Research Article
An Effective Fault Feature Extraction Method for Gas Turbine Generator System Diagnosis

Jian-Hua Zhong,[1] JieJunYi Liang,[2] Zhi-Xin Yang,[1] Pak Kin Wong,[1] and Xian-Bo Wang[1]

[1]*Department of Electromechanical Engineering, University of Macau, Macau*
[2]*School of Electrical, Mechanical and Mechatronic Systems, University of Technology Sydney, Sydney, NSW 2007, Australia*

Correspondence should be addressed to Zhi-Xin Yang; zxyang@umac.mo

Fault diagnosis is very important to maintain the operation of a gas turbine generator system (GTGS) in power plants, where any abnormal situations will interrupt the electricity supply. The fault diagnosis of the GTGS faces the main challenge that the acquired data, vibration or sound signals, contain a great deal of redundant information which extends the fault identification time and degrades the diagnostic accuracy. To improve the diagnostic performance in the GTGS, an effective fault feature extraction framework is proposed to solve the problem of the signal disorder and redundant information in the acquired signal. The proposed framework combines feature extraction with a general machine learning method, support vector machine (SVM), to implement an intelligent fault diagnosis. The feature extraction method adopts wavelet packet transform and time-domain statistical features to extract the features of faults from the vibration signal. To further reduce the redundant information in extracted features, kernel principal component analysis is applied in this study. Experimental results indicate that the proposed feature extracted technique is an effective method to extract the useful features of faults, resulting in improvement of the performance of fault diagnosis for the GTGS.

1. Introduction

The gas turbine generator system (GTGS) is important equipment in power plants. This system was designed to operate 24 hours a day. Any abnormal situations of the GTGS will reduce the electric supply, resulting in enormous economic loss. The traditional manual inspection faces the challenge that the raw acquired signals contain the amount of mixed and redundant information causing the difficulty in accomplishing the fault-monitoring task. To ensure the smooth operation in power plants, development of a suitable fault diagnosis system based on an effective signal preprocessing method for the GTGS is necessary. Many researchers proposed the various methods for doing fault diagnosis of the rotating machinery such as the GTGS [1, 2]. Analysis of the signals is one of the most important means for condition monitoring and fault diagnosis of the rotating machinery. These methods for fault diagnosis of the rotating machinery usually rely on the procedures of (1) signal processing and (2) fault identification/classification [3–6].

In terms of signal processing, vibration signals are usually employed to analyze the rotating machinery faults because the vibration signals are easy to acquire and highly correlate with the condition of the rotating machinery [7, 8]. However, the vibration signals contain high-dimensional data and are enclosed by a lot of irrelevant and redundant information, which degenerate the accuracy as well as the fault identification time of the diagnostic system [9, 10]. To solve the problem, a proper feature extraction technique which can extract useful information of faults from the vibration signal is desirable. At present, there exist many methods to extract the features of faults from vibration signals, such as Fourier transform, short time Fourier transform, and wavelet transform. Fourier transform is only suitable to analyze the linear or stationary signal. However, the signal of the GTGS is characterized by nonstationary signal, which makes Fourier transform unsuitable in this study. The time-frequency analysis methods, short time Fourier transform (STFT) [11] and wavelet transform [12], can process the nonlinear or nonstationary signal. However, STFT has a limitation in

nonstationary signal processing because of using the fixed time windows where it is impossible to achieve good resolution in the time and frequency domain at the same time. As known, the different fault conditions contain various amplitude- and phase-frequency characteristics in the frequency domain. In other words, fault features in some frequency-bands may be enhanced, while restrained in the others. It should be reasonable to assume that there are certain corresponding relationships between the fault feature changes in the frequency-bands and the fault phenomena. Wavelet analysis is proposed to well solve the above problem since it decomposes the signal into the various frequency-bands. To date, various types of wavelet basis have been proposed. Wavelet packet transform (WPT) is a popular and effective tool to decompose the fault signals of the rotating machinery [13, 14]. Compared with the other wavelet transforms, WPT not only iteratively decomposes the approximation signals of lower frequency, but further decomposes the detail signals of higher frequency iteratively as well. To decompose the signal into more details of frequency-bands, WPT is employed in this study. Even though WPT decomposes the signal into different frequency-bands, the data size of different frequency-bands is the same as the raw data, which is usually very big size. Therefore, a proper feature extraction method should be also considered in the phase of data processing to determine representative features from different frequency-bands so as to reduce the input dimension of the classifier. It is well known that the accuracy of the classifier will be degraded if the classifier has many input variables. In the open literatures, to extract the effective fault features from the different frequency-bands, time-domain statistical feature (TDSF) is usually considered [15, 16]. In this paper, WPT combined with TDSF is proposed as the feature extraction method to extract useful fault features from the different frequency-bands.

In practice, the accuracy of fault diagnosis is dependent on the amount of useful information of faults. It means that the more the information of the faults is, the higher the accuracy of fault diagnosis is. To acquire more information of faults, two accelerometers are employed to simultaneously record the vibration signals from the axial and vertical direction. However, it causes that there is still some irrelevant and redundant information in the combined feature after feature is extracted. In order to resolve this problem, a feature selection method should be employed to further wipe off irrelevant and redundant information such that the amount of useless data can be reduced to improve the diagnostic accuracy. The available feature selection approaches include compensation distance evaluation technique (CDET) [16], kernel principal component analysis (KPCA) [17, 18], and genetic algorithm (GA) [19, 20] based methods. Although CDET and GA based methods provide a good solution, the optimal threshold in CDET is difficult to set and the result of GA is unrepeatable. In other words, when a GA is run for two times, two different results will be obtained. In this way, KPCA is considered in this study.

By reviewing the open literatures [15, 21, 22], support vector machine (SVM), relevance vector machine (RVM), back-propagation (BP), and multilayer perception (MLP) are frequently used classifiers to diagnose the rotating machinery faults and other engineering diagnostic applications. In terms of RVM, it is difficult to deal with the problem of large-scale data, like the issue of multiple signal diagnosis presented in this paper. The main reason is that the computational complexity of Hessian matrix attains $O(N^2)$, where N is the number of training data. By contrast with BP and MLP method, support vector machine (SVM) has been successfully applied in pattern classification, due to its high generalization. In this case study, to verify the efficiency of the proposed feature extraction method, the traditional machine learning method, SVM, is considered for the fault identification/classification.

In this paper, to extract the features of faults from the different frequency-bands, an effective feature extraction method which combined wavelet packet transforms (WPT) with time-domain statistical features (TDSF) is proposed. However, multisignal sources cause the extracted fault feature to contain much redundant information which will degrade the diagnosis performance in terms of time and accuracy. Hence, kernel principal component analysis (KPCA) is applied to obtain principal component features which contain most of useful fault information. Finally, a simple and effective machine learning method, SVM, is employed to verify the effectiveness of the proposed feature extraction method.

This paper is organized as follows. Section 2 presents the proposed diagnostic framework and the techniques involved in the framework. Experimental setup and sample data acquisition with a simulated GTGS are discussed in Section 3. Section 4 discusses the experimental results. Finally, a conclusion is given in Section 5.

2. Proposed Framework and Relative Approaches

The flowchart of the proposed diagnostic framework for the rotating machinery, GTGS, is shown in Figure 1. The framework consists of three submodules: (1) data processing; (2) selection of model parameters; (3) performance evaluation.

In the data processing submodule, WPT and TDSF are employed to extract the features from sample dataset that is noted D_{WT}. Considering the existence of irrelevant and redundant information in the extracted features, KPCA is then applied to remove useless information and further reduce the dimensions of D_{WT}. In order to ensure all the features are having even contribution, every feature in D_{KPCA} is normalized with [0, 1]. To train and testify the effectiveness of the proposed framework, normalized data is divided into the training dataset, validation dataset, and unseen signal for testing which are named as $D_{\text{Proc-Train}}$, $D_{\text{Proc-Valid}}$, and $D_{\text{Proc-Test}}$, respectively. The diagnostic model, using support vector machines (SVM), is trained based on the processed training dataset $D_{\text{Proc-Train}}$. The output of the trained classifier, together with processed validation dataset $D_{\text{Proc-Valid}}$, is hired to perform parameters optimization of KPCA and SVM. The optimized parameters are used to construct the SVM-based diagnostic model. Finally, $D_{\text{Proc-Test}}$ is used to evaluate the performance of proposed framework. The details of the three
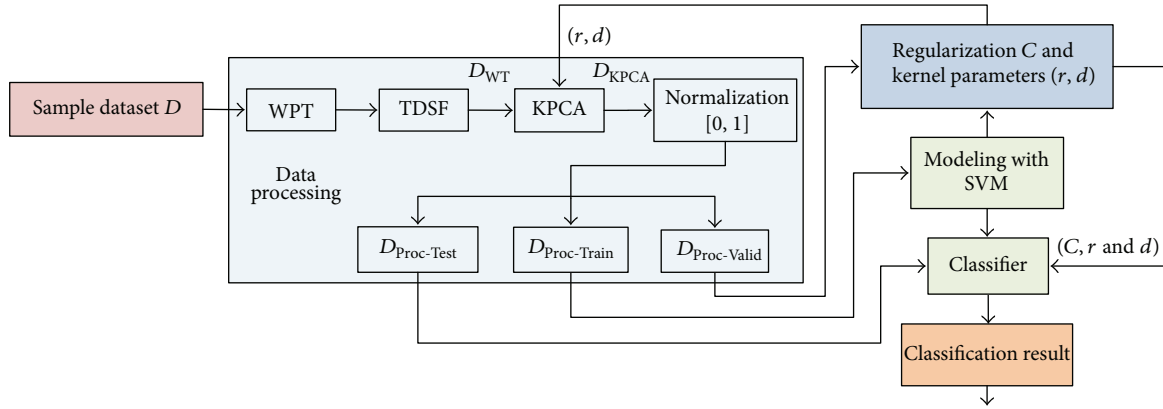
FIGURE 1: The flowchart of the proposed framework.

submodules in the framework are discussed in the following subsections.

2.1. Kernel Principal Component Analysis. Principal component analysis (PCA) is a popular statistical method for analyzing the principal component of information. PCA always performs well in dimensionality reduction when the input variables are linearly correlated. However, for nonlinear cases, PCA usually cannot give good performance. Hence, PCA is extended to nonlinear version which is called kernel PCA (KPCA), which has been used to solve many application problems [23]. KPCA involves solving α in the following set of equations:

$$\theta_{i,j}\alpha = \lambda\alpha, \tag{1}$$

where $\theta_{i,j} = K(D_{\mathrm{WT},i}, D_{\mathrm{WT},j})$ for i or $j = 1, \ldots, N$, N is the number of data for KPCA, $D_{\mathrm{WT},i}, D_{\mathrm{WT},j} \in R^c$, and c is the dimension of the input data. The vector $\alpha = [\alpha_1, \ldots, \alpha_Q]^T$ is the eigenvector of $\theta_{i,j}$ and $\lambda \in R$ is the corresponding eigenvalue. The transformed variables (score variables) z_r for vector D_{WT} become

$$z_r(D_{\mathrm{WT}}) = \sum_{i=1}^{N} \alpha_{r,i} K(D_{\mathrm{WT},i}, D_{\mathrm{WT}}), \tag{2}$$

where $\alpha_{r,i}$ is ith element in the eigenvector α corresponding to rth largest eigenvalue, $r = 1$ to p, and p is the largest number such that eigenvalue λ_p of the eigenvector α_p is nonzero. Therefore, based on the p pairs of (α_r, λ_r), the input vector $D_{\mathrm{WT}} \in R^c$ can be transformed to a nonlinearly uncorrelated variable $z = [z_1, \ldots, z_p]$, where $c > p$. One more point to note is that the eigenvectors α_r should satisfy the normalization condition of unit length:

$$\lambda_r \alpha_r^T \alpha_r = 1, \quad r = 1, 2, \ldots, p, \tag{3}$$

where $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_p > 0$. To produce a further reduced feature vector, a postpruning procedure can be done. That is, after acquiring the p pairs of (α_r, λ_r), all λ_r are normalized to λ_r' which satisfy the constraint $\sum_{r=1}^{p} \lambda_r' = 1$. Based on these normalized eigenvalues λ_r', the smallest λ_r' are deleted

until $\sum_{r=1}^{l} \lambda_r' \leq 0.95$. With the index $l < p$, the eigenvectors α_r ($r = 1$ to l) are selected to produce a reduced feature vector which retains 95% of the information content in the transformed features. Usually a 5% of information loss is a rule of thumb for dimensionality reduction.

2.2. Multiclass Strategies for Classifier. The traditional classifiers and SVM are designed for the issue of the binary classification. However, most of the practical problems are multiclass classifications. Usually, one-versus-all strategy is employed to deal with multiclass classification problems [24]. One-versus-all strategy constructs a group of classifiers $\mathbf{C}_{\mathrm{group}} = [C_1, C_2, \ldots, C_d]$ in a d-label classification problem. The one-versus-all strategy is simple and easy to implement. However, it generally gives a poor result [25, 26] since one-versus-all does not consider the pairwise correlation and hence induces a much larger indecisive region than one-versus-one strategy as shown in Figure 2. One-versus-one strategy also constructs a group of classifiers $\mathbf{C}_{\mathrm{group}} = [C_1, C_2, \ldots, C_d]$ in a d-label classification problem. However, each $\mathbf{C}_i = [C_{i1}, \ldots, C_{ij}, \ldots, C_{id}]$ is composed of a set of $d - 1$ different pairwise classifiers C_{ij}, $i \neq j$. Since C_{ij} and C_{ji} are complementary, there are totally $d(d-1)/2$ classifiers in $\mathbf{C}_{\mathrm{group}}$ as shown in Figure 3. To solve the multiclass classifications, one-versus-one strategy is adopted in this study.

3. Case Study and Experimental Setup

The details of the experiments are discussed in the following subsections, followed by the corresponding results and comparisons. All experiments were done under a PC with a Core 2 Duo E6750 @ 2.13 GHz and 4 GB RAM. All the proposed methods mentioned were implemented using MATLAB R2008a.

3.1. Test Rig and Sample Data Acquisition. The experiments were performed on a test rig as shown in Figure 4, which can simulate the GTGS in power plants. The test rig includes a computer for data acquisition, an electric load simulator, a prime mover, a gearbox, a flywheel, and an asynchronous generator. The test rig can simulate many common faults in the gearbox of the GTGS, such as unbalance, misalignment,

Boundary constructed using one-versus-all

(a)

Boundary constructed using one-versus-one
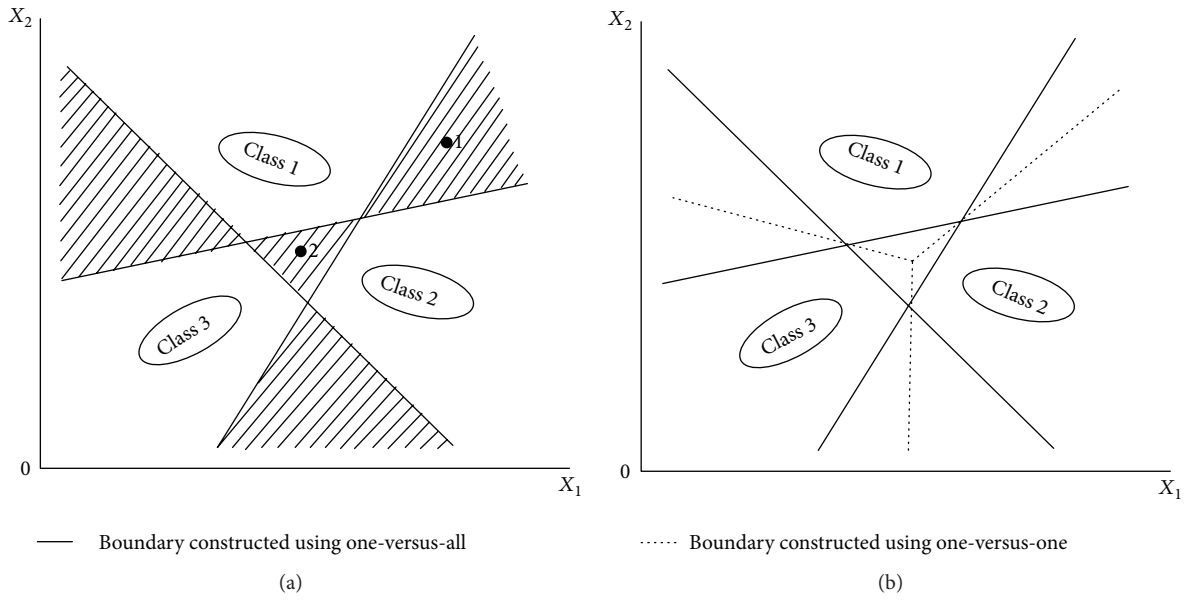
(b)

FIGURE 2: Indecisive regions (shaded regions) using one-versus-all (a) and one-versus-one (b).

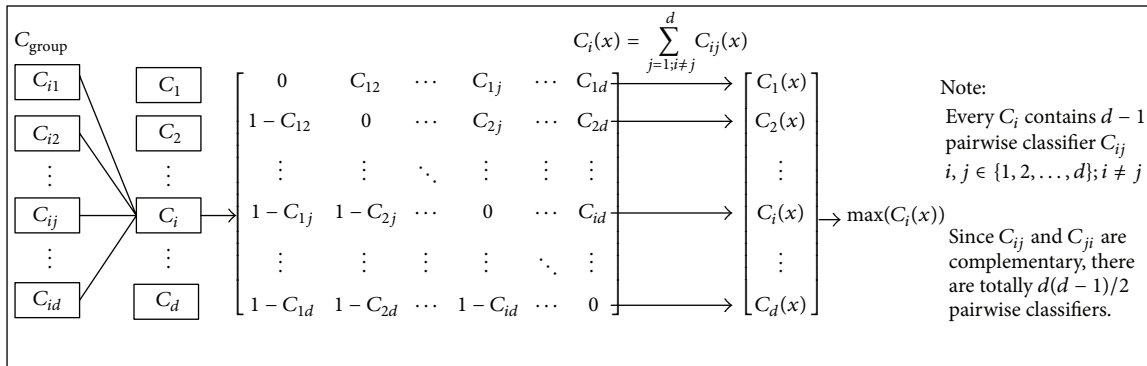FIGURE 3: One-versus-one strategy of multiclass classification.

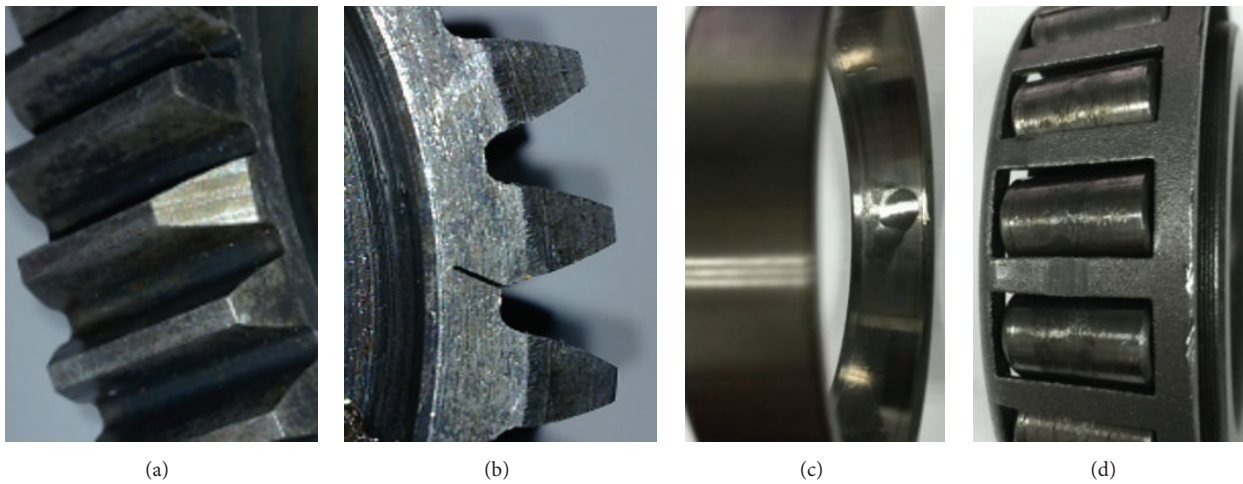(a)   (b)   (c)   (d)

FIGURE 4: Chipped tooth (a), gear crack (b), bearing outer race wear (c), and bearing balls wear (d).

TABLE 1: Description of each fault condition of the gearbox.

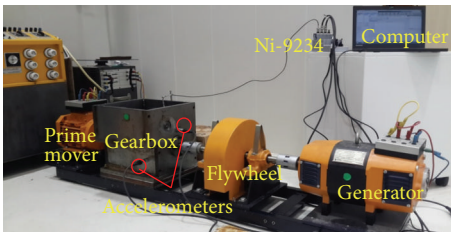| Label	Condition	Fault type
C_1		Unbalance
C_2		Looseness
C_3		Misalignment
C_4	Single-faults	Chipped tooth (a)
C_5		Gear crack (b)
C_6		Bearing outer race wear (c)
C_7		Bearing balls wear (d)
C_8	Simultaneous-faults	Chipped tooth and bearing outer race wear
C_9		Gear crack and bearing balls wear

FIGURE 5: Fault simulator for gas turbine generator system.

gear, and bearing faults. The common faults of gear and bearing are demonstrated in Figure 5. Although the vibration signal has been used to diagnose the faults of the gear or bearing in published papers, it usually contains one kind of signal which limits the number of detectable faults. For example, in this study, the axial vibration signal is not suitable to diagnose gear faults (C_4 and C_5 in Table 1) because the tooth force of the spur gear along the axial direction is insignificant. To generate more reliable diagnostic result and diagnose more faults, a new signal-based diagnostic framework by analyzing the axial and vertical vibration signals is proposed in this study. In the test rig, the signal acquisition module (Ni 9234) with accelerometers acquires the vibration signals along the axial and vertical direction. A total of 9 cases, including 7 single-faults and 2 simultaneous-faults which are described in Table 1, are simulated in the test rig in order to generate sample training and test datasets. Table 1 shows that the gear faults include chipped tooth with 1/4 tooth damage and gear crack with 5 mm crack on tooth face, whereas the bearing faults include medium wear on the rolling elements and outer races. The structural faults contain unbalance, looseness, and misalignment which are simulated by, respectively, adding one eccentric mass on the output shaft, unfastening some screws of the gearbox, and adjusting one height of the gearbox with shims. Besides, samples for fault patterns are shown in Figure 6. Figure 6 shows that all signal profiles are very similar, which make it difficult to be distinguished manually, but their degrees of similarities can be detected using the proposed framework.

To construct and test the diagnostic framework, each simulated single-fault was repeated 160 times. Each time 2 seconds of vibration data was recorded with a sampling rate of 4096 Hz. The sampling rate was set to a frequency higher than the gear meshing frequency, which can ensure no missing signal. In other words, every sampling data for each case has (2 accelerometers × 2 seconds × 4096) 16384 data points. Finally, there were 1440 fault samples data (i.e., (7 single-faults + 2 simultaneous-faults) × 160 samples). The sample data was divided into the different subsets as shown in Table 2, where D_{Valid} denotes the 720 validation sets without the feature extraction and $D_{\text{Proc-Train}}$ denotes the 630 training sets of the extracted features.

3.2. Feature Extraction by WPT and TDSF. Feature extraction method is the determination of a fault feature vector from a signal with minimal loss of useful information of the faults. A feature vector could usually be a reduced-dimensional representation of that signal so as to reduce the modeling complexity and computational cost. Through WPT, a set of 2^L subbands of a signal can be obtained, where L is the level of WPT decomposition.

After decomposing by WPT, the time-domain statistical method is employed to extract the TDSF from the decomposed signals which provide the physical characteristics of time series data. For instance, crest factor indicates how extreme the peak in a signal is, and a high standard deviation indicates that the data points are spread out over a wider range of signal. Generally, different faults have the various amplitudes of signal, resulting in a significant variation of TDSF. It is not only able to effectively indicate the faults such as cracks and wearing in the gear but also independent of loads or speeds of rotating machine. In the opening literatures [5, 15], the time-domain statistical features were applied for the fault detection on gear trains and low speed bearings, such as mean, standard deviation, crest factor, and kurtosis, respectively.

In this study, ten statistical time-domain features are employed to analyze the vibration signals. Table 3 presents the statistical time-domain features. After the feature extraction by WPT and TDSF, the number of extracted features is shown in Table 4.

3.3. Dimension Reduction by KPCA. Although the useful features can be extracted by WPT and TDSF, redundant information and the dimension of these extracted features are still high. Contained redundant information and high dimension input can degrade the diagnostic performance. To tackle this issue, KPCA is applied to obtain a small set of principal components of the extracted features. With the eigenvalues obtained from KPCA, the unimportant transformed features could be deleted. Therefore, only a limited number of the principal components are necessary and 95% of the information in the features can be retained.

3.4. Normalization. To ensure all the features are having even contribution, all reduced features should go through normalization. The interval of normalization is within [0, 1]. The extracted feature is normalized by the following formula:

$$y = \frac{(D_{\text{KPCA}} - D_{\min})}{(D_{\max} - D_{\min})}, \qquad (4)$$

TABLE 2: Division of sample dataset into different subsets.

	Type of dataset	Fault sample (1440)		Type of dataset	Fault sample (1440)
Raw sample data (x_k)	Validation dataset	D_{Valid} (720)	Feature extraction (WPT + TDSF + KPCA)	Validation dataset	$D_{\text{Proc-Valid}}$ (720)
	Training dataset	D_{Train} (630)		Training dataset	$D_{\text{Proc-Train}}$ (630)
	Test dataset	D_{Test} (90)		Test dataset	$D_{\text{Proc-Test}}$ (90)

TABLE 3: Definition of common statistical features in time-domain.

Time-domain feature	Equation
(1) Mean	$x_m = \dfrac{1}{N} \sum_{i=1}^{N} x_i$
(2) Maximum	$x_{\max} = \max \lvert x_i \rvert$
(3) Minimum	$x_{\min} = \min \lvert x_i \rvert$
(4) Standard deviation	$x_{\text{std}} = \sqrt{\dfrac{\sum_{i=1}^{N} (x_i - x_m)^2}{N - 1}}$
(5) Root mean square	$x_{\text{rms}} = \sqrt{\dfrac{1}{N} \sum_{i=1}^{N} x_i^2}$
(6) Kurtosis	$x_{\text{kur}} = \dfrac{\sum_{i=1}^{N} (x_i - x_m)^4}{(N - 1)\, x_{\text{std}}^4}$
(7) Crest factor	$\text{CF} = \dfrac{x_{pk}}{x_{\text{rms}}}$
(8) Clearance factor	$\text{CLF} = \dfrac{x_{pk}}{\left((1/N) \sum_{i=1}^{N} \sqrt{\lvert x_i \rvert} \right)^2}$
(9) Shape factor	$\text{SF} = \dfrac{x_{\text{rms}}}{(1/N) \sum_{i=1}^{N} \lvert x_i \rvert}$
(10) Impulse factor	$\text{IF} = \dfrac{x_{pk}}{(1/N) \sum_{i=1}^{N} \lvert x_i \rvert}$

Note: x_i represents a signal series for $i = 1, 2, \ldots, N$, where N is the number of data points of a raw signal.

where D_{KPCA} is an output feature after going through KPCA and y is the result of normalization. After normalization, a processed dataset is obtained, which is divided into the training dataset, validation dataset, and testing dataset named as $D_{\text{Proc-Train}}$, $D_{\text{Proc-Valid}}$, and $D_{\text{Proc-Test}}$, respectively.

4. Experimental Results and Discussion

4.1. Selection of Kernel Parameters for Feature Extraction. Since there are many combinations of the mother wavelets and kernels for KPCA and SVM, a set of experiments are carried out in the simulated gas turbine generator system to determine the best combination of the system configurations. In the phase of WPT, mother wavelets and the level of decomposition L are selected according to a trial-and-error method. In the family of mother wavelets, the Daubechies wavelet (Db) is the most popular one and hence employed for experiments. In this case study, four Daubechies wavelets from Db3 to Db6 are tried and the range of L is set from 3 to 5. Moreover, three different kernel functions for KPCA and SVM, namely, *linear*, *radial basis function* (*RBF*), and *polynomial*,

are tested. Different kernel functions have the various parameters for adjustment expect the *linear* kernel. Considering there is no parameter in linear kernel, only two optimal kernel parameters, r of the *RBF* and d of *polynomial*, for KPCA need to be determined in this study. The parameter r of the *RBF* based on 2^v is tried for v that ranged from -5 to $+5$, and the parameter d of polynomial kernel is taken from 1 to 8. In the classifier, the kernel of SVM is predefined using *polynomial* and all the parameters are set to be 1.0. Note that the kernel and parameters of SVM are predefined for evaluation only; they are not used in the final fault identification. Under this configuration, 5-fold cross validation with 720 validation datasets, $D_{\text{Proc-Valid}}$, is employed to determine the WPT decomposition level L, *RBF* radius r, and *polynomial* degree d of KPCA.

Due to space limitation, the experimental results only demonstrate the optimal kernel parameter setting of the *RBF* and *polynomial* as shown in Table 5 in which the parameter r of the *RBF* kernel is set to 2 and d of *polynomial* kernel is defined as 4. After comparing the best result of *linear*, *RBF*, and *polynomial* kernel in Table 5, KPCA with the *linear* kernel shows the prior performance compared to *polynomial* and *RBF* kernel. The main reason is that the generalization of KPCA with nonlinear kernel and SVM with *polynomial* kernel (i.e., nonlinear kernel) in current application may be degenerated because the decision surface constructed under this combination (nonlinear kernel + nonlinear kernel) may be overcomplex according to the well-known principle of *Occam's razor*. Referring to the average accuracies in Table 5, this statement is verified for this application. It explains why KPCA with *linear* kernel has higher generalization than the other nonlinear kernels, when combined with nonlinear classifiers such as SVM. Hence, the KPCA with *linear* kernel is adopted in this study. Besides, under the *linear* kernel of KPCA, the mother wavelet of Db4 with level 4 shows the highest diagnostic accuracy (96.26%) in Table 5. In terms of the number of the extracted features, 54 principal components are obtained by using the best combination of the feature extraction techniques. In other words, a raw signal of 16384 data points can be transformed to 54 features as the input variables of the classifiers.

In this study, to ensure the best performance of SVM classifier, grid search (GS) [27], a simple and effective method, is adopted to select the optimal values of the penalty parameter C with *linear* kernel, r of the *RBF* kernel, and d of *polynomial* kernel, respectively. The search region for C is set within the range of 2^i, where i is from -5 to 15 and r of RBF and d of *polynomial* parameters are also selected from 2^v, where v ranged from -5 to $+5$, and taken from 1 to 8, respectively. The best searching result for three different kernel functions

TABLE 4: TDSF extraction under different decomposition levels of WPT.

WPT	Features obtained by WPT	× TDSF	× (locations of accelerometers)	= total features
4 levels	2^4 subbands	10	2	160

TABLE 5: Accuracies of SVM classifier under various kernels of KPCA and mother wavelets.

| Mother wavelet	Level L	KPCA with *linear* kernel		KPCA with *RBF* kernel and $r = 2$		KPCA with *polynomial* kernel and $d = 4$	
		Accuracy	Feature number	Accuracy	Feature number	Accuracy	Feature number
Db3	3	95.48%	34	91.12%	86	67.22%	22
	4	94.20%	62	90.08%	110	64.20%	28
	5	95.13%	110	81.24%	132	87.36%	45
Db4	3	95.12%	25	88.24%	87	70.00%	18
	4	96.26%	54	87.34%	106	81.00 %	27
	5	95.55%	105	83.42%	132	74.26%	39
Db5	3	84.40%	26	82.53%	89	66.36%	20
	4	93.23%	54	85.32%	110	64.06%	27
	5	93.49%	105	80.00%	132	81.04%	38
Db6	3	94.20%	26	84.05%	98	68.34%	19
	4	95.32%	53	85.64%	104	66.78%	24
	5	93.78%	102	88.04%	130	85.30%	38
Average		93.85%	—	85.59%	—	73%	—
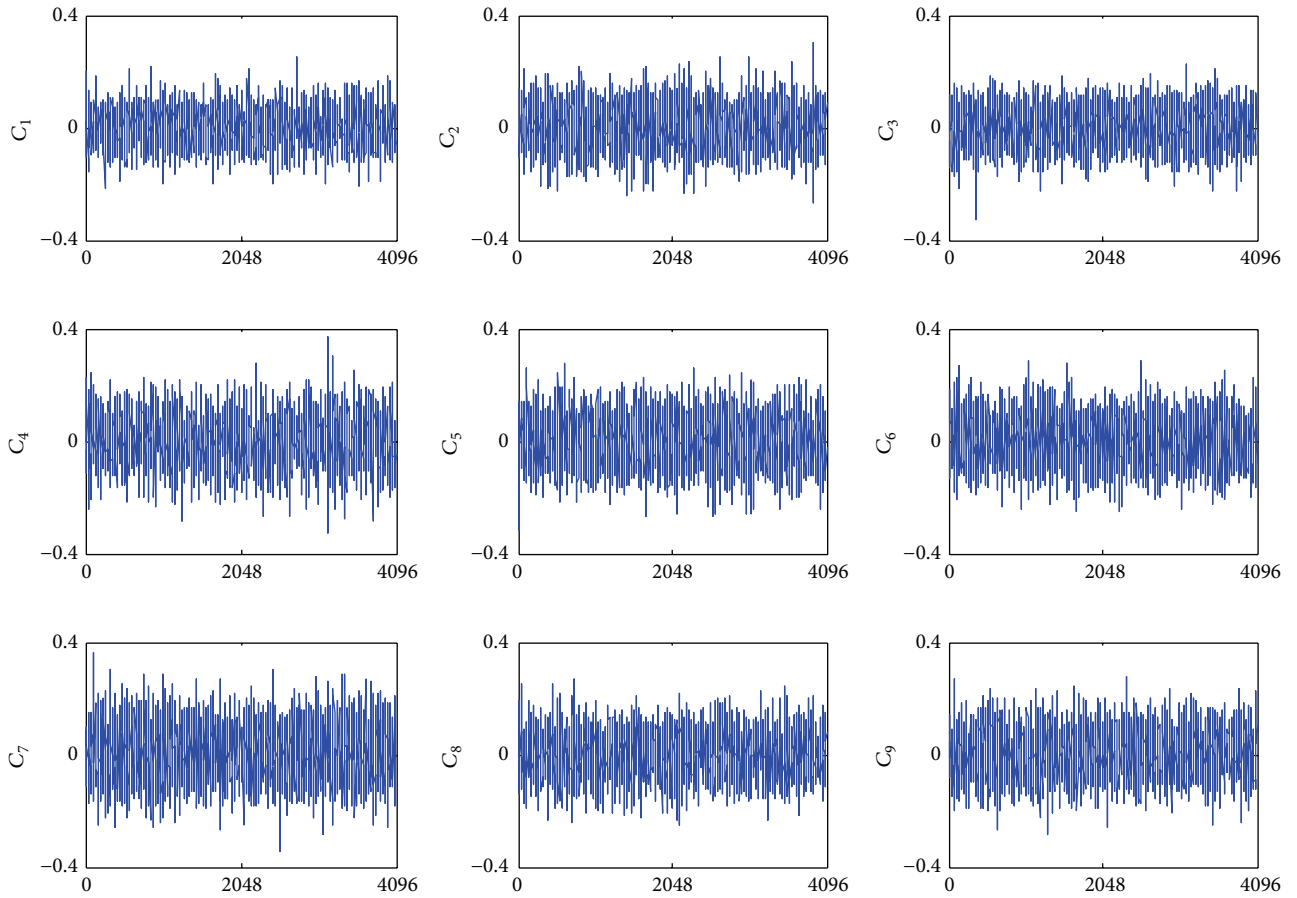
FIGURE 6: Examples of vibration signal at ninth conditions.

Table 6: Accuracies of SVM under the *linear* kernel of KPCA and Db4/L4.

Penalty parameter C	SVM with different kernel functions		
	Linear	*RBF*	*Polynomial*
	—	$r = 2$	$d = 4$
$C = 2^3$	87.16%	92.08%	97.77%

Table 7: Confusion matrix for SVM under *polynomial* kernel with $C = 2^3$ and $d = 4$ without feature extraction.

		Predicted classes									
	Class	1	2	3	4	5	6	7	8	9	$\sum$
	1	**6**	2	2	0	0	0	0	0	0	10
	2	4	**3**	1	0	2	0	0	0	0	10
	3	0	0	**3**	0	4	1	0	2	0	10
	4	0	0	1	**5**	2	0	0	0	2	10
True classes	5	0	0	3	0	**3**	3	1	0	0	10
	6	0	0	0	2	1	**5**	2	0	0	10
	7	0	0	0	0	1	3	**4**	2	0	10
	8	0	0	0	2	0	4	0	**4**	0	10
	9	0	0	0	0	3	0	3	0	**4**	10
	$\sum$	10	5	10	9	16	16	10	8	6	90

Accuracy = correct classification/all cases = 41.11%.

Table 8: Confusion matrix for SVM under *polynomial* kernel with $C = 2^3$ and $d = 4$ using extracted features by WPT + TDSF.

		Predicted classes									
	Class	1	2	3	4	5	6	7	8	9	$\sum$
	1	**10**	0	0	0	0	0	0	0	0	10
	2	0	**8**	2	0	0	0	0	0	0	10
	3	1	1	**8**	0	0	0	0	0	0	10
	4	0	1	0	**7**	0	0	0	2	0	10
True classes	5	0	0	0	0	**7**	0	0	0	3	10
	6	0	0	0	0	0	**10**	0	0	0	10
	7	0	0	0	0	0	0	**7**	0	3	10
	8	0	0	0	1	0	1	0	**8**	0	10
	9	0	0	0	0	1	0	2	0	**7**	10
	$\sum$	11	10	10	8	8	11	9	10	13	90

Accuracy = correct classification/all cases = 80.00%.

is listed in Table 6. Compared with the *linear* and *RBF* with $r = 2$ kernels, SVM using the *polynomial* kernel with $C = 2^3$ and $d = 4$ shows the best accuracy (97.77%) and gives 10.61% and 5.69% improvement, respectively.

4.2. Evaluation of the Proposed Feature Extraction Framework. To verify the effectiveness of the proposed feature extraction method, without feature extraction and feature extraction by WPT + TDSF are employed to compare with the proposed framework. Three corresponding confusion matrixes for the testing datasets are shown in Tables 7–9, respectively, each of which contains the 90 unseen cases. Table 7 reveals that the raw captured data without feature extraction are mostly misclassified because the raw data is chaotic and similar. In Table 8, the performance of classifier using the extracted features (WPT + TDSF) is improved from 41.11% to 80.00%, but still some cases are misclassified. After the feature extraction by the proposed framework (WPT + TDSF + KPCA), the fault diagnostic accuracy is enhanced obviously to attain 97.77% which is demonstrated in Table 9 where only four cases are misclassified.

Therefore, it can be concluded that the proposed feature extracted framework (WPT + TDSF + KPCA) is an effective technique to extract the fault features from the raw data.

5. Conclusion

In this paper, a simple and effective feature extracted framework has been successfully developed to overcome the challenge of faults diagnosis in the GTGS. In the proposed framework, the feature extraction technique is designed by combining WPT + TDSF to extract the features of faults. However, the extracted features are still redundancy and

Table 9: Confusion matrix for SVM under *polynomial* kernel with $C = 2^3$ and $d = 4$ using extracted features by WPT + TDSF + KPCA.

		Predicted classes									
	Class	1	2	3	4	5	6	7	8	9	$\sum$
	1	**10**	0	0	0	0	0	0	0	0	10
	2	0	**10**	0	0	0	0	0	0	0	10
	3	0	0	**10**	0	0	0	0	0	0	10
	4	0	0	0	**10**	0	0	0	0	0	10
True classes	5	0	0	0	0	**10**	0	0	0	0	10
	6	0	0	0	0	0	**10**	0	0	0	10
	7	0	0	0	0	0	0	**10**	0	0	10
	8	0	0	0	0	0	2	0	**8**	0	10
	9	0	0	0	0	1	0	1	0	**8**	10
	$\sum$	10	10	10	10	11	12	11	8	8	90

Accuracy = correct classification/all cases = 97.77%.

disorder. KPCA with *linear* kernel is effective in removing the redundancy information and further reducing the dimension of the extracted features. To verify the effectiveness of the proposed framework, without features extraction and fault feature extraction by WPT + TDSF are employed to evaluate the performance of the proposed framework (WPT + TDSF + KPCA). Experimental results show that the proposed feature extraction framework is effective in extracting the fault features from the GTGS. Since the proposed framework for fault feature extraction in the GTGS is general, it could be applied to the other similar rotating machinery problems.
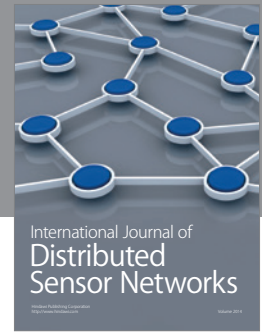
Competing Interests

The authors declare that they have no competing interests.

Acknowledgments

References

[1] X. Wang and S. Yang, "A parallel distributed knowledge-based system for turbine generator fault diagnosis," *Artificial Intelligence in Engineering*, vol. 10, no. 4, pp. 335–341, 1996.

[2] M. Todd, S. D. J. McArthur, J. R. McDonald, and S. J. Shaw, "A semiautomatic approach to deriving turbine generator diagnostic knowledge," *IEEE Transactions on Systems, Man and Cybernetics Part C: Applications and Reviews*, vol. 37, no. 5, pp. 979–992, 2007.

[3] B.-S. Yang, T. Han, and W.-W. Hwang, "Fault diagnosis of rotating machinery based on multi-class support vector machines," *Journal of Mechanical Science and Technology*, vol. 19, no. 3, pp. 846–859, 2005.

[4] Y. Changfeng, H. Zhang, and W. Lixiao, "A novel real-time fault diagnostic system for steam turbine generator set by using strata hierarchical artificial neural network," *Energy and Power Engineering*, vol. 1, no. 1, pp. 7–16, 2009.

[5] J. Xintao, D. Kang, and Y. Songsen, "Fault diagnosis of rotating machinery based on improved wavelet packet algorithm," *International Journal of Advancements in Computing Technology*, vol. 5, no. 6, pp. 267–275, 2013.

[6] Q. Miao and Q. Zhou, "Planetary gearbox vibration signal characteristics analysis and fault diagnosis," *Shock and Vibration*, vol. 2015, Article ID 126489, 8 pages, 2015.

[7] Ł. Jedliński and J. Jonak, "Early fault detection in gearboxes based on support vector machines and multilayer perceptron with a continuous wavelet transform," *Applied Soft Computing*, vol. 30, pp. 636–641, 2015.

[8] W. G. Lee, J. W. Lee, M. S. Hong, S.-H. Nam, Y. Jeon, and M. G. Lee, "Failure diagnosis system for a ball-screw by using vibration signals," *Shock and Vibration*, vol. 2015, Article ID 435870, 9 pages, 2015.

[9] F. Chaari, W. Bartelmus, R. Zimroz, T. Fakhfakh, and M. Haddar, "Gearbox vibration signal amplitude and frequency modulation," *Shock and Vibration*, vol. 19, no. 4, pp. 635–652, 2012.

[10] P. K. Kankar, S. C. Sharma, and S. P. Harsha, "Fault diagnosis of ball bearings using machine learning methods," *Expert Systems with Applications*, vol. 38, no. 3, pp. 1876–1886, 2011.

[11] T. P. Banerjee, S. Das, J. Roychoudhury, and A. Abraham, "Implementation of a new hybrid methodology for fault signal classification using short-time fourier transform and support vector machines," in *Soft Computing Models in Industrial and Environmental Applications, 5th International Workshop (SOCO 2010)*, vol. 73 of *Advances in Intelligent and Soft Computing*, pp. 219–225, Springer, Berlin, Germany, 2010.

[12] R. Yan, R. X. Gao, and X. Chen, "Wavelets for fault diagnosis of rotary machines: a review with applications," *Signal Processing*, vol. 96, pp. 1–15, 2014.

[13] P. K. Wong, Z. Yang, C. M. Vong, and J. Zhong, "Real-time fault diagnosis for gas turbine generator systems using extreme learning machine," *Neurocomputing*, vol. 128, pp. 249–257, 2014.

[14] Q. Liao, X. Li, and B. Huang, "Hybrid fault-feature extraction of rolling element bearing via customized-lifting multi-wavelet packet transform," *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, vol. 228, no. 12, pp. 2204–2216, 2013.

[15] A. Widodo and B.-S. Yang, "Application of nonlinear feature extraction and support vector machines for fault diagnosis of induction motors," *Expert Systems with Applications*, vol. 33, no. 1, pp. 241–250, 2007.

[16] Y. Lei, Z. He, Y. Zi, and X. Chen, "New clustering algorithm-based fault diagnosis using compensation distance evaluation technique," *Mechanical Systems and Signal Processing*, vol. 22, no. 2, pp. 419–435, 2008.

[17] Q. He, F. Kong, and R. Yan, "Subspace-based gearbox condition monitoring by kernel principal component analysis," *Mechanical Systems and Signal Processing*, vol. 21, no. 4, pp. 1755–1772, 2007.

[18] X.-F. Hu, Z. Zhao, S. Wang, F.-L. Wang, D.-K. He, and S.-K. Wu, "Multi-stage extreme learning machine for fault diagnosis on hydraulic tube tester," *Neural Computing and Applications*, vol. 17, no. 4, pp. 399–403, 2008.

[19] C.-L. Huang and C.-J. Wang, "A GA-based feature selection and parameters optimizationfor support vector machines," *Expert Systems with Applications*, vol. 31, no. 2, pp. 231–240, 2006.

[20] B. Samanta, "Artificial neural networks and genetic algorithms for gear fault detection," *Mechanical Systems and Signal Processing*, vol. 18, no. 5, pp. 1273–1282, 2004.

[21] S.-W. Fei and X.-B. Zhang, "Fault diagnosis of power transformer based on support vector machine with genetic algorithm," *Expert Systems with Applications*, vol. 36, no. 8, pp. 11352–11357, 2009.

[22] Y. Lei, Z. Liu, X. Wu, N. Li, W. Chen, and J. Lin, "Health condition identification of multi-stage planetary gearboxes using a mRVM-based method," *Mechanical Systems and Signal Processing*, vol. 60, pp. 289–300, 2015.

[23] C. Alzate and J. A. K. Suykens, "Multiway spectral clustering with out-of-sample extensions through weighted kernel PCA," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 2, pp. 335–347, 2010.

[24] P. K. Wong, J. Zhong, Z. Yang, and C. M. Vong, "Sparse Bayesian extreme learning committee machine for engine simultaneous fault diagnosis," *Neurocomputing*, vol. 174, pp. 331–343, 2016.

[25] T.-F. Wu, C.-J. Lin, and R. C. Weng, "Probability estimates for multi-class classification by pairwise coupling," *Journal of Machine Learning Research*, vol. 5, pp. 975–1005, 2004.

[26] S. Abe, *Support Vector Machines for Pattern Classification*, vol. 53, Springer, London, UK, 2005.

[27] C.-W. Hsu, C.-C. Chang, and C.-J. Lin, *A Practical Guide to Support Vector Classification*, 2003.