

Research Article

Simultaneous-Fault Diagnosis of Gas Turbine Generator Systems Using a Pairwise-Coupled Probabilistic Classifier

Zhixin Yang,¹ Pak Kin Wong,¹ Chi Man Vong,² Jianhua Zhong,¹ and JieJunYi Liang¹

¹ Department of Electromechanical Engineering, University of Macau, Macau

² Department of Computer and Information Science, University of Macau, Macau

Correspondence should be addressed to Zhixin Yang; zxyang@umac.mo

Received 25 April 2013; Accepted 18 June 2013

Academic Editor: Qingsong Xu

Copyright © 2013 Zhixin Yang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

A reliable fault diagnostic system for gas turbine generator system (GTGS), which is complicated and inherent with many types of component faults, is essential to avoid the interruption of electricity supply. However, the GTGS diagnosis faces challenges in terms of the existence of simultaneous-fault diagnosis and high cost in acquiring the exponentially increased simultaneous-fault vibration signals for constructing the diagnostic system. This research proposes a new diagnostic framework combining feature extraction, pairwise-coupled probabilistic classifier, and decision threshold optimization. The feature extraction module adopts wavelet packet transform and time-domain statistical features to extract vibration signal features. Kernel principal component analysis is then applied to further reduce the redundant features. The features of single faults in a simultaneous-fault pattern are extracted and then detected using a probabilistic classifier, namely, pairwise-coupled relevance vector machine, which is trained with single-fault patterns only. Therefore, the training dataset of simultaneous-fault patterns is unnecessary. To optimize the decision threshold, this research proposes to use grid search method which can ensure a global solution as compared with traditional computational intelligence techniques. Experimental results show that the proposed framework performs well for both single-fault and simultaneous-fault diagnosis and is superior to the frameworks without feature extraction and pairwise coupling.

1. Introduction

The gas turbine generator system (GTGS) is commonly used in many power plants. The main components of a GTGS are power turbine, gearbox, flywheel, and asynchronous generator. In the first phase of the GTGS, the power turbine is driven by the exhaust gas; the output of the turbine then drives a gearbox that connects to the flywheel which keeps constant moment of inertia to protect the generator from a sudden stop. Finally, the rotating flywheel drives the asynchronous generator to generate the electric power. The system is designed to run 24 hours per day. Any abnormal situation of the GTGS will interrupt the electricity supply to cause enormous economic loss. The traditional manual inspection on the GTGS can hardly accomplish the fault-monitoring task because the GTGS is complicated and also more than one fault may appear at a time. This kind of problem refers to simultaneous-fault diagnosis. To prevent

the interruption of electricity supply, development of an intelligent fault diagnostic system for single and simultaneous faults of the GTGS is a promising research topic.

According to the existing literature, no physical diagnostic model for simultaneous-fault diagnosis about power generator system like GTGS is available. One possible solution for intelligent fault detection is to use machine learning methods, like neural networks (NNs) and support vector machines, to learn the normal and abnormal vibration signal patterns. A classifier is then built for fault detection of unseen signal patterns.

In recent years, researches on the development of neural network-based monitoring systems for the GTGS and rotating machinery were available [1–4]. However, the neural network classifiers have many drawbacks, such as local minima, time consuming for determination of optimal network structure, and risk of overfitting. To date, a number of researchers have already applied support vector machines (SVMs) to

diagnose rotating machine faults and other engineering diagnosis problems [5–13] and have shown that SVM is superior to traditional NNs [8, 10, 11]. The major advantages of SVM are global optimum and higher generalization capability [8, 11]. Nevertheless, those classification systems can detect single fault only. Even though the simultaneous fault can also be detected by the concept of single-label classification, in which both single and simultaneous-faults are considered as independent labels, they need many simultaneous-fault training patterns. However, transformation of simultaneous-fault diagnosis to single-label classification suffers from two drawbacks. Firstly, the transformation significantly increases the required number of the training data because a set of new classes will be artificially generated via combination of single faults, which makes it impractical to handle industrial problems in medium to large level of faults. For example, with d single-faults (labels), there are $2^d - (d + 1)$ artificial simultaneous-fault labels. Each label requires a certain amount of training dataset. So, it is impractical and prohibitive because acquiring all the possible simultaneous-fault patterns is costly and hard to collect. Secondly, it is very difficult to include new faults in the system later. It is because of the addition of one single fault that the number of combination of single faults for constructing the artificial simultaneous-fault labels will be increased exponentially so as to require a huge amount of corresponding training patterns too. Another way to solve the problem of simultaneous-fault diagnosis is the framework called hierarchical artificial neural network (HANN) which is constructed by several stages of NNs. Each of the HANN stage is constructed with a set of NNs as described in [2, 14–16]. HANN also suffers from the shortcoming that its architecture will become too complex to handle when extending to medium or large scale. With the increase of the number of single-faults, a large amount of NNs must be constructed in every HANN stage.

To overcome the aforesaid limitations, one possible way is to develop a classifier which can diagnose both single and simultaneous faults, while it is trained with single fault data only. This idea refers to multilabel classification. It is believed in the area of GTGS that the features of single-fault signal patterns can be found from the simultaneous-fault patterns, if a proper feature extraction technique can be selected. Should this hypothesis be true, it is possible to develop such kind of classifier by using proper feature extraction techniques. Therefore, one of the originalities of this research is to explore the feasibility of detection of simultaneous-fault vibration patterns of rotary machinery by a classifier which is only trained with single-fault patterns. To exam the feasibility, experimental evaluation is proposed in this study.

As far as feature extraction is concerned, there are currently several feature extraction techniques for fault detection of rotary machinery, including wavelet packet transform (WPT) [4, 8, 17], independent component analysis [6], and time-domain statistical features (TDSF) [13, 18]. By reviewing the literature, WPT and TDSF are commonly used for rotating machinery so that both techniques are considered in this research. After performing feature extraction, there may be still some irrelevant and redundant information in

the extracted features. It is well known that if the number of inputs of a fault classifier is too huge, its accuracy will be degenerated unless a huge number of training data is available, but it is impractical. To resolve this problem, a feature selection method should be employed to wipe off irrelevant and redundant information such that the number of inputs of the classifier can be reduced, resulting in improvement on diagnostic accuracy. Currently, there are some feature selection approaches available, including compensation distance evaluation technique (CDET) [18], kernel principal component analysis (KPCA) [19], and genetic algorithm (GA)-based methods [10, 20]. 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 to reduce the dimensional scale of the information content. The term of feature extraction hereafter refers to feature extraction and feature selection.

From the practical point of view, a proper classifier has to offer the probabilities of all possible faults. Then the user can at least trace the other possible faults according to the rank of their probabilities when the predicted fault(s) from the classifier is (are) incorrect in the problem. Moreover, it is our belief that comparing the similarity of a simultaneous-fault input pattern with every single fault pattern is the main mechanism for simultaneous-fault detection without learning any simultaneous-fault training data. One possible way to represent the degree of similarity of two patterns is the probability of likeness. Therefore, it is logical to employ probabilistic classifier for simultaneous-fault diagnosis because they can determine the probability of each label. Typically, probabilistic neural network (PNN) [21, 22] was employed as a probabilistic classifier. It was shown in [21] that the performance of PNN is superior to SVM-based method for multilabel classification. However, the main drawback of PNN lies in the limited number of inputs because the complexity of the network and the training time are heavily related to the number of inputs. Recently, Widodo et al. [23] proposed to apply an advance classifier, namely, relevance vector machine (RVM), to the fault diagnosis of low speed bearings. They showed that RVM is superior to SVM in terms of diagnostic accuracy. RVM is a statistical learning method proposed by Tipping [24], which trains a probabilistic classifier with sparser model using Bayesian framework. RVM can be extended to multiclass version using one-versus-all (1vA) strategy. However, this strategy was verified to produce a large region of indecision [25, 26]. In view of this drawback, this research proposes to incorporate pairwise coupling, that is, one-versus-one (1v1) strategy, into RVM, namely, pairwise-coupled relevance vector machine (PCRVM). As the pairwise coupling strategy considers the correlation between every pair of fault labels, a more accurate estimate of label probabilities for simultaneous-fault signals can be achieved. The detailed explanation of the advantage of 1v1 over 1vA is discussed in Section 2.3.

If a probabilistic classification is applied to fault detection, the predicted fault is usually inferred as the one with the largest probability. The other alternative approach is

that the probabilistic classifier ranks all the possible faults according to their probabilities and lets the engineer make a decision. These inference approaches work fine with single-fault detection but fail to determine which faults occur simultaneously in the simultaneous-fault problem. It is because the engineer cannot identify the number of simultaneous faults based on the output probability of each label. For instance, an output probability vector for five labels is given as [0.49, 0.5, 0.68, 0.1, 0.7]. In this example, it is difficult for the engineer to judge whether the simultaneous faults are labels 2, 3, and 5. To identify the number of simultaneous faults, a decision threshold must be introduced, and thus a step of decision threshold optimization is proposed in the current framework other than feature extraction and probabilistic classification.

As a summary, a framework combining feature extraction (WPT + TDSF + KPCA), pairwise-coupled relevance vector machine (PCRVM), and effective decision threshold optimization is proposed for simultaneous-fault diagnosis of the GTGS. Even though the authors also proposed a similar framework for simultaneous fault diagnosis of automotive engine ignition systems [27], the framework was designed for simple waveforms, whereas the signal patterns of this application are high frequency oscillating signals. Moreover, one of the classification criteria and decision threshold optimization method in [27], respectively, rely on domain-specific knowledge and computational intelligence techniques, such as genetic algorithm (GA) and particle swarm optimization (PSO), which may produce a local optimal solution. Therefore, the framework in [27] cannot be directly applied; it should be modified significantly, particularly in the phases of feature extraction and the threshold optimization.

This paper is organized as follows. Section 2 presents the proposed framework and the related techniques. Experimental setup and sample data acquisition are discussed in Section 3. Section 4 discusses the experimental results of PCRVM and its comparison with typical PNN [21, 22], an existing single-label SVM classifier [28], and pairwise-coupled probabilistic neural network (PCPNN) in order to show the superiority of RVM and effectiveness of pairwise coupling strategy. Finally, a conclusion is given in Section 5.

2. Proposed Simultaneous-Fault Diagnosis Framework

The proposed simultaneous-fault diagnostic framework for the GTGS and its evaluation approach are shown in Figure 1. The framework consists of four submodules: (1) feature extraction, (2) pairwise-coupled probabilistic classifier, (3) model parameter and decision threshold optimization, and (4) performance evaluation. The proposed framework can be applicable to PCPNN and PCRVM.

In the feature extraction submodule, the sample dataset is divided into three independent groups: validation dataset, training dataset, and test dataset. The validation dataset and test dataset involve the combination of both single-fault patterns and simultaneous-fault patterns, while the training dataset contains single-fault patterns only. WPT and TDSF

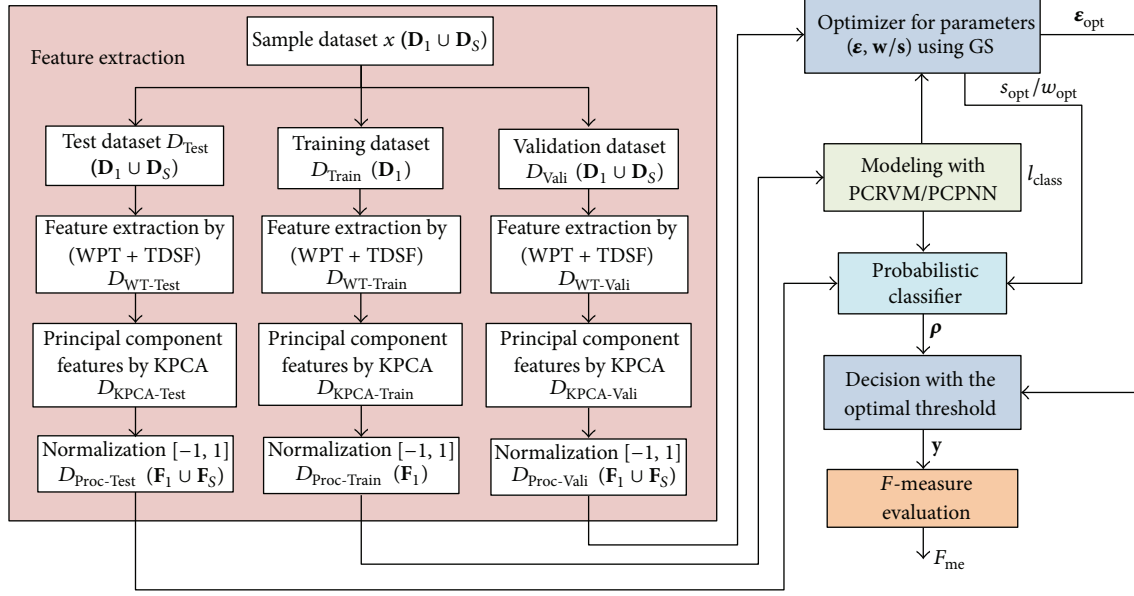
are used for feature extraction, in which the training feature set, $D_{WT-Train}$, validation feature set, $D_{WT-Vali}$, and the unseen vibration signal for testing, $D_{WT-Test}$, are generated. Considering the existence of irrelevant and redundant information in the extracted features, KPCA is then applied to remove useless information and further reduce the dimensions of $D_{WT-Train}$, $D_{WT-Vali}$, and $D_{WT-Test}$. The results are saved as $D_{KPCA-Train}$, $D_{KPCA-Vali}$, and $D_{KPCA-Test}$, respectively. In order to ensure all the features having even contribution, every feature in $D_{KPCA-Train}$, $D_{KPCA-Vali}$, and $D_{KPCA-Test}$ is normalized within $[-1, 1]$. The processed training dataset, validation dataset, and unseen signal for testing are named as $D_{Proc-Train}$, $D_{Proc-Vali}$, and $D_{Proc-Test}$, respectively.

The diagnostic model, using pairwise-coupled probabilistic classifier (PCPNN or PCRVM), is trained using the processed training dataset $D_{Proc-Train}$. The output of the trained classifier, together with processed validation dataset $D_{Proc-Vali}$, is used to perform parameter optimization for the probabilistic classifier (i.e., the spread of PCPNN or the width of PCRVM) and the decision threshold. The optimized parameter for the classifier is used to construct the final PCRVM or PCPNN-based diagnostic model, and the optimal threshold is utilized to identify simultaneous-faults for the unseen test dataset, $D_{Proc-Test}$. The performance evaluation submodule adopts F -measure to test $D_{Proc-Test}$. The details of the four submodules in the framework are discussed in the following subsections.

2.1. Feature Extraction

2.1.1. Wavelet Packet Transform. In the last two decades, wavelet transform (WT) has been widely applied in random signal processing. The transform of a signal is just another form of representing the signal. It does not change the information content presented in the signal. In WT, multiresolution technique is used, and different frequencies are analyzed with different resolutions in order to provide a time-frequency representation of the signal. Wavelet packet transform (WPT) derives from the WT family. WPT is a generalization of wavelet decomposition that offers a richer signal analysis. In the decomposition of a signal by DWT, only the lower frequency band is decomposed, giving a right recursive binary tree structure whose right lobe represents the lower frequency band, and its left lobe is the higher frequency band. In the corresponding WPT decomposition, the lower as well as the higher frequency bands are decomposed giving a balanced binary tree structure. Therefore, the WPT has the same frequency bandwidths in each resolution while the traditional discrete wavelet transform does not have this property. Therefore, WPT is suitable for processing of nonstationary signals, like the vibration signal, because the same frequency bandwidths can provide good resolution regardless of high and low frequencies.

2.1.2. Kernel Principal Component Analysis. Principal component analysis (PCA) is a popular statistical method for principal component feature extraction. PCA always performs well in dimensionality reduction when the input variables are linearly correlated. However, for nonlinear



Note:

D_1 : Dataset of single-faults
 D_S : Dataset of simultaneous-faults
 F_1 : Feature matrix of single-faults
 F_S : Feature matrix of simultaneous-faults
 ρ : Probability vector
 y : Decision vector

GS: Grid search
 s : Spread of PNN
 w : Width of RVM
 ϵ_{opt} : Optimal threshold
 s_{opt} : Optimal spread of PNN
 w_{opt} : Optimal width of RVM

FIGURE 1: Proposed simultaneous-fault diagnostic framework and evaluation approach for GTGS.

cases, PCA cannot give a good performance. Hence, PCA is extended to nonlinear version under SVM formulation and is called kernel PCA (KPCA), which has been used to solve many application problems [29, 30]. KPCA involves solving α in the following set of equations:

$$\Omega_{I,J} \alpha = \lambda \alpha, \quad (1)$$

where $\Omega_{I,J} = K(D_{WTI}, D_{WTJ})$ for $I, J = 1, \dots, Q$, Q is the number of data for KPCA, $D_{WTI}, D_{WTJ} \in R^c$, and c is the dimension of the input data. The vector $\alpha = [\alpha_1, \dots, \alpha_Q]^T$ is the eigenvector of $\Omega_{I,J}$, and $\lambda \in R$ is the corresponding eigenvalue. The transformed variables (score variables) z_r for vector D become

$$z_r(D_{WT}) = \sum_{I=1}^Q \alpha_{r,I} K(D_{WTI}, D_{WT}), \quad (2)$$

where $\alpha_{r,I}$ is the I th element in the eigenvector α corresponding to the r th 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 $D_{WT} \in R^c$ can be transformed to a nonlinearly uncorrelated variable $z = [z_1, \dots, z_p]$ where $c > p$. One more

point to note is that the eigenvectors α_r should satisfy the normalization condition of unit length:

$$\lambda_r \alpha_r^T \alpha_r = 1, \quad r = 1, 2, \dots, p, \quad (3)$$

where $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p > 0$. 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. Relevance Vector Machine. Relevance vector machine (RVM) is a recently available machine learning method. Theoretically, RVM is a statistical learning method utilizing Bayesian learning framework and popular kernels. In this research, predicting the posterior probability of each fault t_n for unseen symptoms \mathbf{f} is conducted by RVM based on experimental data. Given a set of training data $(\mathbf{f}, \mathbf{t}) = \{\mathbf{f}_n, t_n\}$, $n = 1$ to N , $t_n \in \{0, 1\}$, and N is the number of training data. It follows the statistical convention and generalizes the linear

model by applying the logistic sigmoid function $\sigma(y(\mathbf{f})) = 1/(1+\exp(-y(\mathbf{f})))$ to the predicted decision $y(\mathbf{f})$ and adopting the Bernoulli distribution for $P(\mathbf{t} | \mathbf{F})$, and the likelihood of the data is written as [24]

$$P(\mathbf{t} | \mathbf{F}, \mathbf{w}) = \prod_{n=1}^N \sigma\{y(\mathbf{f}_n; \mathbf{w})^{t_n}\} [1 - \sigma\{y(\mathbf{f}_n; \mathbf{w})\}]^{1-t_n}, \quad (4)$$

$$\text{where } y(\mathbf{f}; \mathbf{w}) = \sum_{i=1}^N w_i K(\mathbf{f}, \mathbf{f}_i) + w_0$$

and $\mathbf{w} = (w_0, w_2, \dots, w_N)^T$ are the adjustable parameters. K is a radial basis function (RBF) since RBF kernel is usually adopted for classification problems.

The optimal weight vector \mathbf{w} for the given dataset needs to be computed so as to maximize the probability $P(\mathbf{w} | \mathbf{t}, \mathbf{F}, \boldsymbol{\alpha}) \propto P(\mathbf{t} | \mathbf{F}, \mathbf{w})P(\mathbf{w} | \boldsymbol{\alpha})$, with $\boldsymbol{\alpha} = [\alpha_0, \alpha_1, \dots, \alpha_N]$, a vector of $N+1$ hyperparameters. However, the weights cannot be determined analytically. Thus, the following approximation procedure is chosen, which is based on Laplace's method.

- (a) For the current fixed values of $\boldsymbol{\alpha}$, the most probable weights \mathbf{w}_{MP} are found, which is the location of the posterior mode. Since $P(\mathbf{w} | \mathbf{t}, \mathbf{F}, \boldsymbol{\alpha}) \propto P(\mathbf{t} | \mathbf{F}, \mathbf{w})P(\mathbf{w} | \boldsymbol{\alpha})$, this step is equivalent to the following maximization:

$$\mathbf{w}_{\text{MP}} = \arg \max_{\mathbf{w}} \log \{P(\mathbf{t} | \mathbf{F}, \mathbf{w}) P(\mathbf{w} | \boldsymbol{\alpha})\}$$

$$= \arg \max_{\mathbf{w}} \left\{ \sum_{n=1}^N [t_n \log d_n + (1 - t_n) (1 - \log d_n)] - \frac{1}{2} \mathbf{w}^T \mathbf{A} \mathbf{w} \right\}$$

with $d_n = \sigma\{y(\mathbf{f}_n; \mathbf{w})\}$, $\mathbf{A} = \text{diag}(\alpha_0, \alpha_1, \dots, \alpha_N)$. (5)

- (b) Laplace's method is simply a Gaussian approximation to the log posterior around the mode of the weights \mathbf{w}_{MP} . Equation (5) is differentiated twice to give

$$\nabla_{\mathbf{w}} \nabla_{\mathbf{w}} \log P(\mathbf{w} | \mathbf{t}, \mathbf{F}, \boldsymbol{\alpha}) |_{\mathbf{w}_{\text{MP}}} = -(\boldsymbol{\Phi}^T \mathbf{B} \boldsymbol{\Phi} + \mathbf{A}), \quad (6)$$

where $\mathbf{B} = \text{diag}(\beta_1, \beta_2, \dots, \beta_N)$ is a diagonal matrix with $\beta_n = \sigma\{y(\mathbf{f}_n; \mathbf{w})\}[1 - \sigma\{y(\mathbf{f}_n; \mathbf{w})\}]$ and $\boldsymbol{\Phi}$ is an $N \times (N+1)$ design matrix with $\Phi_{nm} = K(\mathbf{f}_n, \mathbf{f}_{m-1})$ and $\Phi_{n0} = 1$, $n = 1$ to N and $m = 1$ to $N+1$. By inverting (6), the covariance matrix $\boldsymbol{\Sigma} = (\boldsymbol{\Phi}^T \mathbf{B} \boldsymbol{\Phi} + \mathbf{A})^{-1}$ can be obtained.

- (c) The hyperparameter vector $\boldsymbol{\alpha}$ is updated using an iterative reestimation equation. Firstly, randomly guess α_i , and calculate $\gamma_i = 1 - \alpha_i \sum_{ii}$ where \sum_{ii} is the i th diagonal element of the covariance matrix $\boldsymbol{\Sigma}$. Then reestimate α_i as follows:

$$\alpha_i^{\text{new}} = \frac{\gamma_i}{u_i^2}, \quad (7)$$

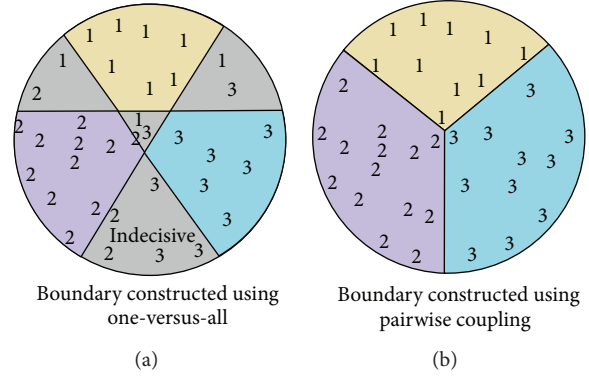


FIGURE 2: Indecisive regions (grey regions) using one-versus-all (a) and pairwise coupling (b).

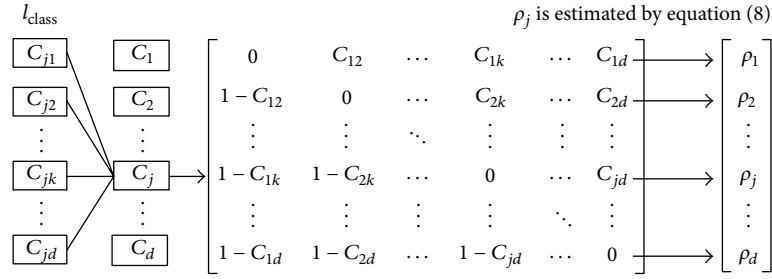
where $\mathbf{u} = \mathbf{w}_{\text{MP}} = \sum \boldsymbol{\Phi}^T \mathbf{B} \mathbf{t}$. Set $\alpha_i \leftarrow \alpha_i^{\text{new}}$, and reestimate γ_i and α_i^{new} again until convergence. Then $\mathbf{w} = \mathbf{w}_{\text{MP}}$ is estimated so that the classification model $y(\mathbf{f}; \mathbf{w}) = \sum_{i=1}^N w_i K(\mathbf{f}, \mathbf{f}_i) + w_0$ is obtained [27].

2.3. Pairwise-Coupled RVM. The traditional RVM formulation is designed only for the issue of binary classification, in which the output is either positive (+1) or negative (-1). In order to resolve the current simultaneous-fault problem, multiclass strategies of one-versus-all and one-versus-one (pairwise coupling) can be employed. Traditionally, one-versus-all strategy constructs a group of classifiers $l_{\text{class}} = [C_1, C_2, \dots, C_j, \dots, C_d]$ in a d -label classification problem. For any undefined input \mathbf{f} , the classification vector $\mathbf{y} = [y_1, y_2, \dots, y_j, \dots, y_d]$, where $y_j = 1$ if $C_j(\mathbf{f}) = +1$ or $y_j = 0$ if $C_j(\mathbf{f}) = -1$. The one-versus-all strategy is simple and easy to implement. However, it generally gives a poor result [25, 26, 31] since one-versus-all does not consider the pairwise correlation and, hence, induces a much larger indecisive region than one-versus-one as shown in Figure 2. On the other hand, pairwise coupling (one-versus-one) also constructs a group of classifiers $l_{\text{class}} = [C_1, C_2, \dots, C_j, \dots, C_d]$ in a d -label classification problem. However, each $C_j = [C_{j1}, C_{j2}, \dots, C_{jk}, \dots, C_{jd}]$ is composed of a set of $d-1$ different pairwise classifiers C_{jk} , $j \neq k$. Since C_{jk} and C_{kj} are complementary, there are totally $d(d-1)/2$ classifiers in l_{class} as shown in Figure 3.

There are several available methods for pairwise coupling strategy [25], which are, however, unsuitable for simultaneous-fault diagnosis because of the constraint $\sum \rho_j = 1$. Note that the nature of simultaneous-fault diagnosis is that $\sum \rho_j$ is unnecessarily equal to 1. Therefore, the following simple pairwise coupling strategy for simultaneous-fault diagnosis is proposed. Every ρ_j is calculated as

$$\rho_j = C_j(\mathbf{f}) = \frac{\sum_{k=1:k \neq j}^d n_{jk} C_{jk}(\mathbf{f})}{\sum_{k=1:k \neq j}^d n_{jk}} = \frac{\sum_{k=1:k \neq j}^d n_{jk} \rho_{jk}}{\sum_{k=1:k \neq j}^d n_{jk}}, \quad (8)$$

where n_{jk} is the number of training feature vectors with either j th or k th labels. Hence, the probability can be more accurately estimated from $\rho_{jk} = C_{jk}(\mathbf{f})$ because the



Note.

Every C_j contains $d - 1$ pairwise classifier C_{jk} , $k \in \{1, 2, \dots, d\}; k \neq j$.

Since C_{jk} and C_{kj} are complementary, there are totally $d(d - 1)/2$ pairwise classifiers.

FIGURE 3: Pairwise coupling strategy of probabilistic classification.

pairwise correlation between the labels is taken into account. With the previous pairwise coupling strategy, the proposed framework, PCRVM, could estimate the probability vector ρ in high accuracy level and, hence, generates a higher classification accuracy for simultaneous-fault diagnosis.

2.4. Decision Threshold. The pairwise-coupled probabilistic classification could produce a probability vector $\rho = [\rho_1, \rho_2, \dots, \rho_j, \dots, \rho_d]$, where d is the number of the single-fault labels, indicating the probabilistic occurrence of every single-fault. This probability vector ρ can be provided to the user as a quantitative measure for reference and further processing. Practically, it is desirable to identify the multilabel decision as a decision vector $\mathbf{y} = [y_1, y_2, \dots, y_j, \dots, y_d]$ for all $y_j \in \{0, 1\}$, which is, however, not directly available from the probability vector ρ . By applying a decision threshold ε , \mathbf{y} could be derived from ρ , that is, $\mathbf{y} = \text{DT}(\rho) = [\varepsilon(\rho_1), \varepsilon(\rho_2), \dots, \varepsilon(\rho_j), \dots, \varepsilon(\rho_d)]$, where

$$\varepsilon(\rho_j) = \begin{cases} 1 & \text{if } \rho_j \geq \varepsilon \\ 0 & \text{otherwise} \end{cases}, \quad \text{for } j = 1 \text{ to } d, \quad (9)$$

where $\varepsilon \in [0 \dots 1]$ is a decision threshold and y_j indicates if \mathbf{f} belongs to the j th label or not, $j = 1$ to d . For example, if $\varepsilon = 0.5$ and $\rho = l_{\text{class}}(\mathbf{f}) = [0.62, 0.71, 0.35, 0.46, 0.85]$, then $\mathbf{y} = \text{DT}(\rho) = [1, 1, 0, 0, 1]$. Therefore, \mathbf{f} is diagnosed as a simultaneous-fault (1, 2, 5). Note that the decision threshold ε is the major factor affecting the classification accuracy. The decision threshold is domain specific and sensitive to the diagnostic accuracy. The diagnosis of GTGS requires an efficient searching algorithm for optimal setting of the decision threshold. In this study, a simple and effective method, grid search (GS), is adopted. The search region for the decision threshold was set within the range of 0 to 1. By applying a reasonable small interval, say 0.01, a series of candidate thresholds were generated. After evaluating the performance index, F -measure, of various candidate thresholds using the validation dataset, the best threshold ε_{opt} can be obtained. Even though GS method is time-consuming, it can ensure a global solution, if the grids can cover the whole searching space. As a result, GS will not stick in a local optimal solution

as compared with the computational intelligent techniques such as GA and PSO in [27, 32]. Thankfully, the threshold lies between 0 and 1, so the trial time is not very long.

2.5. Evaluation. The traditional statistical measure of classification accuracy only considers exact matching of the decision vector \mathbf{y} against the true vector \mathbf{t} . This evaluation is, however, unsuitable for simultaneous-fault diagnosis where partial matching is preferred. Therefore, a well-known and common evaluation method called F -measure [32–34] is employed. F -measure is mostly used as a performance evaluation for information retrieval systems where a document may belong to a single or multiple tags simultaneously, which is very similar to our current study. By using F -measure, the evaluation of both single-fault and simultaneous-fault test cases can be fairly examined. The definition of F -measure is given in (10). The larger the F -measure value, the higher the diagnostic accuracy will be

$$F_{\text{me}} = \frac{2 \sum_{j=1}^d \sum_{i=1}^{N_t} y_{ij} t_{ij}}{\sum_{j=1}^d \sum_{i=1}^{N_t} y_{ij} + \sum_{j=1}^d \sum_{i=1}^{N_t} t_{ij}} \in [0, 1], \quad (10)$$

where $\mathbf{y}_i = [y_{i1}, y_{i2}, \dots, y_{ij}, \dots, y_{id}]$ and $\mathbf{t}_i = [t_{i1}, t_{i2}, \dots, t_{ij}, \dots, t_{id}]$ are the predicted decision vector and the true decision vector, respectively, for $j = 1$ to d and $i = 1$ to N_t and for all $y_{ij}, t_{ij} \in \{0, 1\}$. N_t is the number of single-fault and simultaneous-fault test patterns.

2.6. Summary of the Proposed Framework. The proposed framework and techniques are summarized in Figure 4. Figure 4(a) shows the workflow of combining technologies including WPT, TDSE, and KPCA as feature extraction submodule. Every dataset for training, validation, and testing is required to go through the feature extraction process. Figure 4(b) shows the architecture of the classifier l_{class} , where the pairwise coupling is deployed as depicted in Figure 3. Then the classifier is passed to a grid search optimizer to search for the optimal parameter of the classifier (i.e., w for PCRVM/RVM or s for PNN/PCPNN) and the decision

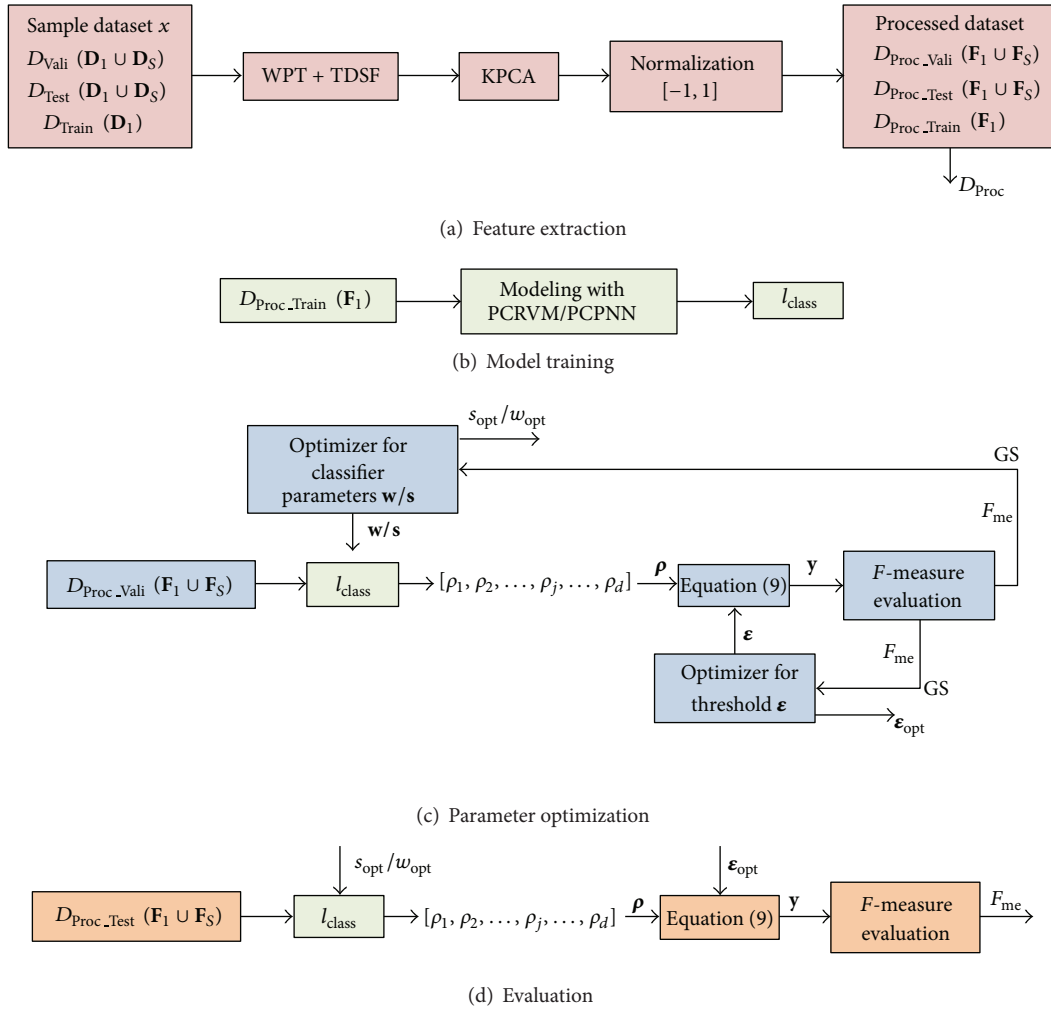


FIGURE 4: Detailed workflow of the proposed framework.

threshold based on the validation set $D_{Pro_Vali} (F_1 \cup F_S)$ and F -measure F_{me} , (Figure 4(c)). l_{class} produces the probability vector $\rho = [\rho_1, \rho_2, \dots, \rho_j, \dots, \rho_d]$ for each case in $D_{Pro_Vali} (F_1 \cup F_S)$. To optimize the parameters of the classifier and decision threshold, the fitness of each candidate parameter is obtained and evaluated based on the F -measure F_{me} over $D_{Pro_Vali} (F_1 \cup F_S)$ under 5-fold cross-validation strategy. Finally, the test dataset $D_{Pro_Test} (F_1 \cup F_S)$ is used for evaluating the performance of the proposed framework based on the optimal classifier parameter and decision threshold obtained.

3. Experimental Setup and Data Preparation for a Case Study

To obtain representative sample data for model construction and verify the effectiveness of the proposed framework, experiments were carried out. The details of the experiments are discussed in the following subsections, followed by the corresponding results and comparisons. All the proposed methods mentioned were implemented by using Matlab

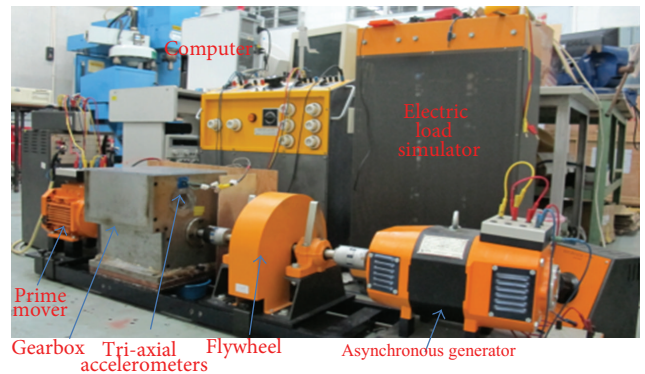


FIGURE 5: Fault simulator for gas turbine generator system.

R2008a and executed on a PC with a Core 2 Duo E6750 @ 2.13 GHz with 4 GB RAM onboard.

3.1. Test Rig and Sample Data Acquisition. The experiments were performed on a test rig as shown in Figure 5, which can simulate the GTGS of the Macau power plant. The test

TABLE 1: Sample single-faults and possible simultaneous-faults in the gearbox of GTGS.

Case no.	Condition	Description of fault
C1	Normal	Normal
C2	Single-faults	Unbalance
C3		Looseness
C4		Mechanical misalignment
C5		Wear of cage and rolling elements of bearing
C6		Wear of outer race of bearing
C7		Gear tooth broken
C8		Gear crack
C9		Chipped tooth
C10	Simultaneous-faults	Gear tooth broken and chipped tooth
C11		Chipped tooth & wear of outer race of bearing
C12		Gear tooth broken and wear of cage and rolling elements of bearing
C13		Gear tooth broken and wear of cage and rolling elements of bearing & wear of outer race of bearing

rig includes a computer for data acquisition, an electric load simulator, a prime mover, a gearbox, a flywheel, and an asynchronous generator. As it is not realistic to implement the diagnostic system to monitor all the components in the GTGS in one study, this research selected the fault detection of the gearbox as a case study. The test rig can simulate many common faults in the gearbox of the GTGS, such as unbalance, misalignment, and gear crack. A total of 13 cases, including one normal case, 8 single-faults, and 4 simultaneous-faults in the gearbox, were simulated in the test rig in order to generate sample training and test dataset. Some samples for single-fault and simultaneous-fault patterns are shown in Figures 6 and 7, respectively. Figures 6 and 7 show that the signal profiles between single-faults and simultaneous-faults are very similar, which make them difficult to be distinguished manually, but their degrees of similarities can be detected using the proposed framework.

Table 1 shows the detailed descriptions of the thirteen simulated cases in which the mechanical misalignment of the gearbox was simulated by adjusting one height of the gearbox with shims, and the mechanical unbalance was simulated by adding one eccentric mass on the output shaft. The sample vibration data were acquired by two triaxial accelerometers located on the outer case of the gearbox as shown in Figure 5. The accelerometers are used to record the gearbox vibration signals along the horizontal and vertical directions, respectively. In the axial direction, the vibration signal is ignored since the test rig uses spur gears in which the vibration along the axial direction is not obvious. To construct and test the diagnostic framework, each simulated single fault was repeated 200 times and 100 times for each simultaneous-fault under various random electric loads. Each time, 2 seconds of vibration data was recorded with a sampling rate of 2048 Hz.

TABLE 2: Division of the sample dataset into different subsets.

	Type of dataset	Single-faults (1800)	Simultaneous-faults (400)
Raw sample data	Validation dataset	$D_{Valid.1}$ (360)	$D_{Valid.S}$ (120)
	Training dataset	$D_{Train.1}$ (1080)	$D_{Train.S}$ (160)
	Test dataset	$D_{Test.1}$ (360)	$D_{Test.S}$ (120)
Feature extraction (WPT + TDSF + KPCA)	Validation dataset	$D_{Proc.Valid.1}$ (360)	$D_{Proc.Valid.S}$ (120)
	Training dataset	$D_{Proc.Train.1}$ (1080)	$D_{Proc.Train.S}$ (160)
	Test dataset	$D_{Proc.Test.1}$ (360)	$D_{Proc.Test.S}$ (120)

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 \times 2 measurement directions \times 2 seconds \times 2048) 16384 data points. Finally, there were 1800 single-fault sample data (i.e., (1 normal case + 8 kinds of single-faults) \times 200 samples) and 400 simultaneous-fault sample data (i.e., 4 kinds of simultaneous-faults \times 100 samples). In order to test the diagnostic performance for both single-faults and simultaneous-faults, the sample data was divided into different subsets as shown in Table 2, where $D_{Valid.1}$ denotes the validation set of 360 single-fault patterns without feature extraction, and $D_{Valid.S}$ denotes the validation set of the extracted features of 120 simultaneous-fault patterns.

3.2. Feature Extraction by WPT and TDSF. Feature extraction is the determination of a feature vector from a signal with minimal loss of important information. 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, and L is the level of WPT decomposition.

With reference to the literature [8], statistical characteristics can be used for representing signal subbands effectively so that the dimensionality of a feature vector extracted from vibration signals can be reduced. For each signal subband, the statistical characteristics include the following:

- (1) maximum of the wavelet coefficients,
- (2) minimum of the wavelet coefficients,
- (3) mean of the wavelet coefficients,
- (4) standard deviation of the wavelet coefficients.

In this case study, for a vibration signal of 4096 sample points, there are 4096 wavelet coefficients and $2^4 = 16$ subbands after $L = 4$ levels of WPT decomposition. However, using the previous statistics, there are only $2^4 \times 4 = 64$ features which greatly reduce the input complexity (from 4096 inputs to 64 inputs) for the next stage. For a good presentation,

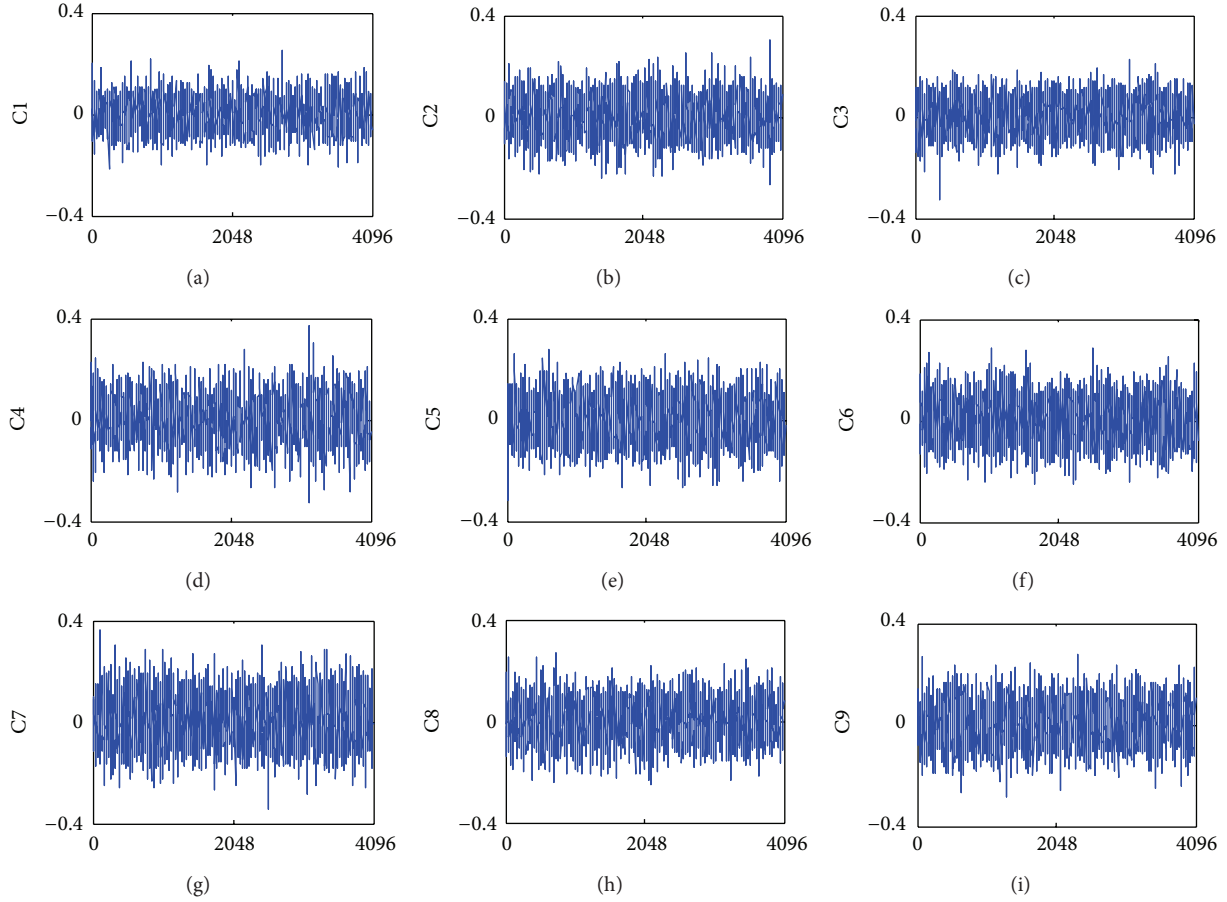


FIGURE 6: Sample normalized single-fault and normal patterns of GTGS.

the term of WPT hereafter refers to the process of wavelet packet decomposition together with the previous calculating statistics.

After decomposition by WPT, time-domain statistical method is usually further employed to extract time-domain features of the raw signals which provide the physical characteristics of time series data. For instance, references [12, 13] applied time-domain statistical features for fault detection on gear trains and low speed bearings, such as mean, standard deviation, skewness, crest factor, and kurtosis, respectively. In this study, 10 statistical time-domain features are employed to analyze the vibration signal. Table 3 presents the statistical time-domain features. After 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, the dimension of these extracted features is still high. Such high dimension 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 having even contribution, all reduced features should go through normalization. The interval of normalization is within $[-1, 1]$. The extracted feature is normalized by the following formula:

$$y = \frac{(D_{\text{KPCA}} - D_{\text{min}})}{(D_{\text{max}} - D_{\text{min}})}, \quad (11)$$

where D_{KPCA} is an output feature after going through KPCA and y is the result of normalization. After normalization, a processed dataset D_{Proc} is obtained. The pairwise-coupled probabilistic classification algorithm can then be employed to construct the fault classifier based on $D_{\text{Proc-TRAIN}}$.

4. Experimental Results and Discussion

To verify the effectiveness of the proposed framework, a set of experiments using different combinations of methods were carried out by using the validation and test datasets. The performance evaluations for all experiments were done based on the F -measure.

4.1. Selection of Kernel Parameters, Kernel Functions, and Decision Thresholds. Since there are many combinations of mother wavelets and kernels for the probabilistic classifier,

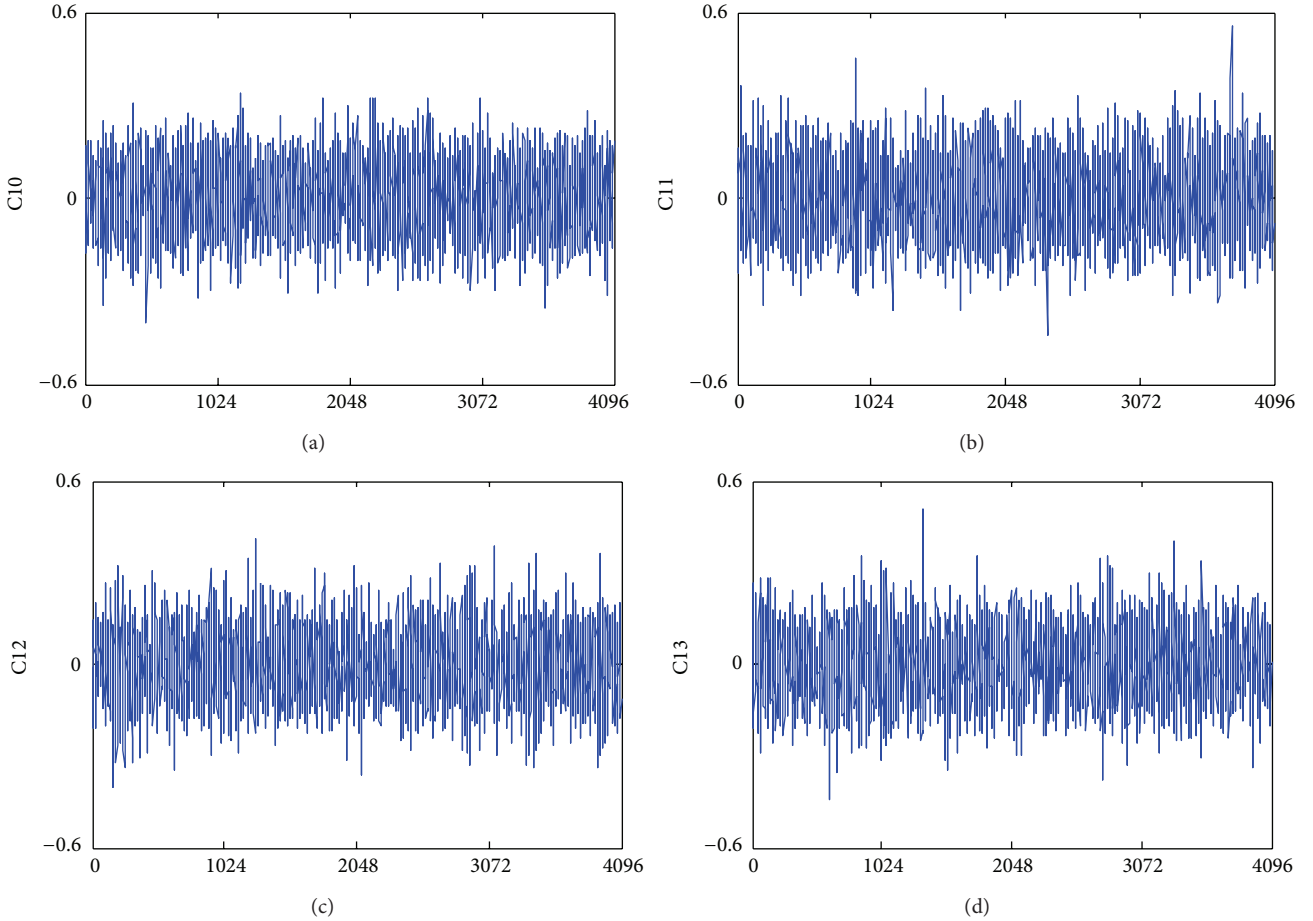


FIGURE 7: Sample normalized simultaneous-fault patterns of GTGS.

a set of experiments were carried out to determine the best combination of the system configuration. In the phase of WPT, mother wavelet and the level of decomposition L were 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, three Daubechies wavelets Db3, Db4, and Db5 were tried, and, hence, the range of L was set from 3 to 5. Moreover, three different kernel functions for KPCA, namely, *linear*, *radial basis function (RBF)*, and *polynomial*, were tested. Different kernel functions have various hyperparameters for adjustment. However, it is very time consuming to try all the different values of hyperparameters. To reduce the number of trials, the hyperparameter R of RBF based on 2^ν was tried for ν ranged from -3 to $+3$, and the hyperparameter d' of polynomial kernel was taken from 2 to 5. Moreover, the common width for both RVM and PCRVM and the common spread for both PNN and PCPNN and the decision threshold were, respectively, assumed to be 1, 1 and 0.5 in the course of determination of the best configuration of feature extraction module. Under this configuration, the best combination of kernel function and its parameter for feature extraction (WPT + TDSF + KPCA) is presented in Table 5.

According to experimental results not listed here, the classifiers using KPCA under the *polynomial* kernel with $d' = 4$ and the mother wavelet of Db4 with level 4 reach the highest diagnostic accuracies, and, hence, this combination of feature extraction techniques was finally selected. Table 5 also indicates that 77 principal components are obtained by using this combination of feature extraction techniques. In other words, a raw signal of 16384 data points can be transformed to 77 features as the input variables of the classifiers.

After determining the configuration of the feature extraction module, the next step is to determine the optimal parameters of the classifiers by using the grid search. As mentioned previously, different probabilistic classifiers have their own hyperparameters for tuning. PNN/PCPNN uses spread s and RVM/PCRVM employs width w . In this case study, the value of s was examined from 1 to 3, at an interval of 0.5, and w was selected from 1 to 8 at an interval of 0.5. To find the optimal decision threshold, the search region was set within 0 to 1, at an interval of 0.01. Then 5-fold cross-validation was applied to $D_{\text{Proc.Vali}}$ to determine the best combination of the parameters. Finally, the best hyperparameters and thresholds for PCRVM (w, ϵ) and PCPNN (s, ϵ) were found to be (6.5, 0.67) and (1, 0.76), respectively.

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

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

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

TABLE 4: Features extracted under different decomposition levels of WPT.

WPT	Features obtained by WPT	+ TDSF	× (measurement directions × locations of accelerometers)	= Total features
3 Level	2^3 subbands × 4	10	2×2	168
4 Level	2^4 subbands × 4	10	2×2	296
5 Level	2^5 subbands × 4	10	2×2	552

4.2. Evaluation of the Proposed Framework. To evaluate the effectiveness of the proposed probabilistic classifier, pairwise coupling, and feature extraction strategies, eight experiments were conducted, and their results are presented in Table 6. The objectives are to compare the diagnostic framework with versus without feature extraction, classification method using single-label SVM classifier versus multilabel PNN and RVM classifiers, and the framework with versus without pairwise coupling strategy. Since different aspects of the proposed framework are investigated for single-faults and simultaneous-faults, various combinations of training and test datasets are employed. For example, experiments 1, 3, 5, and 7 have no feature extraction so that the training dataset only uses the raw dataset $D_{\text{Train}} (\mathbf{D}_1 \cup \mathbf{D}_S)$, whereas the feature extracted dataset $D_{\text{Proc}} (\mathbf{F}_1 \cup \mathbf{F}_S)$ was selected as training data for Experiments 2, 4, 6 and 8. As described in Section 4.1,

the best decision threshold for both PNN and PCPNN is 0.76, while 0.67 is for both RVM and PCRVM. Table 6 reveals that the experiments with feature extraction 2, 4, 6, and 8, show an increase of diagnostic accuracy by 4%–8% as compared with the experiments without feature extraction (1, 3, 5, and 7). These results indicate that the proposed feature extraction method (WPT + TDSF + KPAC) is effective.

To verify the effectiveness of the pairwise coupling strategy, a set of experiments without pairwise coupling were carried out using one-versus-all strategy, and the experimental results are shown in Table 6 (experiments 1 to 4). Comparing the results with the results of experiments 5 to 8 where pairwise coupling strategy is employed, the accuracies in the experiments without pairwise coupling are generally 2% to 5% worse. The main reason is that only d binary classifiers were constructed for d labels in the one-versus-all strategy, so that there are many indecision regions between pairs of classes. Therefore, when a test case lies on these regions, the classifiers mostly fail to classify the faults correctly. However, the classifiers with pairwise coupling strategy (i.e., PCPNN and PCRVM) can minimize those indecision regions. Hence, a probabilistic classifier with pairwise coupling strategy is an effective approach to improve the diagnosis accuracy. In this case study, the proposed framework in experiment 8 achieves the best F -measure of 0.9129. Note that the proposed framework only employs a training set of single-fault patterns to construct the classifier while the overall performance is evaluated over both single-fault and simultaneous-fault test patterns. Therefore, the proposed method can successfully detect simultaneous-faults without costly simultaneous-fault training patterns. The effectiveness of the proposed framework is further evaluated by looking at the size of the training dataset. In the training of the proposed diagnostic model, only single-fault patterns are processed. From Table 2, there are 160 simultaneous-fault patterns in D_{Train_S} or D_{Proc_S} while there are totally 1240 (i.e., 1080 + 160) patterns of both single-faults and simultaneous-faults. When the 160 simultaneous-fault patterns are not required, a significant reduction of training patterns at 12.9% (i.e., $(160/1240) \times 100\%$) is achieved. In fact, the time reduction is even more significant because simultaneous-faults are more costly and difficult to acquire.

In order to further verify the superiority of the proposed framework, the latest single-label binary classification framework based on SVM [28] was also applied to the same datasets for comparison. The evaluation results are shown in Table 7, in which the overall diagnostic accuracy of the single-label method is lower than that of the proposed PCRVM framework. Therefore, it can be concluded that the proposed framework is currently the best method for single and simultaneous-fault diagnosis of the GTGS.

5. Conclusions

In this paper, simultaneous-fault diagnosis for the GTGS is studied. A systematic framework combining feature extraction, pairwise coupling probabilistic classification, and parameter optimization based on a partial-match assessment has been developed to overcome the challenges

TABLE 5: Accuracies of various classifiers under the best combination of polynomial kernel of KPCA and mother wavelet of Db4/L4.

Mother wavelet	Level L	No. of features reduced	KPCA with polynomial kernel and $d' = 4$								
			Single-faults			Simultaneous-faults			Overall faults		
			SVM F_{me}	PCPNN F_{me}	PCRVM F_{me}	SVM F_{me}	PCPNN F_{me}	PCRVM F_{me}	SVM F_{me}	PCPNN F_{me}	PCRVM F_{me}
Db4	4	77	0.9594	0.9298	0.9336	0.4546	0.7526	0.8364	0.7725	0.8824	0.9129

TABLE 6: Evaluation of different combinations of techniques using the best model parameters obtained.

Classifier	Feature extraction	Strategy	Experiment no.	Decision threshold	Accuracies for test cases (F_{me})		
					Single-faults	Simultaneous-faults	Overall faults
PNN	None	One-versus-all	1	$\varepsilon = 0.76$	0.8364	0.7156	0.7576
	WPT + TDSF + KPCA	One-versus-all	2	$\varepsilon = 0.76$	0.8646	0.7246	0.8121
RVM	None	One-versus-all	3	$\varepsilon = 0.67$	0.8299	0.7464	0.8125
	WPT + TDSF + KPCA	One-versus-all	4	$\varepsilon = 0.67$	0.8960	0.7964	0.8536
PCPNN	None	Pairwise coupling	5	$\varepsilon = 0.76$	0.8923	0.7364	0.7824
	WPT + TDSF + KPCA	Pairwise coupling	6	$\varepsilon = 0.76$	0.9198	0.7526	0.8661
PCRVM	None	Pairwise coupling	7	$\varepsilon = 0.67$	0.8795	0.7843	0.8497
	WPT + TDSF + KPCA	Pairwise coupling	8	$\varepsilon = 0.67$	0.9336	0.8364	0.9129

TABLE 7: Comparison of SVM, PCPNN, and PCRVM.

Classifier	Feature extraction	Strategy	Experiment no.	Decision threshold	Accuracies for test cases (F_{me})		
					Single-faults	Simultaneous-faults	Overall faults
SVM	WPT + TDSF + KPCA	One-versus-all	9	Not applicable because of binary outputs of SVM	0.9594	0.4546	0.7725
PCPNN	WPT + TDSF + KPCA	Pairwise coupling	10	$\varepsilon = 0.76$	0.9198	0.7526	0.8661
PCRVM	WPT + TDSF + KPCA	Pairwise coupling	11	$\varepsilon = 0.67$	0.9336	0.8364	0.9129

of simultaneous-fault diagnosis for complicated mechanical systems. In the proposed framework, the feature extraction module is designed by combining the techniques of WPT + TDSF + KPCA to effectively capture the single-fault components in the simultaneous-fault vibration patterns. A pairwise coupling strategy is also employed to deal with the interaction between the independent labels, which outperforms the approaches without pairwise coupling by 2% to 5% in simultaneous-fault diagnosis. In short, this research work makes contributions to four aspects: (1) it is the first research to find that the features of single-fault signal patterns of the GTGS can be found from their simultaneous-fault patterns by the proposed feature extraction method; (2) a high diagnostic accuracy for both unseen single-fault and simultaneous-fault patterns is achieved by the proposed framework; (3) the proposed framework can achieve a high diagnostic efficiency while the request for large amount of expensive simultaneous-fault training data is no longer necessary; (4) it is the original application of the proposed framework to the problem of GTGS diagnosis. Since the proposed framework

for simultaneous-fault diagnosis is general, it could be applied to other similar industrial problems.

Nomenclature

A :	Diagonal matrix of hyperparameters
B :	Diagonal matrix in RVM
C_i :	i th probabilistic classifier
$C_i(\mathbf{f})$:	Probability of \mathbf{f} belonging to the i th label
C_{ik} :	Pairwise classifier
$C_{ik}(\mathbf{f})$:	Probability of \mathbf{f} belonging to the i th label against the j th label
\mathbf{D}_1 :	Dataset of single-faults
\mathbf{D}_S :	Dataset of simultaneous-faults
D_{Proc} :	Feature matrix after feature extraction
$D_{Proc-Train}$:	Training feature matrix after feature extraction
$D_{Proc-Vali}$:	Validation feature matrix after feature extraction

$D_{\text{Proc-Test}}$: Test feature matrix after feature extraction
 D_i : i th raw signal data
 D_{max} : Upper limit of feature
 D_{min} : Lower limit of feature
 D_{WPT} : Feature set extracted by WPT and TDSF
 d : Number of single-fault labels
 d' : Hyperparameter of polynomial kernel function of KPCA
 F_1 : Feature matrix of single-faults
 F_S : Feature matrix of simultaneous-faults
 F_{me} : F -measure
 f : Feature vector
 f_n : n th feature vector
 F : Set of feature vectors
 $K(\cdot)$: Kernel function of KPCA
 L : Decomposition level of Db mother wavelet
 l_{class} : Probabilistic classifier
 N : Number of training data
 $P(\cdot)$: Probability
 s : Spread of PNN
 s_{opt} : Optimal spread of PNN
 t : Set of fault labels in training dataset
 t_n : Fault label of the n th training case
 w : Width of RVM
 w_{opt} : Optimal width of RVM
 w_{MP} : Most probable weight vector in RVM
 x : Raw signal data
 y : Predicted label vector
 y_j : j th predicted label
 y : Normalized feature
 $y(f)$: Predicted decision
 $z_r(D_{\text{WPT}})$: Transformed variables z_r for vector D_{WPT}
 α : Hyperparameter vector of RVM
 α_i : i th hyperparameter in vector α
 $\alpha_{r,I}$: I th element in eigenvector α corresponding to the r th largest eigenvalue
 λ_r : r th eigenvalue of KPCA
 ε : Decision threshold
 ε_{opt} : Optimal decision threshold
 ρ : Probability vector
 ρ_j : Probability of the j th vector
 Σ : Covariance matrix in RVM
 Σ_{ii} : i th diagonal element of covariance matrix Σ
 $\sigma(\cdot)$: Logistic sigmoid function
 Φ : Design matrix in RVM.

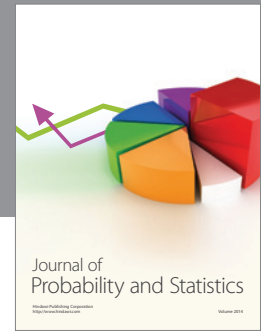
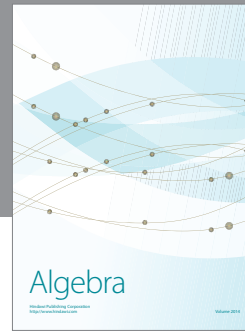
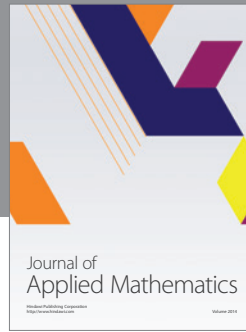
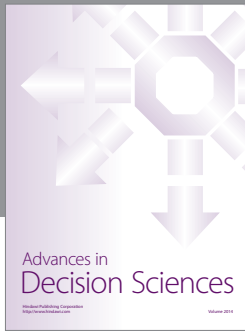
Acknowledgment

The authors would like to thank the funding support by the University of Macau, Grant nos. MYRG153(Y1-L2)-FST11-YZX and MYRG075(Y2-L2)-FST12-VCM.

References

- [1] C. Ma, X. Gu, and Y. Wang, "Fault diagnosis of power electronic system based on fault gradation and neural network group," *Neurocomputing*, vol. 72, no. 13–15, pp. 2909–2914, 2009.
- [2] C. F. Yan and H. Zhang, "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, pp. 7–16, 2009.
- [3] J. Rafiee, F. Arvani, A. Harifi, and M. H. Sadeghi, "Intelligent condition monitoring of a gearbox using artificial neural network," *Mechanical Systems and Signal Processing*, vol. 21, no. 4, pp. 1746–1754, 2007.
- [4] J.-D. Wu and J.-J. Chan, "Faulted gear identification of a rotating machinery based on wavelet transform and artificial neural network," *Expert Systems with Applications*, vol. 36, no. 5, pp. 8862–8875, 2009.
- [5] 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.
- [6] A. Widodo, B.-S. Yang, and T. Han, "Combination of independent component analysis and support vector machines for intelligent faults diagnosis of induction motors," *Expert Systems with Applications*, vol. 32, no. 2, pp. 299–312, 2007.
- [7] J. Cheng, D. Yu, J. Tang, and Y. Yang, "Application of SVM and SVD technique based on EMD to the fault diagnosis of the rotating machinery," *Shock and Vibration*, vol. 16, no. 1, pp. 89–98, 2009.
- [8] C. M. Vong and P. K. Wong, "Engine ignition signal diagnosis with wavelet packet transform and multi-class least squares support vector machines," *Expert Systems with Applications*, vol. 38, no. 7, pp. 8563–8570, 2011.
- [9] A. Widodo and B.-S. Yang, "Support vector machine in machine condition monitoring and fault diagnosis," *Mechanical Systems and Signal Processing*, vol. 21, no. 6, pp. 2560–2574, 2007.
- [10] B. Samanta, "Gear fault detection using artificial neural networks and support vector machines with genetic algorithms," *Mechanical Systems and Signal Processing*, vol. 18, no. 3, pp. 625–644, 2004.
- [11] S. Abbasion, A. Rafsanjani, A. Farshidianfar, and N. Irani, "Rolling element bearings multi-fault classification based on the wavelet denoising and support vector machine," *Mechanical Systems and Signal Processing*, vol. 21, no. 7, pp. 2933–2945, 2007.
- [12] Q. Hu, Z. J. He, Z. S. Zhang, and Y. Y. Zi, "Fault diagnosis of rotating machinery based on improved wavelet package transform and SVMs ensemble," *Mechanical Systems and Signal Processing*, vol. 21, no. 2, pp. 688–705, 2007.
- [13] 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.
- [14] K. Watanabe, S. Hirota, L. Hou, and D. M. Himmelblau, "Diagnosis of multiple simultaneous fault via hierarchical artificial neural networks," *AIChE Journal*, vol. 40, no. 5, pp. 839–848, 1994.
- [15] R. Eslamloueyan, "Designing a hierarchical neural network based on fuzzy clustering for fault diagnosis of the Tennessee-Eastman process," *Applied Soft Computing Journal*, vol. 11, no. 1, pp. 1407–1415, 2011.
- [16] R. Eslamloueyan, M. Shahrokhi, and R. Bozorgmehri, "Multiple simultaneous fault diagnosis via hierarchical and single artificial neural networks," *Scientia Iranica*, vol. 10, no. 3, pp. 300–310, 2003.
- [17] C.-M. Vong, P.-K. Wong, and W.-F. Ip, "Case-based expert system using wavelet packet transform and kernel-based feature

- manipulation for engine ignition system diagnosis,” *Engineering Applications of Artificial Intelligence*, vol. 24, no. 7, pp. 1281–1294, 2011.
- [18] Y. G. Lei, Z. J. He, Y. Y. Zi, and X. F. 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.
- [19] Q. B. 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.
- [20] C.-L. Huang and C.-J. Wang, “A GA-based feature selection and parameters optimization for support vector machines,” *Expert Systems with Applications*, vol. 31, no. 2, pp. 231–240, 2006.
- [21] P. M. Ciarelli, E. Oliveira, C. Badue, and A. F. de Souza, “Multi-label text categorization using a probabilistic neural network,” *International Journal of Computer Information Systems and Industrial Management Application*, vol. 1, no. 1, pp. 133–144, 2009.
- [22] P. Refregier and F. Vallet, “Probabilistic approach for multiclass classification with neural network,” in *Proceedings of the International Conference on Artificial Networks*, pp. 1003–1007, 1991.
- [23] A. Widodo, E. Y. Kim, J.-D. Son et al., “Fault diagnosis of low speed bearing based on relevance vector machine and support vector machine,” *Expert Systems with Applications*, vol. 36, no. 3, pp. 7252–7261, 2009.
- [24] M. E. Tipping, “Sparse Bayesian learning and relevance vector machine,” *Journal of Machine Learning Research*, vol. 1, no. 3, pp. 211–244, 2001.
- [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*, Advances in Pattern Recognition, Springer, London, UK, 2nd edition, 2010.
- [27] C. M. Vong, P. K. Wong, W. F. Ip, and C. C. Chiu, “Simultaneous fault diagnosis of automotive engine ignition systems using prior domain knowledge and relevance vector machine,” *Mathematical Problems in Engineering*, vol. 2013, Article ID 974862, 19 pages, 2013.
- [28] I. Yélamos, M. Graells, L. Puigjaner, and G. Escudero, “Simultaneous fault diagnosis in chemical plants using a multilabel approach,” *AIChE Journal*, vol. 53, no. 11, pp. 2871–2884, 2007.
- [29] 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.
- [30] P.-K. Wong, C.-M. Vong, L.-M. Tam, and K. Li, “Data preprocessing and modelling of electronically-controlled automotive engine power performance using kernel principal components analysis and least squares support vector machines,” *International Journal of Vehicle Systems Modelling and Testing*, vol. 3, no. 4, pp. 312–330, 2008.
- [31] T. Hastie and R. Tibshirani, “Classification by pairwise coupling,” *The Annals of Statistics*, vol. 26, no. 2, pp. 451–471, 1998.
- [32] C. M. Vong, P. K. Wong, and W. F. Ip, “A new framework of simultaneous-fault diagnosis for time-dependent patterns,” *IEEE Transactions on Industrial Electronics*, vol. 60, no. 8, pp. 3372–3385, 2013.
- [33] R. Baeza-Yates and B. Ribeiro-Neto, *Modern Information Retrieval*, ACM Press, Addison-Wesley, England, UK, 1999.
- [34] C. J. V. Rijsbergen, *Information Retrieval*, Butterworths, London, UK, 1979.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

