Fault diagnosis of rotating machinery using an improved HHT based on EEMD and sensitive IMFs. *Meas Sci Technol* 20(12):125701. <https://doi.org/10.1088/0957-0233/20/12/125701>
- Lei Y, Liu Z, Ouazri J, Lin J (2017) A fault diagnosis method of rolling element bearings based on CEEMDAN. *Proc Inst Mech Eng C* 10:1804–1815. <https://doi.org/10.1177/0954406215624126>
- Li Z, Yan X, Tian Z, Yuan C, Peng Z, Li L (2013) Blind vibration component separation and nonlinear feature extraction applied to the nonstationary vibration signals for the gearbox multi-fault diagnosis. *Measurement* 46(1):259–271. <https://doi.org/10.1016/j.measurement.2012.06.013>
- Li Y, Xu M, Wei Y, Huang W (2015) An improvement EMD method based on the optimized rational Hermite interpolation approach and its application to gear fault diagnosis. *Measurement* 63:330–345. <https://doi.org/10.1016/j.measurement.2014.12.021>
- Lu L, Yan J, de Silva CW (2015) Dominant feature selection for the fault diagnosis of rotary machines using modified genetic algorithm and empirical mode decomposition. *J Sound Vib* 344:464–483. <https://doi.org/10.1016/j.jsv.2015.01.037>
- Mallat SG (1989) A theory for multiresolution signal decomposition: the wavelet representation. *IEEE Trans Pattern Anal Mach Intell* 11(7):674–693. <https://doi.org/10.1109/34.192463>
- Nezamivand Chegini S, Bagheri A, Najafi F (2018) PSOSCALF: a new hybrid PSO based on sine cosine algorithm and Levy flight for solving optimization problems. *Soft Comput, Appl*. <https://doi.org/10.1016/j.asoc.2018.09.019>
- Nezamivand Chegini S, Bagheri A, Najafi F (2019) Application of a new EWT-based denoising technique in bearing fault diagnosis. *Measurement* 144:275–297. <https://doi.org/10.1016/j.measurement.2019.05.049>
- Nguyen P, Kang M, Kim JM, Ahn BH, Ha JM, Choi BK (2015) Robust condition monitoring of rolling element bearings using de-noising and envelope analysis with signal decomposition techniques. *Expert Syst Appl* 42(22):9024–9032. <https://doi.org/10.1016/j.eswa.2015.07.064>
- Shan Y, Zhou J, Jiang W, Liu J, Xu Y, Zhao Y (2019) A fault diagnosis method for rotating machinery based on improved variational mode decomposition and a hybrid artificial sheep algorithm. *Meas Sci Technol* 30(5):055002
- Shi Y, Eberhart R (1998) A modified particle swarm optimizer. In: International conference on evolutionary computation proceedings. IEEE world congress on computational intelligence, pp 69–73. IEEE. <https://doi.org/10.1109/ICEC.1998.699146>
- Song Q, Jiang H, Liu J (2017) Feature selection based on FDA and F-score for multi-class classification. *Expert Syst Appl* 81:22–27. <https://doi.org/10.1016/j.eswa.2017.02.049>
- Sperduti A, Starita A (1997) Supervised neural networks for the classification of structures. *IEEE Trans Neural Netw* 8:714–735
- Tabrizi A, Garibaldi L, Fasana A, Marchesiello S (2015) Early damage detection of roller bearings using wavelet packet decomposition, ensemble empirical mode decomposition and support vector machine. *Meccanica* 50(3):865–874. <https://doi.org/10.1007/s11012-014-9968-z>
- Vakharia V, Gupta VK, Kankar PK (2016) Bearing fault diagnosis using feature ranking methods and fault identification algorithms. *Proc Eng* 144:343–350. <https://doi.org/10.1016/j.proeng.2016.05.142>
- Vapnik V (1995) The nature of statistical learning theory. Springer, New York
- Wang CS, Sha CY, Su M, Hu YK (2017) An algorithm to remove noise from locomotive bearing vibration signal based on self-adaptive EEMD filter. *J Cent South Univ* 24(2):478–488. <https://doi.org/10.1007/s11771-017-3450-8>
- Wei Z, Wang Y, He S, Bao J (2017) A novel intelligent method for bearing fault diagnosis based on affinity propagation clustering

- and adaptive feature selection. *Knowl Based Syst* 116:1–12. <https://doi.org/10.1016/j.knsys.2016.10.022>
- Xue X, Zhou J, Xu Y, Zhu W, Li C (2015) An adaptively fast ensemble empirical mode decomposition method and its applications to rolling element bearing fault diagnosis. *Mech Syst Signal Process* 62:444–459. <https://doi.org/10.1016/j.ymsp.2015.03.002>
- Yan X, Jia M (2018) A novel optimized SVM classification algorithm with multi-domain feature and its application to fault diagnosis of rolling bearing. *Neurocomputing* 313(3):47–64. <https://doi.org/10.1016/j.neucom.2018.05.002>
- Yin H, Qiao J, Fu P, Xia XY (2014) Face feature selection with binary particle swarm optimization and support vector machine. *J Inf Hiding Multimed Signal Process* 5(4):731–739
- Zeng Z, Zhang H, Zhang R, Yin C (2015) A novel feature selection method considering feature interaction. *Pattern Recognit* 48(8):2656–2666. <https://doi.org/10.1016/j.patcog.2015.02.025>
- Zhang X, Chen W, Wang B, Chen X (2015) Intelligent fault diagnosis of rotating machinery using support vector machine with ant colony algorithm for synchronous feature selection and parameter optimization. *Neurocomputing* 167:260–279. <https://doi.org/10.1016/j.neucom.2015.04.069>
- Zhang X, Zhang Q, Chen M, Sun Y, Qin X, Li H (2018) A two-stage feature selection and intelligent fault diagnosis method for rotating machinery using hybrid filter and wrapper method. *Neurocomputing* 275:2426–2439. <https://doi.org/10.1016/j.neucom.2017.11.016>
- Zhao DZ, Li JY, Cheng WD, Wang TY, Wen WG (2016) Rolling element bearing instantaneous rotational frequency estimation based on EMD soft-thresholding denoising and instantaneous fault characteristic frequency. *J Cent South Univ* 23(7):1682–1689. <https://doi.org/10.1007/s11771-016-3222-x>
- Ziani R, Felkaoui A, Zegadi R (2017a) Bearing fault diagnosis using multiclass support vector machines with binary particle swarm optimization and regularized Fisher's criterion. *J Intell Manuf* 28(2):405–417. <https://doi.org/10.1007/s10845-014-0987-3>
- Ziani R, Mahgoun H, Fedala S, Felkaoui A (2017b) Feature selection scheme based on Pareto method for gearbox fault diagnosis. *Signal Process Appl Rotating Mach Diagn*. [https://doi.org/10.1007/978-3-319-96181-1\\_1](https://doi.org/10.1007/978-3-319-96181-1_1)

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