



Fault severity recognition of aviation piston pump based on feature extraction of EEMD paving and optimized support vector regression model

Chuanqi Lu^{a,b}, Shaoping Wang^a, Viliam Makis^{b,*}

^a School of Automation Science and Electrical Engineering, Beihang University, Beijing 100191, China

^b Department of Mechanical and Industrial Engineering, University of Toronto, Toronto M5S 3G8, Canada



ARTICLE INFO

Article history:

Received 26 July 2016

Received in revised form 25 November 2016

Accepted 30 March 2017

Available online 5 April 2017

Keywords:

EEMD paving

Principal component analysis

Support vector regression

Fault severity identification

Aviation hydraulic pump

ABSTRACT

Recognizing an early fault for aviation hydraulic pump and evaluating its size is essential in this industrial application. This paper proposes a new method which combines ensemble empirical mode decomposition (EEMD) paving and optimized support vector regression (SVR) model to detect faults and estimate the fault sizes of a piston pump. Different from other feature extraction methods in which the information of intrinsic mode functions (IMFs) is not being fully utilized, the collected pressure signals are first decomposed by EEMD, and then some useful IMFs are selected by calculating the correlation coefficients between the signals reconstructed by the chosen IMFs and the original signals. These selected IMFs are referred to as EEMD paving. Subsequently, some new fault features considering time domain, frequency domain, and time-frequency domain are extracted from the paving of EEMD. To acquire the most sensitive fault features, principal component analysis (PCA) is then employed to reduce the dimensionality of the original feature vectors. Finally, SVR model is constructed to identify different fault sizes of aviation pump. To achieve higher recognition accuracy, a new method combining genetic algorithm (GA) with grid search is adopted to optimize the parameters of the SVR model. The effectiveness of the proposed method is verified by two datasets collected from a test rig under different conditions. The results demonstrate that the fault features based on the proposed method can be used to characterize the pump fault severity more accurately, and the constructed SVR model has higher recognition accuracy and better prediction ability when compared with previously published methods. The proposed method can also be readily used in other industrial applications.

© 2017 Elsevier Masson SAS. All rights reserved.

1. Introduction

The aviation hydraulic piston pump, as the key component of the hydraulic system, has been widely used in the modern large civil aircraft thanks to its advantages of high rated pressure, high force-to-weight ratio, high efficiency, etc. [1,2]. Failures of piston pump might cause unexpected machine breakdowns, resulting in economic loss and, even worse, catastrophic consequences. Therefore, effective and feasible fault diagnosis methods for piston pump are vital for successful operation of the whole aircraft hydraulic system [3,4].

Discharge pressure, one of the most important parameters, can reflect the performance of piston pump. When a pump is working, the pump discharge pressure signals comprise variety of state in-

formation. As a result, fault diagnosis methods based on pressure signals have been widely investigated [4–6]. Although these methods have been proved to be effective and powerful for fault diagnosis of pump, most studies have been focusing on distinguishing different fault patterns. As a matter of fact, the degradation of a normal pump is a progressive process [7]. The failure usually occurs when an early pump fault grows to a severe extent. So if the defect sizes could be reliably estimated to track the pump's health status, the fault diagnosis would be more effective. So far, fault severity recognition of the piston pump based upon pressure signals has rarely been reported.

To recognize the fault sizes of piston pump, effective fault features need to be extracted first. Signal processing-based feature extraction methods have been used in the fault diagnosis of pump, including wavelet packet decomposition (WPT) [8], and Empirical Mode Decomposition (EMD) [5,7]. Unlike other methods, EMD [9], as one of the most powerful time-frequency analysis techniques, is self-adaptive and especially suitable for processing the

* Corresponding author.

E-mail address: makis@mie.utoronto.ca (V. Makis).

non-stationary and non-linear signals [10]. In addition, the obtained Intrinsic Mode Functions (IMFs) decomposed by EMD are based on the local characteristic scales of a signal itself and contain valuable information. When pump fault occurs, it affects the piston pump discharge pressure signal. This fault will generate some obvious abnormal signal patterns in some frequency bands, and because each IMF obtained from EMD includes different frequency bands, so EMD is suitable for faults feature extraction of pumps. However, there still exist some drawbacks, one of which is mode mixing. Mode mixing will affect the decomposition accuracy which is closely associated with the feature extraction. To overcome this problem, Ensemble Empirical Mode Decomposition (EEMD) has been proposed [11]. The research results show that the IMFs obtained from EEMD is more accurate [12–14]. Though these EEMD-based feature extraction methods have shown outstanding performance in fault pattern recognition, the information of IMFs is not being fully utilized and single time-frequency domain features are not sufficient for fault severity recognition [15]. Consequently, more proper features for fault size identification still need to be constructed.

After the fault features have been extracted, the dimensionality of the features needs to be reduced because there is no guarantee that all these features are equally useful or even necessary in reflecting the pump health status. Additionally, too many inputs may increase the computational burden [16]. Popular dimensionality reduction methods include principal component analysis (PCA) [17], independent component analysis (ICA) [18], and kernel principal component analysis (KPCA) [19]. The previous studies indicate that PCA has the best performance in dimensionality reduction and it leads to the highest prediction accuracies among the methods which have been applied in fault severity recognition [20]. Accordingly, PCA is introduced to obtain sensitive and informative fault features.

Finally, after selecting the sensitive fault features, there is still a challenge how to recognize the fault sizes by using these features. Artificial neural networks (ANN) developed by mimicking the way in which the human brain processes information, have been applied in the fault severity recognition field [21,22]. Though ANN-based methods have achieved good performance, some insufficiencies, such as weak generalization ability and slow convergence have been found. In addition, they are not suitable for processing small samples for training [23]. Different from ANN methods, support vector machine (SVM) proposed by Vapnik [24], which is based upon statistical learning theory, has no need for a large number of samples for training due to its superior generalization ability [23]. But the traditional SVM can only be applied to handle a binary problem. Thus, support vector regression (SVR) whose target values are continuous has been developed. SVR has been originally used in time series prediction [25], and recently it has been used in the field of fault size recognition as it can establish a stable nonlinear relationship between inputs and outputs successfully [26–29]. Although SVR-based methods have shown remarkable performance, the parameter selection of SVR is an ongoing research issue [30]. As the SVR model accuracy is closely related to the parameters selection, an optimized SVR model needs to be created to obtain the higher recognition accuracy.

In this study, a new fault severity recognition scheme based on EEMD paving and optimized SVR model is proposed to estimate the fault sizes of aviation hydraulic pump. First, novel multi-domain fault features considering time domain, frequency domain, and time-frequency domain features, are extracted from the paving of EEMD to characterize the pump health status. To select more sensitive fault features, PCA is employed to reduce the dimensionality of the original extracted feature vectors. Then, a parameter optimization selection method combining genetic algorithm (GA) with grid search is introduced to determine the opti-

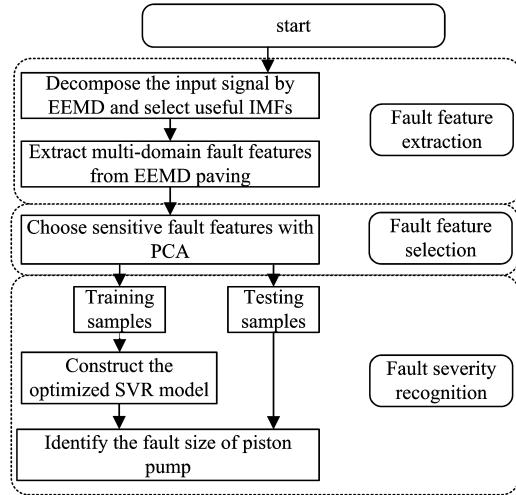


Fig. 1. Framework of the proposed fault severity recognition scheme.

mal parameters of SVR model. Finally, the fault sizes of pump are reliably estimated via the constructed SVR model. The proposed scheme is validated using the pump dataset considering different levels of fault severity under different conditions.

The remainder of this paper is organized as follows. Section 2 presents the new fault severity recognition method based on EEMD paving and an optimized SVR model. In Section 3, the proposed method is verified by experimental data collected from a pump test platform. A comparison analysis with other methods is investigated in Section 4. Conclusions are drawn in Section 5.

2. The fault severity recognition method

The proposed fault severity recognition scheme includes the following three steps: multi-domain fault feature extraction, sensitive fault feature selection and fault size identification. The detailed description of each step is given in the following subsections. The framework of the proposed method is depicted in Fig. 1.

2.1. Multi-domain fault feature extraction

As previously noted, to prevent the machine's breakdown over time and minimize economic loss, recognizing fault severity of pump is very essential. Hence, extracting useful fault features which can reflect the actual performance of pump from the collected signals is a critical task for the later processing.

It follows from previous research that when the fault of piston pump occurs, it is reflected in the pump discharge pressure signal. So the pump outlet pressure signal is used to extract fault features. However, the operation of pumps usually leads to a dynamic behavior that generates non-stationary pressure signals with strong background noise. Especially when the aviation piston pump operates in a condition of high pressure, high speed and heavy load, non-stationary characteristics of pressure signals will be more obvious. This makes the traditional temporal and frequency analysis insufficient in the case of non-stationary pressure signals. Fortunately, this limitation can be solved by the time-frequency analysis methods [8,31]. Different from other time-frequency methods, EMD is self-adaptive and especially suitable for processing the non-stationary and non-linear signals. In addition, the obtained IMFs decomposed by EMD are based on the local characteristic scales of a signal itself and contain valuable information. As a result, EMD has been widely used in feature extraction of piston pump. Nevertheless, a problem of EMD algorithm is mode mixing which affects the decomposition accuracy. To illustrate this prob-

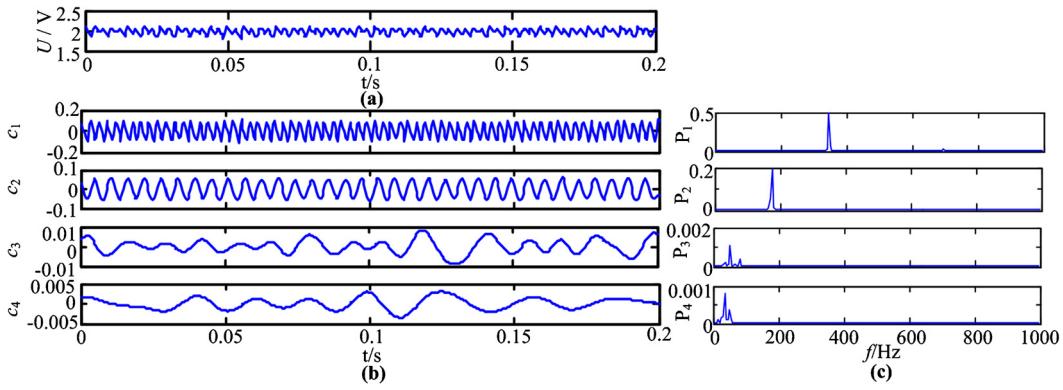


Fig. 2. The decomposition results based on EMD (a) the original pump discharge pressure signal (b) the first four IMFs (c) the power spectra of the first four IMFs.

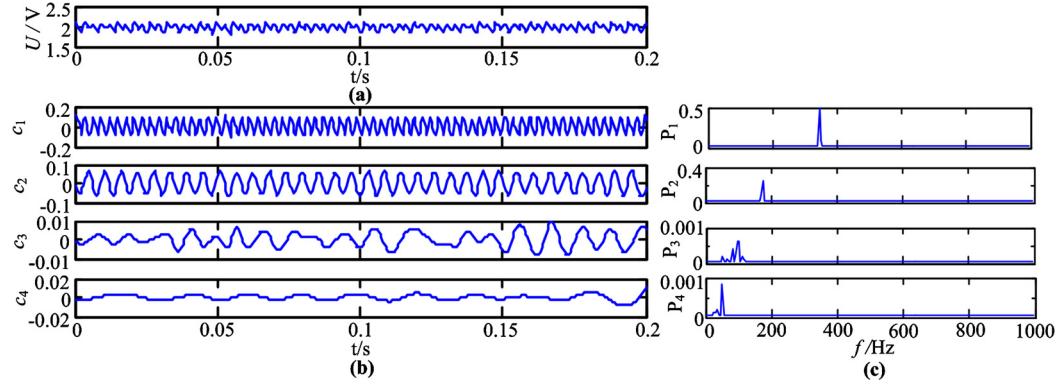


Fig. 3. The decomposition results based on EEMD (a) the original pump discharge pressure signal (b) the first four IMFs (c) the power spectrum of the first four IMFs.

lem, a pressure signal collected from a piston pump is considered and shown in Fig. 2(a).

First, the original pressure signal is decomposed by EMD. The obtained first four IMFs are shown in Fig. 2(b). To further analyze the frequency components of each IMF, the power spectra of the first four IMFs are calculated and presented in Fig. 2(c). From Fig. 2(c), it is obvious that the same frequency components are contained in IMFs c_3 and c_4 , which means that mode mixing occurs. Thus, both IMFs c_3 and c_4 obtained from EMD fail to characterize the characteristics of the original signal. As a result, EEMD is proposed to alleviate the problem of mode mixing [10].

2.1.1. The basic theory of EEMD

Compared with the original EMD, the means of an ensemble of trials are considered as the true IMFs in EEMD algorithm. In each trial, the normally distributed white noise of finite amplitude is added to the original signal. This new method is mainly based on the principle that the added white noise can populate the whole time-frequency space uniformly according to the constituting components of different scales [10]. On the basis of this analysis, the EEMD algorithm can be presented [30]:

- (1) Determine the number of ensemble M .
- (2) Give the amplitude of the added normally distributed white noise, and $j = 1$.
- (3) Add the white noise series with the given amplitude to the original signal $x(t)$ to obtain a new signal:

$$x_j(t) = x(t) + n_j(t) \quad (1)$$

where $n_j(t)$ symbolizes the j th added white noise series, and $x_j(t)$ denotes the noise-added signal of the j th trial.

- (4) Perform the original EMD to decompose the obtained signal $x_j(t)$ into some IMFs, and the obtained signal can be written as follows:

$$x_j(t) = \sum_{p=1}^{\rho} c_{j,p}(t) + r_{j,\rho}(t) \quad (2)$$

where ρ is the number of IMFs, $r_{j,\rho}(t)$ is the residue function which stands for the mean trend of the signal, and $c_{j,p}(t)$ denotes the obtained p th IMF of the j th trial.

- (5) Repeat steps (3) and (4) M times with different white noise series each time, and an ensemble of IMFs denoted by $\{c_{1,p}(t)\}, \{c_{2,p}(t)\}, \dots, \{c_{M,p}(t)\}$ can be obtained.
- (6) Compute the ensemble means of M trials for each IMF as the final result:

$$c_p(t) = \frac{1}{M} \sum_{j=1}^M c_{j,p}(t) \quad j = 1, 2, \dots, M, p = 1, 2, \dots, \rho \quad (3)$$

where $c_p(t)$ represents the p th IMF obtained from EEMD.

To verify the advantage of EEMD, the pressure signal shown in Fig. 2(a) is processed again. The signal is first decomposed by EEMD. In EEMD algorithm, the number of ensemble is 100 and the amplitude of added white noise is 0.1 standard deviation of the signal. Subsequently, the obtained first four IMFs and their power spectra are depicted in Fig. 3(b) and Fig. 3(c). Compared with Fig. 2(c), it shows that there is nearly no same frequency component in different IMFs. Thus, it can be demonstrated that EEMD can resolve the problem of mode mixing and obtain an improved decomposition results with physical meaning.

2.1.2. Feature extraction based on EEMD paving

Although previous studies indicate that EEMD-based extracted features have been considered in the field of fault diagnosis, it is difficult to get a better recognition accuracy based on simple time domain or frequency domain features according to the research results in [15]. Besides, the IMFs information is not fully utilized and

Table 1
The time-domain features.

Square root amplitude value: $\left(\frac{1}{T} \sum_{q=1}^T \sqrt{ x_q }\right)^2$	Impulsive index: $\max(x_q) / \left(\frac{1}{T} \sum_{q=1}^T x_q \right)$
Shape index: $\sqrt{\frac{1}{T} \sum_{q=1}^T x_q^2} / \left(\frac{1}{T} \sum_{q=1}^T x_q \right)$	Clearance index: $\max(x_q) / \left(\frac{1}{T} \sum_{q=1}^T \sqrt{ x_q }\right)^2$
Crest index: $\max(x_q) / \sqrt{\frac{1}{T} \sum_{q=1}^T x_q^2}$	Kurtosis index: $\sum_{q=1}^T x_q^4 / T \left(\sqrt{\frac{1}{T} \sum_{q=1}^T x_q^2}\right)^4$
Skewness index: $\sum_{q=1}^T x_q^3 / T \left(\sqrt{\frac{1}{T} \sum_{q=1}^T x_q^2}\right)^3$	where T is the length of the data, x_q is a signal series of the selected IMF.

Table 2
The frequency domain and time–frequency domain features.

Frequency center: $\sum_{z=1}^N f_z b_z / \sum_{z=1}^N b_z$	Root mean square frequency: $\sqrt{\sum_{z=1}^N f_z^2 b_z / \sum_{z=1}^N b_z}$
Position change parameters of frequency: $\sum_{z=1}^N f_z^2 b_z / \sqrt{\sum_{z=1}^N b_z \sum_{z=1}^N f_z^4 b_z}$	
Hilbert marginal spectrum-based energy entropy: $-\sum_k^n p_k \lg(p_k) \quad p_k = E_k/E, E = \sum_{k=1}^n E_k$	

where N is the number of the marginal spectrum lines of the reconstructed signal obtained from the selected IMFs, f_z is the frequency value of the z th spectrum line, b_z is a marginal spectrum amplitude of the z th spectrum line, n is the number of the selected IMFs, E_k is the calculated Hilbert marginal spectrum-based feature energy of the k th IMF, E is the sum of Hilbert marginal spectrum-based feature energy of all the selected IMFs, p_k is the ratio of the feature energy of the k th IMF and the sum of feature energy of all the selected IMFs.

single time–frequency domain features are not sufficient for fault severity identification [15]. For these reasons, multi-domain feature extractions based on EEMD, including time domain, frequency domain and time–frequency domain, have been investigated.

As is well known, the decomposition results are closely related to the size of ensemble M and the amplitude of the added white noise a in EEMD algorithm. Wu [11] investigated the relationship between the results and the two parameters and proposed the evaluation criteria. Based on this, M and a are set to 100 and 0.1, respectively. After the parameters of EEMD are determined, the collected discharge pressure signals are first processed by EEMD and a set of IMFs is obtained. In our previous research, the analysis results indicate that the energy of the signal is concentrated in the first few IMFs. To select the proper IMFs to extract the fault features, the correlation coefficient between the signal reconstructed by the selected IMFs and the original signal is calculated. The IMFs can then be determined by setting a threshold δ . Based on previous research [7], δ is set to 0.95. The selected IMFs are referred to as EEMD paving in this paper. When a pump fault occurs, it results in some obvious abnormal signal patterns in some frequency bands. Meanwhile, each IMF contains different frequency components, so it is more reasonable to extract the fault features from the EEMD paving. The seven time-domain features listed in Table 1 are first extracted from the paving of EEMD. As a result, a time-domain feature set containing $7n$ features for each sample will be obtained when the number of selected IMFs is n .

For the pump failures, the frequency and energy distribution of the signal may change with the increase of the fault severity. To describe the change of distribution, some fault features should be extracted from frequency and time–frequency domain. The frequency center, root mean square frequency, and position change parameters of frequency can measure the locations of the main frequency components, so these three features have been used in the actual application. However, there are still some limitations. For example, most of these extracted features are based on FFT anal-

ysis, and it is difficult to describe the change accurately because FFT, a typical global transform, is more suitable for processing linear and stable signals. Unlike the traditional Fourier spectrum analysis, Hilbert spectrum of the obtained IMFs represents the time–frequency distribution of the amplitude and the corresponding marginal spectrum can measure total amplitude distribution from each frequency value, so the Hilbert marginal spectrum can describe the change of frequency distribution accurately. Therefore, Hilbert marginal spectrum is introduced to compute the frequency domain fault characteristics. Similarly, the Hilbert marginal spectrum-based energy entropy which can measure the change of energy distribution is used to extract the time–frequency domain features. A more detailed explanation can be found in our previous research [6,7]. The frequency domain and time–frequency domain features are summarized in Table 2.

Based on Table 1 and Table 2, it can be found that the number of the extracted fault features is $(7n + 4)$, where n is the number of the selected IMFs.

2.2. Selection of sensitive fault features

In the previous section, multi-domain fault features based on EEMD paving have been extracted. Though these features have their own particular meanings to characterize the different aspects of a pump's health status, what should be noted is that each feature has different sensitivity contribution for fault recognition. Besides, it has been found in previous studies [16] that too many input parameters may result in low recognition accuracy and increase the computational burden. Accordingly, it is essential to reduce the dimensionality of the input features to obtain most sensitive fault features. As one of the most effective dimensionality reduction methods, PCA has been used successfully in the fault patterns recognition of pump. Hence, PCA is employed in this paper to select the most sensitive features. The detailed steps are given as follows:

- (1) Suppose we have a $Z \times D$ dataset $X = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_Z)^T$ which is the extracted feature vector dataset, where \mathbf{x}_i is a D -dimensionality sample, Z is the number of samples, and D is the dimensionality of each sample. To reduce the effect of different dimension and order of magnitude, the original dataset X is normalized and new dataset is denoted by X^{new} . Then the $D \times D$ correlation coefficient matrix R of the dataset X^{new} is calculated.
- (2) Solve the eigenvalue problem.
- (3) Compute the cumulative contribution rate of the largest m eigenvalues, and the eigenvalues can be selected by setting a proper threshold η . Generally speaking, the satisfactory result can be obtained when η is bigger than 0.85 [32]. Therefore, η is set to 0.95 [19] in this paper.
- (4) Construct the linear mapping matrix Q based on the eigenvectors corresponding to the selected eigenvalues.
- (5) Calculate the low-dimensional data representations Y by using equation $Y = XQ$, and the obtained features matrix will be used as training samples of SVR model.

In order to enhance the generalization ability of the algorithm, testing samples obtained from PCA is straightforward using equation $Y^* = X^*Q$. At this time, the high-dimensional dataset X^* has been transformed to a new low-dimensional dataset Y^* successfully.

2.3. Fault severity recognition based on optimized SVR model

2.3.1. The basic principles of SVR theory

SVR theory is developed in accordance with the principle of SVM and it is available for time series prediction. Given a dataset $\{x_i, y_i\}$, $i = 1, 2, \dots, B$, where x_i represents an input feature vector, y_i is the expected target value, and B denotes the number of training samples. In ε -insensitive support vector regression, the objective is to acquire a function $f(x)$ which can predict the output y_i within the error limit of ε . This regression function can be expressed as follows:

$$f(x) = w \cdot x + b \quad (4)$$

where w denotes the weight vector and b represents the offset value. On the basis of previous research, this problem can be transformed into an optimization problem, where the objective function and constraints are given as follows:

$$\begin{aligned} \min J &= \frac{1}{2} \|w\|^2 + C \sum_{i=1}^B (\xi_i^* + \xi_i) \\ &\left\{ \begin{array}{l} y_i - w \cdot x_i - b \leq \varepsilon + \xi_i \\ w \cdot x_i + b - y_i \leq \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0 \end{array} \right. \end{aligned} \quad (5)$$

ξ_i and ξ_i^* represent the slack variables, C is a positive constant which penalizes the errors larger than $\pm\varepsilon$ using ε -insensitive loss function depicted in Fig. 4.

Based on Fig. 4, this loss function can be defined as follows:

$$|\xi|_\varepsilon = \begin{cases} 0 & \text{if } |\xi| \leq \varepsilon \\ |\xi| - \varepsilon & \text{otherwise} \end{cases} \quad (6)$$

After solving the optimization problem described in Eq. (5), a linear regression function can be obtained as follows:

$$f(x) = \sum_{i=1}^B (\alpha_i - \alpha_i^*) (x_i \cdot x) + b \quad (7)$$

where α_i and α_i^* are the Lagrange multipliers.

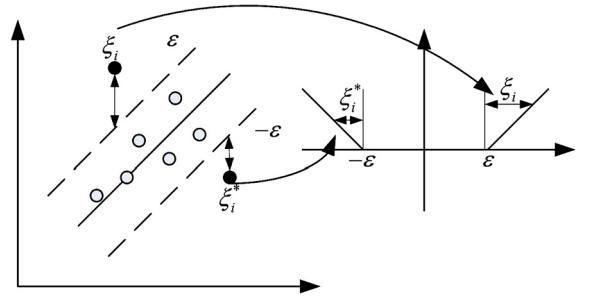


Fig. 4. ε -Insensitive loss function.

In fact, it is difficult to obtain satisfactory results when linear regression function is used to process the non-linear problem. To solve this problem, the kernel function is employed here to map low-dimensional input data into high-dimensional feature space and thus the regression function can be expressed as

$$f(x) = \sum_{i=1}^B (\alpha_i - \alpha_i^*) K(x_i, x) + b \quad (8)$$

where $K(x_i, x)$ is the kernel function.

From Eq. (8), it can be observed that the performance of SVR is closely related to the type of kernel function. Many previous studies have proved that the SVR model can achieve good performance when using radial basis function (RBF) [26–30], so the RBF is adopted in this paper.

2.3.2. The fault severity recognition based on optimized SVR

Given a dataset $\{x_i, y_i\}$, $i = 1, 2, \dots, B$, where x_i is the input multi-domain feature extracted from the outlet pressure signals under different fault sizes of piston pump, y_i is the actual pump fault size and B is the number of training samples. By solving the mathematical problem illustrated in Eq. (5), a SVR model that constructs the relationship between the extracted features and the corresponding fault size is trained and established. Subsequently, the fault size nonlinear analysis function can be described by the following equation:

$$f_{size}(x) = \sum_{i=1}^B (\alpha_i - \alpha_i^*) \exp\left(\frac{-\|x_i - x\|^2}{2\sigma^2}\right) + b \quad (9)$$

A large number of research results have shown that the SVR model performance is closely associated with three key parameters, which are regularization parameter C , kernel function parameter σ and ε -insensitive loss function parameter ε . Consequently, to improve the model's accuracy, three key parameters need to be properly determined. Many methods have been proposed to select proper parameters, such as grid search, GA [33] and PSO [29], among others. Although these methods are available, there still exist some insufficiencies. For instance, grid search-based method can find the optimal parameters, but it requires large enough search space and small enough step, so it will be very time-consuming. In contrast, as one of heuristics methods, GA and PSO are able to obtain the optimal results and they don't need to traverse all parameter groups. Nevertheless, sometimes the algorithm can easily produce local optimum [34]. Thus, a new parameter optimization method is proposed in this section.

The studies indicate that the model's accuracy of grid search will be low in most of search intervals of the parameters C and σ . However, the accuracy turns to be high on the condition that the parameters vary in a specific interval. As a result, if the parameters optimization intervals can be pre-located, there will be a great increase in the search efficiency and the probability of obtaining optimal parameters in grid search. For the purpose of solving this

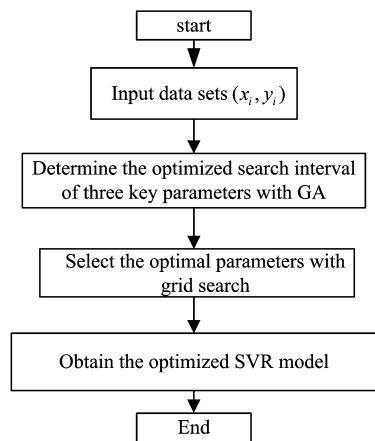


Fig. 5. Flowchart of the parameters selection of optimized SVR model.

problem, GA algorithm is first adopted to determine three parameters defined as first-time optimal parameters due to the strong global optimization ability. To obtain satisfactory results, the search interval of GA is relatively expanded. Meanwhile, considering the algorithm's randomness, algorithm operates h times repeatedly and h first-time optimal parameter sets can be achieved. Based on obtained parameter sets, the final optimal intervals of parameters C and σ can be defined as follows:

$$\begin{aligned} [C_{\min}^*, C_{\max}^*] &= [2^{\lfloor \log_2(C_{\min}) \rfloor}, 2^{\lceil \log_2(C_{\max}) \rceil}] \\ [\sigma_{\min}^*, \sigma_{\max}^*] &= [2^{\lfloor \log_2(\sigma_{\min}) \rfloor}, 2^{\lceil \log_2(\sigma_{\max}) \rceil}] \end{aligned} \quad (10)$$

where C_{\min} and C_{\max} are the minimum and maximum of the parameter C obtained from GA, respectively. Similarly, σ_{\min} and σ_{\max} are the minimum and maximum of the parameter σ obtained from GA, respectively. $\lfloor \cdot \rfloor$ and $\lceil \cdot \rceil$ stand for round down and round up operations to the nearest integer, respectively.

Based upon experimental research, as the parameter ε has a small fluctuation during the process of GA, the optimal parameter of ε can be described as:

$$\varepsilon^* = \frac{\sum_{i=1}^h \varepsilon_i}{h} \quad (11)$$

where ε_i represents the obtained loss function parameter originated from each GA operation.

After the optimized loss function parameter ε^* and the numerical ranges of the parameters C and σ are determined, grid search is then used to choose more reasonable values of C and σ . In grid search method, the search step can be set as small as possible because the search interval has been optimized by GA. Subsequently, the mean square error (MSE) of the predicted outputs based on the popular K-fold cross-validation scheme is employed as a stopping criterion. When calculated MSE satisfies the stopping criterion, the

Table 3
Some parameters of tested pumps.

Parameter	
Number of pistons	11
Rated pressure (MPa)	21
Rated Speed (r/min)	4000
Pressure of zero flow rate (MPa)	21
Pressure of large flow rate (MPa)	20

obtained parameters will be considered as the input parameters of the SVR model. Then, optimized SVR model can be constructed. The flowchart of the proposed optimization strategy is presented in Fig. 5.

3. Results and experimental validation

3.1. Experimental platform

The proposed method was validated on an experimental platform, as shown in Fig. 6. This platform mainly consisted of an AC motor, an actual aviation piston pump whose parameters were shown in Table 3, a gear reducer, a grease pump, and some other auxiliary devices. For the tested piston pump, the defect sizes of loose piston shoes were set to 8 μm, 15 μm, 20 μm, 25 μm. With the objective of studying the applicability of the proposed algorithm under different conditions, two cases including large flow rate and zero flow condition were selected. A pressure sensor, with a range of 0 to 30 MPa and a bandwidth of 0 to 20 kHz, was used to collect the discharge pressure signals of tested pump under two conditions. Data acquisition system composed of an industrial computer and a National Instrument (NI) USB-6221 board. Data acquisition software was developed with NI LabVIEW® 8.6. The data sampling rate was 2 kHz, and each set of data collection lasted for 80 s for each fault size under the same condition. Consequently, 4 groups of pressure signals were collected under each condition. In this paper, the collected dataset of zero flow and large flow rate condition were named dataset 1 and dataset 2, respectively. In our experimentation, the data obtained from each fault size was divided into 50 groups for each condition. Among these 50 groups data, 20 groups were used for training and the remaining were used for testing. The description of the pump data for each dataset was shown in Table 4.

3.2. Experimental results and analysis

The former two pump datasets are analyzed according to the flowchart of the proposed fault severity recognition scheme illustrated in Fig. 1. The signals from two datasets are first decomposed by EEMD, and then a set of IMFs can be obtained. To select sensitive IMFs, the correlation coefficients between the signals reconstructed by selected IMFs and the original signals are calculated.

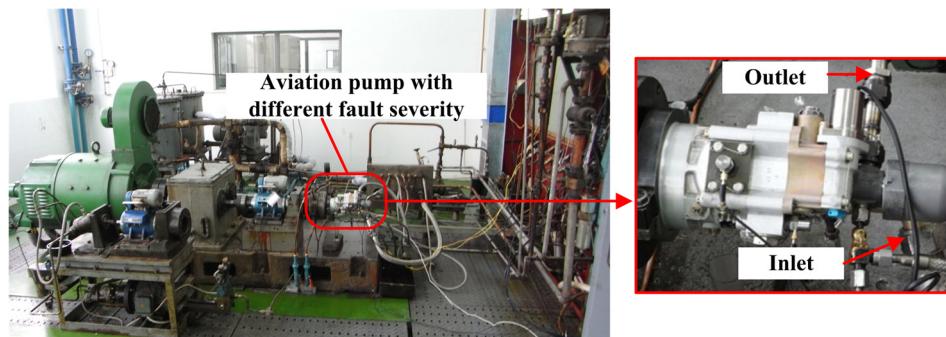


Fig. 6. The aviation hydraulic pump with different fault severity test stand.

Table 4

Description of the pump data for each dataset.

Dataset	Fault size (μm)	Training sets	Testing sets	Target values of SVR model
1	8	20	30	8
	15	20	30	15
	20	20	30	20
	25	20	30	25
	8	20	30	8
2	15	20	30	15
	20	20	30	20
	25	20	30	25

When the first four IMFs are chosen, the obtained correlation coefficients are 0.9617, 0.9728, 0.9570, and 0.9673 for dataset 1. Similarly, the correlation coefficients obtained from dataset 2 can be worked out. Hence, the first four IMFs are used to extract fault features as the computed results have exceeded the pre-set threshold. Based on the proposed feature extraction method, 32 multi-domain features which constitute a feature vector are obtained. Repeating the process of feature extraction, 50 sets of feature vectors are selected for each fault size from each dataset. Subsequently, PCA is applied to reduce the dimensionality of the extracted feature vectors to alleviate the computational burden and select more sensitive features. When applying PCA algorithm, feature vectors obtained from training sets are considered as the original training samples. Fig. 7 shows the cumulative contribution rate of different principal components from two datasets after PCA.

From Fig. 7, it can be seen that the cumulative contribution rate has exceeded the threshold 0.95 when the number of the principal components is 9. Therefore, the obtained 9 principal components are used to replace the original 32 dimensional feature vector. The original training samples with high dimensionality are transformed into the low dimensional dataset for each fault size. In a similar way, the low dimensional dataset of the original testing samples of each fault size can also be acquired. For each fault size, these two low dimensional datasets will be considered as the training and testing samples of SVR model, respectively.

As described in previous section, to achieve the better performance of SVR model, three key parameters C , σ , and ε need to be optimized. Based upon the proposed optimization strategy, GA algorithm is first adopted to determine the interval of searching for the optimal parameters, and some parameters of GA optimization algorithm are presented in Table 5. Then the algorithm is run repeatedly 5 times, and 5 sets of parameters are acquired. On the basis of Eqs. (10)–(11), the parameter ε and the numerical range

Table 5

Some set parameters of GA optimization.

Parameter	The value
The maximum number of generations	200
The maximum number of population	20
Crossover rate	0.7
Mutation rate	0.1
The numerical range of C	[0, 1000]
The numerical range of σ	[0, 100]
The numerical range of ε	[0.001, 1]

Table 6

The results obtained from GA optimization.

Dataset	The optimization interval of C	The optimization interval of σ	The value of ε
1	[$2^0, 2^6$]	[$2^{-3}, 2^0$]	0.00104
2	[$2^{-1}, 2^5$]	[$2^{-3}, 2^{-1}$]	0.00108

of C and σ can be determined. The results are summarized in Table 6. Compared with Table 5, it can be observed that the interval of searching for the optimal parameters shown in Table 6 has been greatly reduced.

After the optimization intervals are selected, grid search is adopted to choose the optimal parameter values of C and σ . For obtaining more reasonable results, the search step of C and σ are set to 0.1 and 0.01, respectively. Besides, 5-fold cross validation is also used in grid search. Subsequently, the optimal parameters $C = 1$, $\sigma = 0.4796$ from dataset 1 and $C = 1$, $\sigma = 0.5$ from dataset 2 are respectively obtained. With the optimal parameters being determined, the optimal SVR model is constructed.

Fig. 8 and Fig. 9 illustrate the training and testing results obtained from the constructed SVR model for dataset 1 and dataset 2, respectively. As indicated in Fig. 8 and Fig. 9, it can be observed that the model can recognize the fault sizes effectively and the deviation of the predicted values and actual values is very small for both training samples and testing samples.

To evaluate the performance of the constructed SVR model quantitatively, some statistical indexes which include the maximum relative error (MRE), average relative error (ARE), and mean square error (MSE) are given as:

$$\begin{aligned} \text{MRE} &= \max(|f_{\text{predicted}} - f_{\text{actual}}|/f_{\text{actual}}) \\ \text{ARE} &= \frac{1}{H} \sum_{i=1}^H (|f_{\text{predicted}} - f_{\text{actual}}|/f_{\text{actual}}) \end{aligned} \quad (12)$$

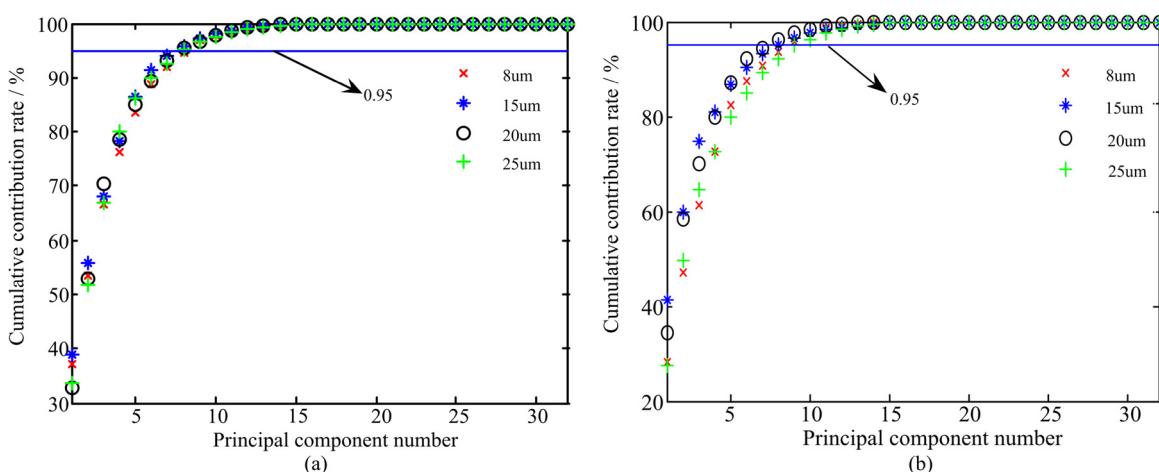


Fig. 7. Cumulative contribution rate of different principal components from (a) dataset 1 and (b) dataset 2.

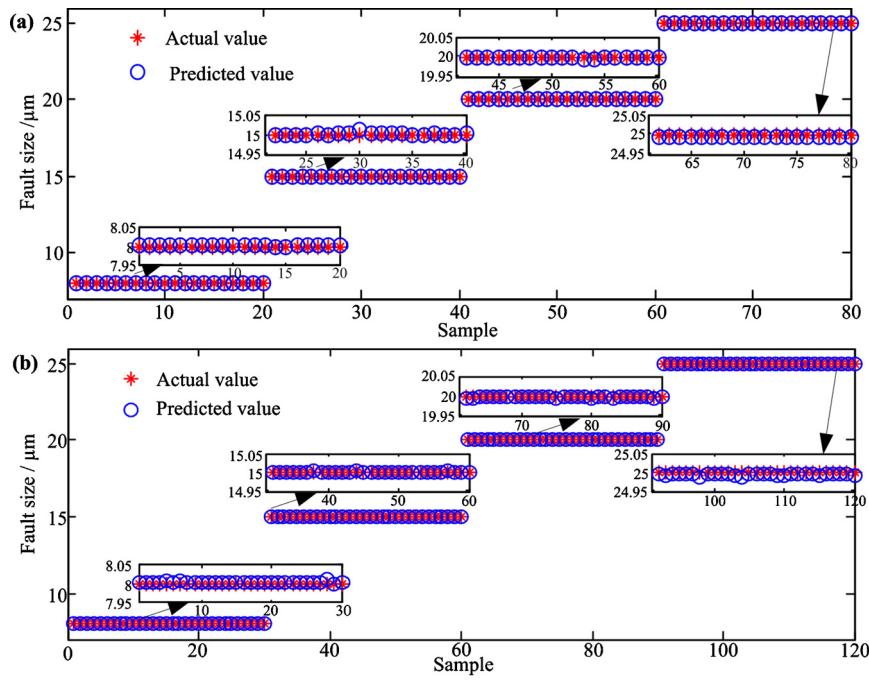


Fig. 8. For dataset 1, the predicted fault sizes of the (a) training samples and (b) testing samples.

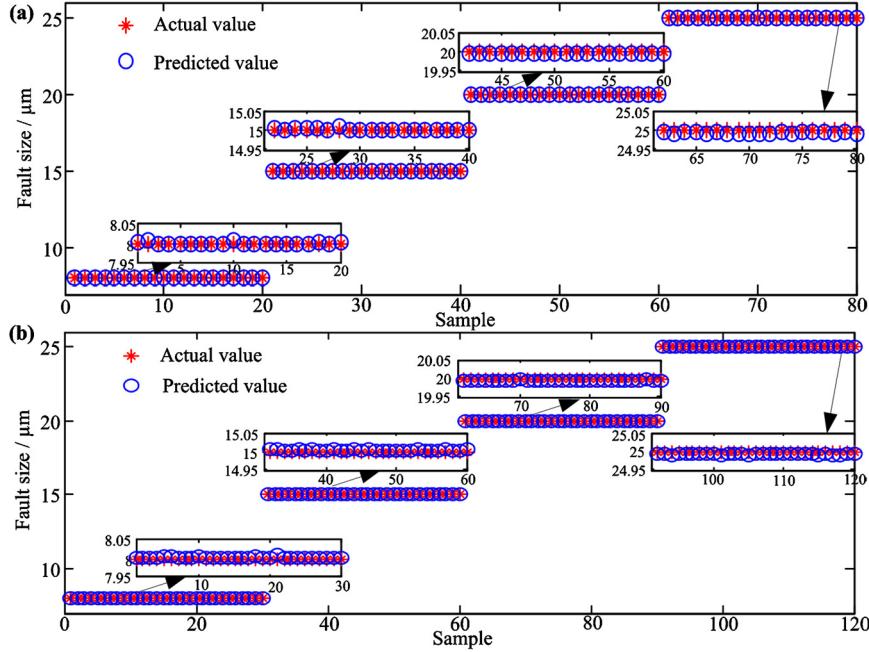


Fig. 9. For dataset 2, the predicted fault sizes of the (a) training samples and (b) testing samples.

$$\text{MSE} = \frac{1}{H} \sum_{i=1}^H (f_{\text{predicted}} - f_{\text{actual}})^2$$

where H is the number of the training or testing samples, $f_{\text{predicted}}$ is the predicted fault size, and f_{actual} is the actual fault size.

Table 7 presents the calculated statistical indexes of SVR model. In **Table 7**, the errors between the actual and predicted values are very small. Moreover, the errors for the predicted results of training samples are nearly equal to those of testing samples. Hence, it can be concluded that the constructed SVR model has strong generalization ability and the proposed SVR-based identification method achieves great performance.

Table 7
The obtained statistical indexes of constructed SVR model.

Dataset	Samples	MRE	ARE	MSE
1	Training	0.0007	0.00024	0.0000167
	Testing	0.0016	0.00025	0.000019
2	Training	0.0009	0.00028	0.0000195
	Testing	0.0011	0.00028	0.0000214

4. Comparison with other methods

The experimental results demonstrate that the proposed method as a systematic scheme can recognize pump fault sizes with high

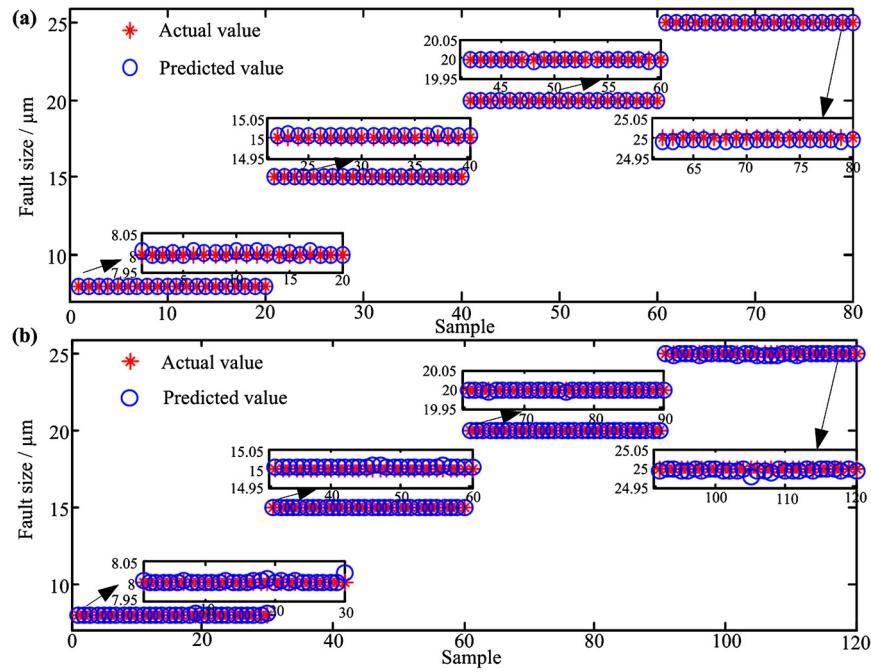


Fig. 10. For dataset 1, the predicted fault sizes of the (a) training samples and (b) testing samples based on EMD with optimized SVR.

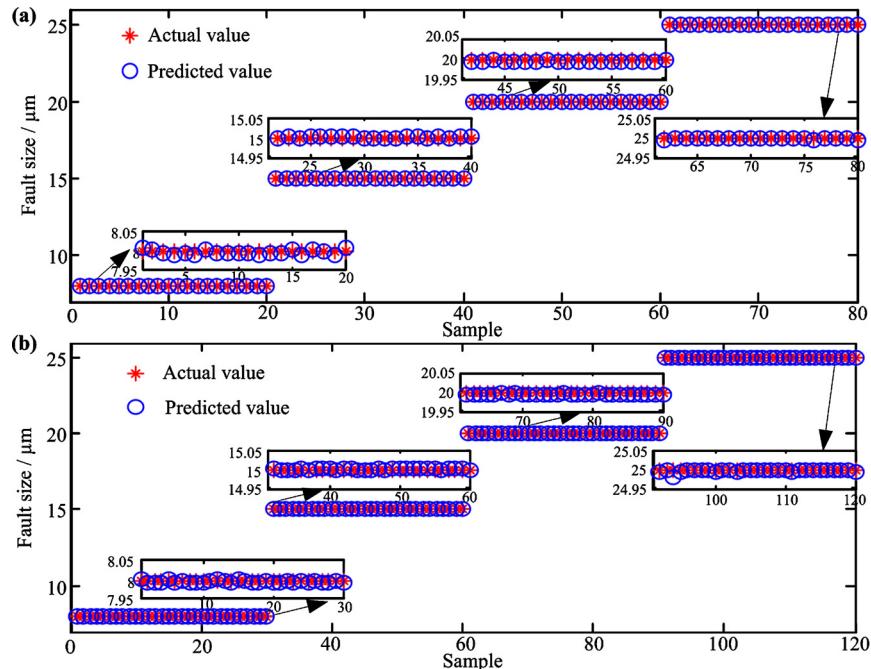


Fig. 11. For dataset 2, the predicted fault sizes of the (a) training samples and (b) testing samples based on EMD with optimized SVR.

accuracy under different conditions. To prove its superiority over other methods, some comparisons are presented in this section. The comparisons are conducted considering two aspects. On one hand, for the purpose of verifying the advantages of the proposed feature extraction method, it is compared with the EMD-based feature extraction. On the other hand, BP and traditional grid search-based SVR (GS-SVR) methods which can be used to predict the fault size are also selected for a comparison.

Fig. 10 and Fig. 11 present the predicted results of samples based on EMD and optimized SVR for dataset 1 and dataset 2, respectively. Comparing Figs. 10–11 with Figs. 8–9, it can be seen that the results shown in Figs. 10–11 fluctuate more violently than those shown in Figs. 8–9. To provide more details

Table 8
The obtained statistical indexes based on EMD and optimized SVR.

Dataset	Samples	MRE	ARE	MSE
1	Training	0.001	0.00026	0.0000214
	Testing	0.0031	0.00024	0.0000256
2	Training	0.0011	0.0003	0.0000217
	Testing	0.0011	0.00028	0.0000236

for this difference, some statistical indexes obtained from EMD and optimized SVR model are calculated as presented in Table 8. In comparison with Table 8, we have found that the values of these indexes shown in Table 7 are better. The comparison results

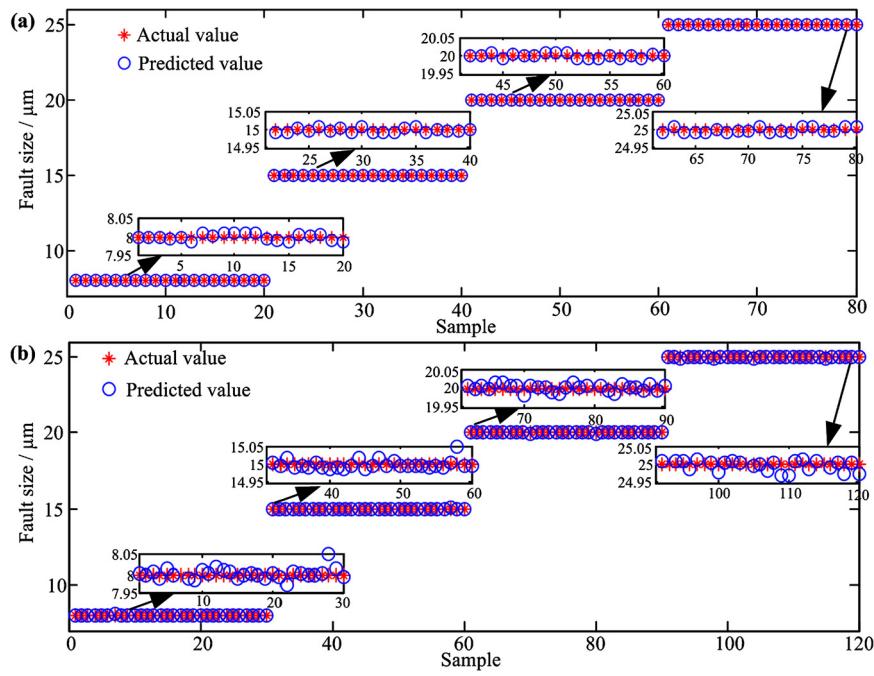


Fig. 12. For dataset 1, the predicted fault sizes of the (a) training samples and (b) testing samples based on EEMD paving and GS-SVR.

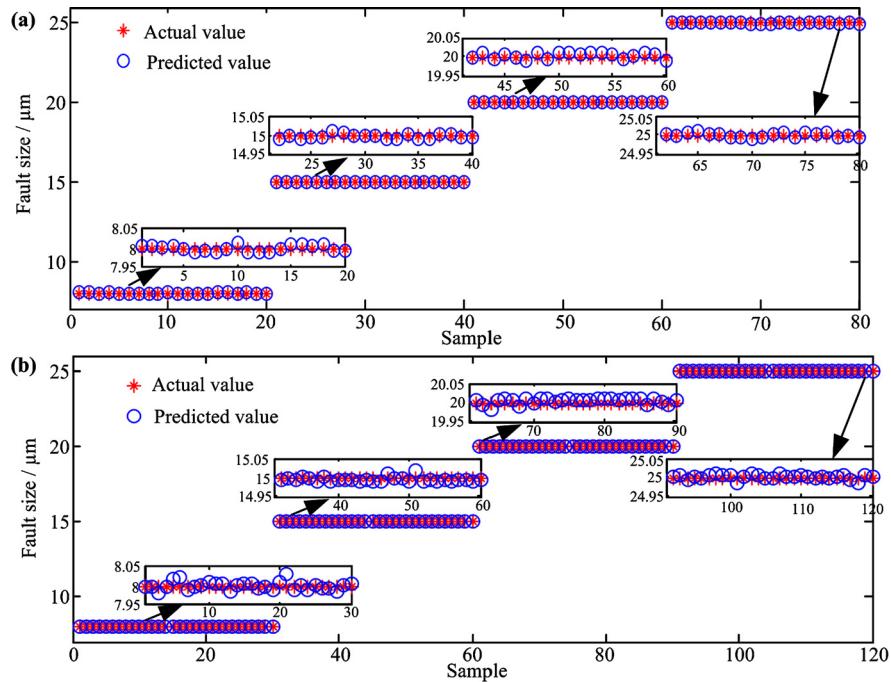


Fig. 13. For dataset 2, the predicted fault sizes of the (a) training samples and (b) testing samples based on EEMD paving and GS-SVR.

prove that the extracted fault features obtained from the proposed method can characterize the health status of the pump more accurately.

When finishing the comparisons of the feature extraction methods, the popular GS-SVR and BP methods are used for a comparison with the proposed SVR model. To summarize the comparison results, only the final statistical indexes of predicted results based on EMD and two comparison methods are reported in the following paragraphs. As presented previously, GS-SVR has been successfully applied in the fault diagnosis field. In this study, $C = [2^{-10}, 2^7]$, $\sigma = [2^{-5}, 2^5]$, and $\varepsilon = 0.001$ are selected as the numerical ranges for the parameters grid search [26].

Fig. 12 and **Fig. 13** depict the predicted results of the training and testing samples from dataset 1 and dataset 2. As illustrated in **Fig. 12** and **Fig. 13**, the predicted fault sizes of testing samples fluctuate more violently around the actual fault sizes than those of training samples. Especially, from **Fig. 12(b)**, it can be seen that the maximum error for the predicted results of testing samples has reached $0.05 \mu\text{m}$. Meanwhile, comparing **Figs. 8–9** with **Figs. 12–13**, we can find that the errors of predicted results based on EEMD paving and GS-SVR are bigger than those based on the proposed method. It indicates that the generalization ability of the constructed SVR model is better and thus higher prediction accuracy can be achieved.

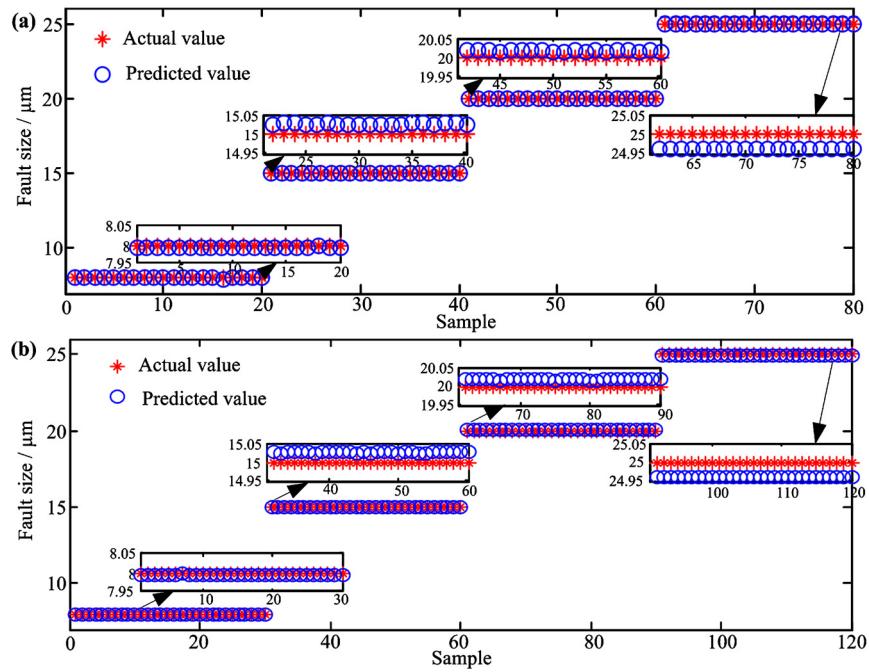


Fig. 14. For dataset 1, the predicted fault sizes of the (a) training samples and (b) testing samples based on EEMD paving and BP.

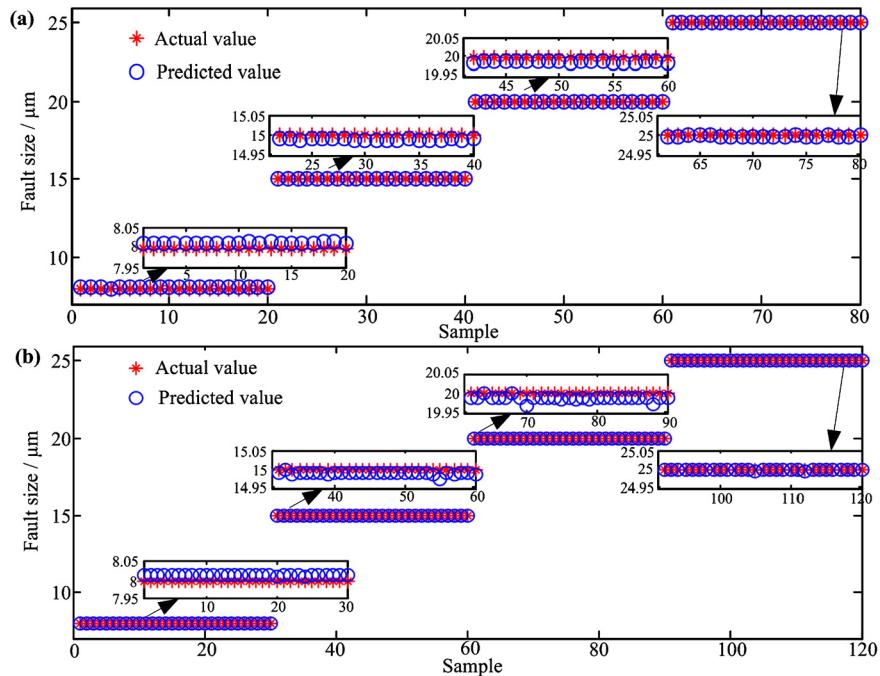


Fig. 15. For dataset 2, the predicted fault sizes of the (a) training samples and (b) testing samples based on EEMD paving and BP.

After comparing with GS-SVR, BP algorithm is employed here for further comparison. In BP method of this study [29], the Trainlm training function and Tansig transfer function are adopted. The number of the maximum training steps is 1000; the learning rate of BP network is set to 0.01; the goal of training error is set to 0.0001; the number of hidden nodes is 10. Subsequently, BP method is used to estimate the fault sizes of pump. The results obtained by the BP method for dataset 1 and dataset 2 are illustrated in Fig. 14 and Fig. 15, respectively.

Fig. 14 clearly indicates that there is a big deviation between the predicted fault sizes and actual fault sizes especially when the

fault size is 25 μm compared with the results shown in Fig. 8. The same conclusions can be obtained by comparing Fig. 9 with Fig. 15.

To quantify the accuracy of the compared methods, the statistical indexes obtained in BP and GS-SVR using EEMD and EMD extraction are summarized in Table 9 and Table 10. Comparing Table 9 and Table 7, the values of these indexes obtained from the proposed method are better than those obtained from BP and GS-SVR methods. From Table 10 and Table 8, it can also be seen that the errors of predicted values and actual values based on optimized SVR are smaller than those based on BP and GS-SVR when using EMD extraction in two datasets. Thus, it can be concluded that the optimized SVR model can capture the relationship be-

Table 9

The statistical indexes obtained by BP and GS-SVR using EEMD extraction.

Dataset	Samples	GS-SVR method			BP method		
		MRE	ARE	MSE	MRE	ARE	MSE
1	Training	0.0017	0.00046	0.00005	0.0037	0.0037	0.0007
	Testing	0.0016	0.00088	0.00039	0.0037	0.0009	0.0007
2	Training	0.0018	0.00046	0.00005	0.0029	0.0029	0.0004
	Testing	0.0037	0.00053	0.00007	0.0028	0.00087	0.00038

Table 10

The statistical indexes obtained by BP and GS-SVR using EMD extraction.

Dataset	Samples	GS-SVR method			BP method		
		MRE	ARE	MSE	MRE	ARE	MSE
1	Training	0.0021	0.0005	0.00006	0.0043	0.0046	0.0034
	Testing	0.0083	0.00063	0.00015	0.0043	0.0017	0.0034
2	Training	0.0018	0.00053	0.00006	0.0037	0.0037	0.00066
	Testing	0.0059	0.00062	0.00029	0.0037	0.00094	0.00065

tween the extracted features and fault sizes better. In addition, we can find that most of the indexes of the samples based on EEMD extraction shown in Table 9 are better than those obtained from EMD extraction presented in Table 10 when using BP and GS-SVR. The results once again prove that EEMD-based feature extraction method is more effective.

5. Conclusions

In this study, a new effective fault severity recognition approach has been proposed. Unlike the traditional fault identification methods reported in many other studies, the main objective of this study is the development of the methods which can reliably estimate the fault sizes of the aviation piston pump. Because the deterioration process of pump is progressive, the proposed method is very important in this industrial application as it enables the determination of the critical fault occurrence accurately and to properly plan the preventive maintenance. According to the aforementioned illustration, the proposed scheme involves multi-domain fault features extraction based on EEMD paving from discharge pressure signals, a PCA for sensitive features selection, and a prediction of fault sizes of pump based on optimized SVR. Based on the experimental results and the comparison with previous approaches, we can conclude the following:

- (1) The extracted multi-domain fault features based on EEMD paving can characterize the fault severity of pump more accurately than the obtained features based on EMD.
- (2) The proposed optimized SVR model can recognize the fault sizes of pump effectively with high accuracy under different conditions, which can help to track the health status of pump.
- (3) The proposed method can be generalized and used in many other industrial applications and it has higher prediction accuracy than the traditional methods, such as GS-SVR and BP methods.
- (4) In the future, some research will be done on how to obtain the remaining useful life (RUL) prediction for aviation hydraulic pump on the basis of the proposed fault severity recognition.

Conflict of interest statement

The authors declare that there is no conflict of interest regarding the publication of this paper.

Acknowledgements

This research was supported by the National Basic Research and Development Program of China (Grant No. 2014CB046402),

the Natural Science Foundation of China (Grant No. 51575019, 51305011) and 111 Program of China. The authors gratefully acknowledge financial support from China Scholarship Council (award to Chuanqi Lu for one year's study abroad at the University of Toronto).

References

- [1] X.J. Wang, S.R. Lin, S.P. Wang, et al., Remaining useful life prediction based on the Wiener process for an aviation axial piston pump, *Chin. J. Aeronaut.* 29 (3) (2015) 779–788.
- [2] Z.P. Wang, C. Lu, Z.L. Wang, Chaotic information-geometric support vector machine and its application to fault diagnosis of hydraulic pumps, *J. Vibroeng.* 16 (2) (2014) 1033–1041.
- [3] L. Han, S.P. Wang, C. Zhang, A partial lubrication model between valve plate and cylinder block in axial piston pumps, *Proc. Inst. Mech. Eng., Part C, J. Mech. Eng. Sci.* 229 (17) (2015) 3201–3217.
- [4] J. Du, S.P. Wang, H.Y. Zhang, Layered clustering multi-fault diagnosis for hydraulic piston pump, *Mech. Syst. Signal Process.* 36 (2) (2013) 487–504.
- [5] Z. Zhao, M.X. Jia, F.L. Wang, et al., Intermittent chaos and sliding window symbol sequence statistics-based early fault diagnosis for hydraulic pump on hydraulic tube tester, *Mech. Syst. Signal Process.* 23 (5) (2009) 1573–1585.
- [6] C.Q. Lu, S.P. Wang, C. Zhang, Fault diagnosis of hydraulic piston pumps based on a two-step EMD method and fuzzy C-means clustering, *Proc. Inst. Mech. Eng., Part C, J. Mech. Eng. Sci.* 230 (16) (2016) 2913–2928.
- [7] C.Q. Lu, S.P. Wang, M. Tomovic, Fault severity recognition of hydraulic piston pumps based on EMD and feature energy entropy, in: *IEEE 10th International Conference on Industrial Electronics and Applications*, Auckland, 15–17 June, 2015, pp. 489–494.
- [8] Y.J. Gao, Q. Zhang, A wavelet packet and residual analysis based method for hydraulic pump health diagnosis, *Proc. Inst. Mech. Eng., Part D, J. Automob. Eng.* 220 (6) (2006) 735–745.
- [9] N.E. Huang, Z. Shen, S.R. Long, et al., The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis, *Proc. R. Soc. Lond.* 454 (1998) 903–995.
- [10] Y.G. Lei, J. Lin, Z.J. He, et al., A review on empirical mode decomposition in fault diagnosis of rotating machinery, *Mech. Syst. Signal Process.* 35 (2013) 108–126.
- [11] Z.H. Wu, N.E. Huang, Ensemble empirical mode decomposition: a noise assisted data analysis method, *Adv. Adapt. Data Anal.* 1 (1) (2009) 1–41.
- [12] Y. Amirat, V. Choqueuse, M. Benbouzid, EEMD-based wind turbine bearing failure detection using the generator stator current homopolar component, *Mech. Syst. Signal Process.* 41 (1) (2013) 667–678.
- [13] Z. He, Y. Shen, Q. Wang, et al., Optimized ensemble EMD-based spectral features for hyperspectral image classification, *IEEE Trans. Instrum. Meas.* 63 (5) (2014) 1041–1056.
- [14] Z.K. Liu, Y.H. Liu, H.K. Shan, et al., A fault diagnosis methodology for gear pump based on EEMD and Bayesian network, *PLoS ONE* 10 (5) (2015) 0125703.
- [15] S. Hong, Z. Zhou, Zio Enrico, et al., Condition assessment for the performance degradation of bearing based on a combinatorial feature extraction method, *Digit. Signal Process.* 27 (2015) 159–166.
- [16] C.Q. Shen, D. Wang, F.R. Kong, P.W. Tse, Fault diagnosis of rotating machinery based on the statistical parameters of wavelet packet paving and a generic support vector regressive classifier, *Measurement* 46 (4) (2013) 1551–1564.
- [17] P. Bect, Z. Simeu-Abazi, P.L. Maisonneuve, Diagnostic and decision support systems by identification of abnormal events: application to helicopters, *Aerosp. Sci. Technol.* 46 (2015) 339–350.

- [18] T. Yang, H. Pen, Z. Wang, et al., Feature knowledge based fault detection of induction motors through the analysis of stator current data, *IEEE Trans. Instrum. Meas.* 65 (3) (2016) 549–558.
- [19] I. Jaffel, O. Taouali, M.F. Harkat, et al., Kernel principal component analysis with reduced complexity for nonlinear dynamic process monitoring, *Int. J. Adv. Manuf. Technol.* (2016), <http://dx.doi.org/10.1007/s00170-016-8987-4>.
- [20] X. Wan, D. Wang, P.W. Tse, et al., A critical study of different dimensionality reduction methods for gear crack degradation assessment under different operating conditions, *Measurement* 78 (2016) 138–150.
- [21] J.B. Ali, N. Fnaiech, L. Saidi, et al., Application of empirical mode decomposition and artificial neural network for automatic bearing fault diagnosis based on vibration signals, *Appl. Acoust.* 89 (2015) 16–27.
- [22] L.L. Cui, C.Q. Ma, F.B. Zhang, et al., Quantitative diagnosis of fault severity trend of rolling element bearings, *Chin. J. Mech. Eng.* 28 (6) (2015) 1254–1260.
- [23] A. Widodo, B.S. Yang, Support vector machine in machine condition monitoring and fault diagnosis, *Mech. Syst. Signal Process.* 21 (6) (2007) 2560–2574.
- [24] V.N. Vapnik, *The Nature of Statistical Learning Theory*, Springer, Berlin, 1995.
- [25] A.J. Smola, B. Scholkopf, A tutorial on support vector regression, *Stat. Comput.* 14 (3) (2004) 199–222.
- [26] C.Q. Shen, D. Wang, Y.B. Liu, et al., Recognition of rolling bearing fault patterns and sizes based on two-layer support vector regression model, *Smart Struct. Syst.* 13 (3) (2014) 453–471.
- [27] A. Soualhi, K. Medjaher, N. Zerhouni, Bearing health monitoring based on Hilbert–Huang transform, support vector machine, and regression, *IEEE Trans. Instrum. Meas.* 64 (1) (2015) 52–62.
- [28] H.M. Liu, J.C. Zhang, C. Lu, Performance degradation prediction for a hydraulic servo system based on Elman network observer and GMM–SVR, *Appl. Math. Model.* 39 (19) (2015) 5882–5895.
- [29] S.J. Dong, T.H. Luo, Bearing degradation process prediction based on the PCA and optimized LS-SVM model, *Measurement* 46 (2013) 3143–3152.
- [30] X.Y. Zhang, Y.T. Liang, J.Z. Zhou, et al., A novel bearing fault diagnosis model integrated permutation entropy, ensemble empirical mode decomposition and optimized SVM, *Measurement* 69 (2015) 164–179.
- [31] M. Grasso, P. Pennacchi, B.M. Colosimo, Empirical mode decomposition of pressure signal for health condition monitoring in waterjet cutting, *Int. J. Adv. Manuf. Technol.* 72 (2014) 347–364.
- [32] Y.G. Lei, M.J. Zuo, Gear crack level identification based on weighted K nearest neighbor classification algorithm, *Mech. Syst. Signal Process.* 23 (5) (2009) 1535–1547.
- [33] C. Ding, J.P. Xu, L. Xu, ISHM-based intelligent fusion prognostics for space avionics, *Aerosp. Sci. Technol.* 29 (2013) 200–205.
- [34] A. Sarosh, Y.F. Dong, The GA-ANN expert system for mass-model classification of TSTO surrogates, *Aerosp. Sci. Technol.* 48 (2016) 146–157.