

## Research Article

# Simultaneous-Fault Diagnosis of Automotive Engine Ignition Systems Using Prior Domain Knowledge and Relevance Vector Machine

Chi-Man Vong,<sup>1</sup> Pak-Kin Wong,<sup>2</sup> Weng-Fai Ip,<sup>3</sup> and Chi-Chong Chiu<sup>1</sup>

<sup>1</sup> Department of Computer and Information Science, University of Macau, Taipa, Macau

<sup>2</sup> Department of Electromechanical Engineering, University of Macau, Taipa, Macau

<sup>3</sup> Faculty of Science and Technology, University of Macau, Macau

Correspondence should be addressed to Chi-Man Vong; [cmvong@umac.mo](mailto:cmvong@umac.mo)

Received 30 November 2012; Accepted 12 December 2012

Academic Editor: Qingsong Xu

Copyright © 2013 Chi-Man Vong 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.

Engine ignition patterns can be analyzed to identify the engine fault according to both the specific prior domain knowledge and the shape features of the patterns. One of the challenges in ignition system diagnosis is that more than one fault may appear at a time. This kind of problem refers to simultaneous-fault diagnosis. Another challenge is the acquisition of a large amount of costly simultaneous-fault ignition patterns for constructing the diagnostic system because the number of the training patterns depends on the combination of different single faults. The above problems could be resolved by the proposed framework combining feature extraction, probabilistic classification, and decision threshold optimization. With the proposed framework, the features of the single faults in a simultaneous-fault pattern are extracted and then detected using a new probabilistic classifier, namely, pairwise coupling relevance vector machine, which is trained with single-fault patterns only. Therefore, the training dataset of simultaneous-fault patterns is not necessary. Experimental results show that the proposed framework performs well for both single-fault and simultaneous-fault diagnoses and is superior to the existing approach.

## 1. Introduction

*1.1. Background of Engine Ignition Patterns.* Although automotive engine ignition systems vary in construction, they are similar in basic operation. All of them have a primary circuit that causes a spark in the secondary circuit, which is then delivered to the correct spark plug at the proper time. The conditions inside the ignition system and the cylinder also affect the ignition pattern in the secondary circuit. Consequently, the ignition patterns reflect the conditions within the ignition system and help pinpoint their faults [1], such as wide or narrow spark-plug gaps and open spark-plug cables. After capturing the ignition pattern, the automotive mechanic compares the features of the captured pattern with samples from handbooks for diagnosis [2, 3]. This procedure is called ignition system diagnosis. However, there are several challenges for the automotive mechanic which are as follows.

- (1) The engine ignition pattern is time dependent. Different engine models produce the ignition patterns of various amplitude and duration for the same kind of fault. Even for the same engine, it may produce slightly different shapes of ignition patterns for each engine cycle due to engine speed fluctuation and various testing conditions. Therefore, there is no exact scale and duration for sample patterns in the handbooks. Hence, the traditional diagnosis merely relies on prior domain knowledge and the engineer's experience.
- (2) Practically, the engine ignition-system diagnosis is a simultaneous-fault problem, but many handbooks only provide single-fault patterns for reference. To determine simultaneous faults, the engineer can only extract and analyze some specific features of

single-fault patterns from a simultaneous-fault pattern, such as frequency, firing voltage, and burn time, and make a decision about the presence of simultaneous faults according to their experience and knowledge.

- (3) As suggested in the existing literature [1–3], the ignition-system diagnosis based on the shape features and the prior domain knowledge of the ignition pattern cannot conclude a definite answer. It is because many possible faults may occur individually or simultaneously. The handbooks do not provide the rank of the probability of each possible fault. Therefore, to find out a fault based on ignition patterns, many trials for disassembling and assembling of engine parts are often necessary unless the engineer has very rich experience.

To tackle these challenges, an effective feature extraction method for engine ignition patterns is required, which combines domain knowledge (DK), time-frequency decomposition, and dimensional reduction techniques. Moreover, an advanced probabilistic classifier is necessary to provide the rank of each possible fault and reliable diagnostic results. In recent years, some intelligent diagnostic methods based on pattern recognition have been developed for multiclass fault diagnosis (i.e., single-fault diagnosis because only a single fault is identified) of mechanical systems [4–9]. Generally, these methods include two steps: feature extraction and classification.

*1.2. Feature Extraction Methods.* Feature extraction is very important because the in-depth and hidden features of single-fault patterns can be detected through frequency subband decomposition. Referring to the existing literature, many classical feature extraction techniques were applied to fault diagnosis; the most typical one is the fast Fourier transform (FFT) [10–13]. However, its main drawback is the unsuitability for nonstationary patterns. Wavelet packet transform (WPT) [1, 4, 14–19] is another popular time frequency localization analysis method that received a widespread utilization in the past decade. By means of multiscale analysis, WPT can be successfully applied to nonstationary patterns, based on subband coding and a systematic decomposition of a pattern into its subband levels for pattern analysis. Therefore, WPT is employed in this research for feature extraction.

Nevertheless, one drawback for WPT is that the size of the extracted features is larger or equal to that of its original pattern. If the original pattern is of a high dimension, there is a large amount of extracted features that may incur two issues: (1) the high complexity of the trained classifiers because of the huge amount of inputs; (2) there may be many redundant and unimportant extracted features so that noise can be induced. Both of the issues can degrade the classifier performance. Therefore, compensating the drawback by employing dimensional reduction technique such as principal component analysis (PCA) [20–22] is suggested. In this research, PCA is selected as the dimensional reduction technique for a simple illustration purpose. More advance techniques could be considered in the future. Compared to other

dimensional reduction techniques, PCA has three advantages: (1) it has no hyperparameter; (2) PCA eliminates the interaction of variables because the principal components are independent of each other; (3) the principal components are sorted by their information weights, so some unimportant principal components can be further reduced. Then, the feature extraction approach of WPT+PCA can transform an original ignition pattern into a reduced dimensional feature vector while retaining most of the information content.

*1.3. Classification Methods.* For classification, a fault can be considered as a label, no matter whether it is a single fault or simultaneous fault. To date, there are only a few researches on simultaneous-fault diagnosis. The typical classification method for simultaneous-fault diagnosis is to build a number of classifiers according to the combination of all possible faults; this method is called monolabel classification [23]. However, it is practically difficult to obtain the training data of all possible combinations particularly for ignition patterns. Normally, the number of combination of all faults in an engineering problem is very large that affects the diagnostic accuracy because the complexity of the classifiers will also be immensely increased. Moreover, if a new single fault is added in the future, the number of required training simultaneous-fault patterns grows significantly. To overcome this drawback, Yélamos et al. [23] proposed a binarization strategy using support vector machine (SVM) and applied to simultaneous-fault diagnosis of a simulated chemical process based on time-independent data, in which the labels of the single faults or simultaneous faults were processed as binary vectors, that is, 0 or 1 only. For each label, a binary classifier was constructed using SVM with one-versus-all splitting strategy. Given an unknown pattern, the classifier would output a vector of binary results (0 or 1). From this approach, only single-fault patterns are used for training the classifiers while simultaneous-fault patterns are not necessary. The experimental results showed that the overall accuracy of their binarization approach is almost the same as that of the traditional monolabel approach. This kind of binarization approach sounds good but still suffers from several drawbacks: (1) the approach assumes that informative features are obvious and available that is not always the case for time-dependent signal patterns, so this approach cannot be suitable for ignition patterns; (2) the one-versus-all strategy ignores the pairwise correlation between the labels and hence the classification accuracy is mostly degenerated; (3) the approach only considers the presence of a fault, if its corresponding output is close to the classification margin which lacks confidence of correct classification, that is, the degree of belief of faults.

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 incorrect in the problem. Therefore, it is better to employ probabilistic classifier for simultaneous-fault diagnosis. The probabilistic structure is also suitable for the fault with uncertainty such as engine ignition-system diagnosis. Typically, probabilistic neural network (PNN) [24, 25] was

employed as a probabilistic classifier. It was shown in [24] 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. [6] proposed to apply an advanced classifier, namely, relevance vector machine (RVM) to fault diagnosis of low speed bearings. They showed that RVM is superior to SVM in terms of diagnostic accuracy. Besides, RVM can also handle regression problem [26]. RVM is a statistical learning method proposed by Tipping [27], 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 [28, 29]. In view of this drawback, this research is the first in the literature to incorporate pairwise coupling, that is, one-versus-one (1v1) strategy, into RVM, namely pairwise coupled relevance vector machine (PCRVM). As PCRVM considers the correlation between every pair of fault labels, a more accurate estimate of label probabilities for simultaneous-fault signals can be achieved.

*1.4. Decision Threshold Optimization.* 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 [0.21, 0.5, 0.69, 0.01, 0.6]. 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 new step of decision threshold optimization is proposed in the current framework other than feature extraction and probabilistic classification.

*1.5. Research Objectives and the Proposed Framework.* Currently, very little research examines whether the features of single-fault ignition patterns can be reflected in the ignition patterns of some simultaneous faults. If it is feasible, some rational (not all) simultaneous faults are likely to be identified based on the prior domain knowledge and the features of single-fault ignition patterns. In other words, the features about the single faults in a simultaneous-fault pattern could be detected and then classified using the probabilistic classifier trained with the single-fault patterns only. Under this concept, the simultaneous-fault patterns are not necessary for training the classifiers. Once a new single fault is added in the future, the diagnostic system can be easily extended because the issue of combinatory single faults has been eliminated. To verify the feasibility and determine the best feature extraction method, this research proposes to extract the important

knowledge-specific, time-domain, and frequency-domain features of the single-fault patterns using the combination of WPT+PCA, FFT, and DK. Then the pairwise coupled probabilistic classifier is trained using a training dataset of these extracted single-fault features in order to identify simultaneous faults for reasonable unseen patterns. Therefore, a feasibility study on this idea for simultaneous-fault diagnosis is an important contribution of this research. Another important contribution of the research is the reduction of required training patterns for simultaneous-fault diagnosis.

This paper is organized as follows. The proposed framework and the related techniques are described in Section 2. In Section 3, the experimental setup is presented, followed by the results and a comparison with latest approach [23] in Section 4 and discussion in Section 5. Finally, a conclusion is given in Section 6.

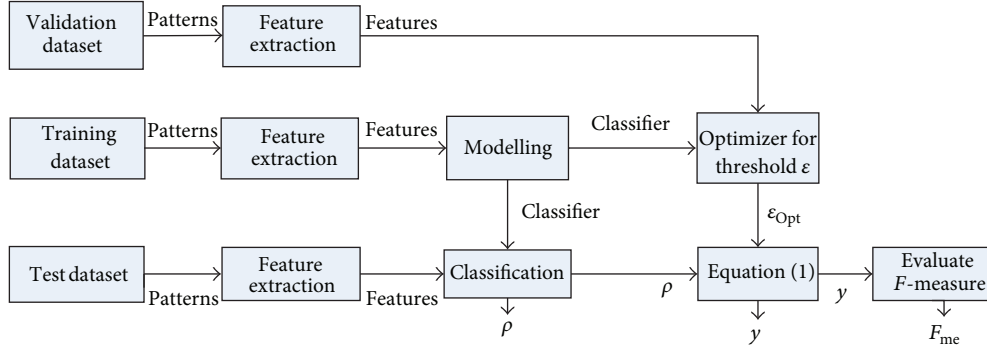
## 2. Proposed Framework and Related Techniques

The proposed diagnosis framework (Figure 1) includes three steps: feature extraction, classification, and threshold optimization. The framework is general so that different feature extraction, probabilistic classification, and threshold optimization techniques could be adopted. In this paper, FFT, WPT, and PCA are examined in the step of feature extraction and their detailed descriptions can be, respectively, found in [22, 30, 31]. In addition, these techniques are combined, respectively, with time-related domain knowledge (DK) for a comprehensive comparison.

*2.1. Formulation of the Proposed Framework.* Given a sample dataset  $\mathbf{D} = \{(\mathbf{x}_i, \mathbf{l}_i)\}$  of (single-fault or simultaneous-fault) patterns,  $i = 1$  to  $N_D$ ,  $\mathbf{x}_i \in \mathbf{R}^n$  and  $\mathbf{l}_i = [l_{i1}, l_{i2}, \dots, l_{id}]$  is a vector of labels of the corresponding single-fault pattern of  $\mathbf{x}_i$  and  $d$  is the number of single faults. Here there may be more than one fault in  $\mathbf{l}_i$  so that  $\sum_{g=1}^d l_{ig} \geq 1$ ,  $l_{ig} \in \{0, 1\}$  for  $g = 1$  to  $d$ . In Figure 1, the sample dataset is divided into three groups: training dataset, validation dataset, and test dataset where training dataset only involves single-fault patterns.

After applying feature extraction techniques to the patterns  $\{\mathbf{x}_i\}$ , a set of feature vectors  $\mathbf{F} = \{(\mathbf{f}_i, \mathbf{l}_i)\}$  is produced. A training dataset of single-fault patterns only (no simultaneous-fault patterns are necessary) is selected to train a multilabel classifier  $f_{\text{class}}$  by using probabilistic classification algorithm. Then  $f_{\text{class}}$  takes an unknown feature vector  $\mathbf{f}$  as input and outputs a probability vector  $\boldsymbol{\rho} = [\rho_1, \rho_2, \dots, \rho_d]$  where  $d$  is the number of the single-fault labels. Here  $\rho_j = P(l_j | \mathbf{f}) \in [0, 1]$  denotes the probability that  $\mathbf{f}$  belongs to the  $j$ th label for  $j = 1$  to  $d$ . Since every  $\rho_j$  is an independent probability,  $\sum \rho_j$  is not necessarily equal to one. At this stage, the diagnostic system can provide the probability vector  $\boldsymbol{\rho}$  to the user as a quantitative measure for reference and further use. Afterwards, the multilabel decision vector  $\mathbf{y} = [y_1, y_2, \dots, y_d]$  is constructed from  $\boldsymbol{\rho}$  using (1):

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



Both  $\rho$  and  $y$  will be offered to the user for decision making

FIGURE 1: Proposed framework of the simultaneous engine ignition-fault diagnosis system and its evaluation.

where  $\varepsilon \in (0, 1)$  is a user-defined decision threshold and  $y_j$  indicates that  $\mathbf{f}$  belongs to the  $j$ th label or not (Figure 2). For example, if  $\varepsilon = 0.5$  and  $\rho = f_{\text{class}}(\mathbf{f}) = [0.72, 0.42, 0.82, 0.28, 0.86]$ , then  $\mathbf{y} = \varepsilon(\rho) = [1, 0, 1, 0, 1]$ . Therefore,  $\mathbf{f}$  is diagnosed as a simultaneous-fault (1, 3, 5). Notice that  $\mathbf{y} = [0, 0, 0, 0, 0]$  indicates that no fault has been found, and hence the unseen instance  $\mathbf{f}$  is diagnosed as a normal pattern.

**2.2. Extraction of Prior Domain Knowledge Features for Ignition Patterns.** When an engine starts firing, its secondary coil produces a rapid high voltage to cause spark plug to produce spark. This high voltage is called the *firing voltage*. Then the spark voltage decreases until zero. The spark voltage represents the voltage required to maintain spark for the duration of the spark line. The duration is called the *burn time*. After the burn time, the energy in the ignition coil nearly exhausts, and the residual energy forms slight oscillation in the ignition coil. The entire procedure is shown in Figure 3. Using the ignition pattern to diagnose the engine fault is a common diagnostic method for automotive engineers. With reference to some handbooks [2, 3], the following prior domain knowledge for a pattern can be observed for engine fault diagnosis (Figure 3):

- (1) firing voltage ( $F_1$ );
- (2) burn time ( $F_2$ );
- (3) average spark voltage ( $F_3$ ).

In this study, all patterns start from the firing voltage ( $F_1$ ) which is at the first sampling point:

$$F_1 = x_1, \quad (2)$$

where  $x_1$  is the voltage of the first sampling point. Ideally, the burn time ( $F_2$ ) starts from the spark voltage and ends at the position where the spark voltage falls to zero. However, in practice, the voltage could slightly oscillate after the burn time so that exact zero value may not be reached. In this study, when the voltage falls to 0.1% of the firing voltage, it is considered as zero and the burn time ends. The feature  $F_2$  can be obtained as illustrated in Figure 4, where  $a$  indicates the end

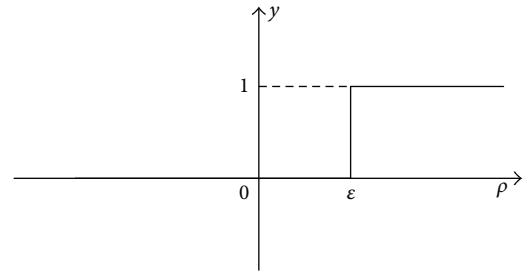


FIGURE 2: Decision function based on threshold  $\varepsilon$ .

point of burn time, and  $L_p$  is the length of patterns. With the index  $a$  and time step  $q$ , the average spark voltage of the spark line ( $F_3$ ) can be calculated as follows:

$$F_3 = \frac{1}{a} \sum_{q=1}^a x_q. \quad (3)$$

**2.3. Feature Extraction Using WPT and PCA and Combined Feature Vector.** WPT is a generalization of wavelet decomposition that offers a richer signal analysis [31]. It is well known that WPT can extract time-frequency features of a signal pattern. Given a set of patterns  $\mathbf{X} = \{\mathbf{x}_i\}$ ,  $i = 1$  to  $N_D$ , WPT transforms an ignition pattern  $\mathbf{x}_i \in R^s$  into a set of  $2^J$  coefficient packets  $v_{i,b} \in R^m$ , and  $m$  is the ceiling function of  $s/2^J$  at level  $J$  ( $b = 1$  to  $2^J$ ). Then, these packets  $v_{i,b}$  are concatenated as  $\mathbf{v}_i = [v_{i,1}, v_{i,2}, v_{i,3}, \dots, v_{i,2^J}]$  as the extracted features of the pattern  $\mathbf{x}_i$ . It is believed that the in-depth and hidden features of the single fault patterns can be detected through the coefficient packets  $\mathbf{v}_i$  after WPT decomposition. WPT is applied to every  $\mathbf{x}_i$  to form a set of features  $\mathbf{V} = \{\mathbf{v}_i\}$ ,  $i = 1$  to  $N_D$ .

Usually, the dimension of  $\mathbf{v}_i$  is large and a certain amount of the features may be redundant. Therefore, PCA is employed for dimension reduction of  $\mathbf{v}$  while retaining its important information. The details of PCA can be found in [22]. After applying PCA to  $\mathbf{V}$ , a set of eigen vectors  $\mathbf{h}_j$  and eigen values  $e_j$  are returned, which represent the transformation vectors

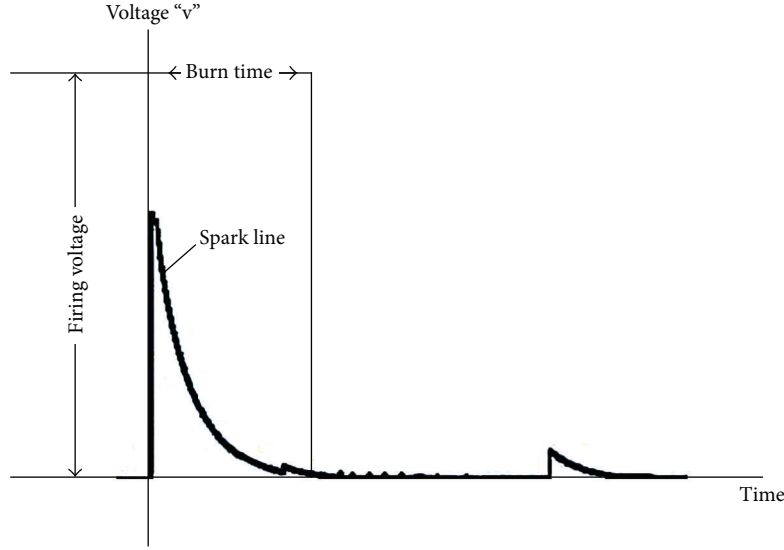


FIGURE 3: Key domain knowledge features of the normal engine ignition signal.

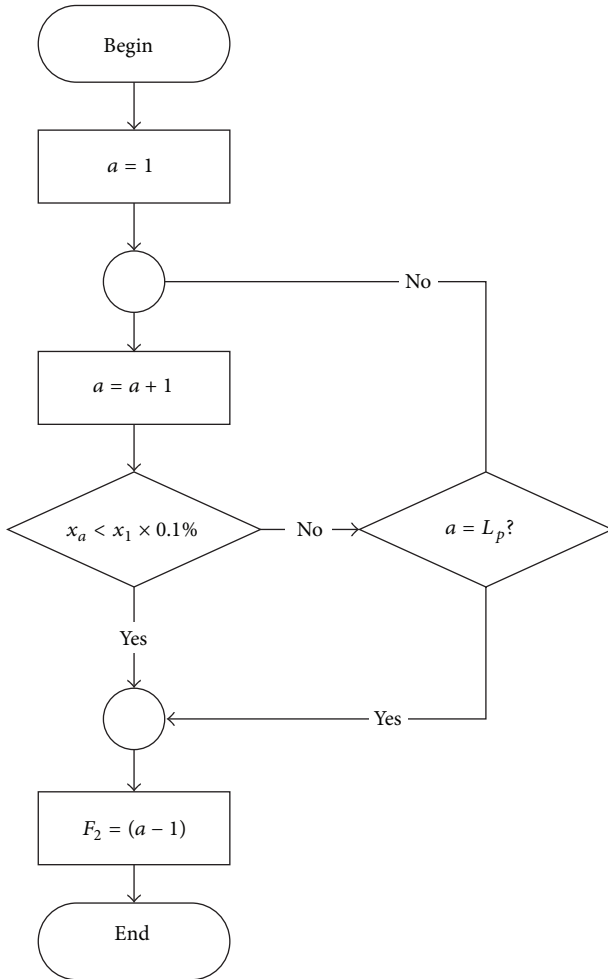


FIGURE 4: Procedure for calculating the burn time ( $F_2$ ).

and the importance of the transformed dimensions, where  $j = 1$  to  $p$ ,  $e_1 \geq e_2 \geq \dots \geq e_p \geq 0$ . The  $k$  ( $k < p$ ) most important dimensions are selected based on the criterion of  $\sum_{j=1}^k e'_j \leq 0.99$ , that is, a 1% of information loss is allowed, where  $e'_j = e_j / \sum_{j=1}^p e_j$  is a normalized eigen value. Knowing the value of  $k$ , the corresponding transformation matrix  $\mathbf{H} = [\mathbf{h}_1 \mathbf{h}_2 \dots \mathbf{h}_k]$  is then formed. So  $\mathbf{F}' = \mathbf{H}^T \mathbf{V}$  is the reduced feature dataset. For any unseen ignition pattern  $\mathbf{x}$  in the future, its feature vector can be obtained by  $\mathbf{f}' = \mathbf{H}^T \mathbf{v}$ , where  $\mathbf{v} = \text{WPT}(\mathbf{x})$ . By combining the prior domain knowledge, the final feature vector as the classifier inputs is given in the following:

$$\mathbf{f} = [\mathbf{f}', F_1, F_2, F_3]. \quad (4)$$

**2.4. Relevance Vector Machine.** Relevance vector machine [27] is a statistical learning method utilizing Bayesian learning framework and popular kernel methods. In fault diagnosis, RVM is designed to predict the posterior probability of the binary class membership (i.e., either positive or negative) for an unseen input  $\mathbf{f}$ , 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  $\sigma(z(\mathbf{f})) = 1/(1 + \exp(-z(\mathbf{f})))$  to the predicted decision  $z(\mathbf{f})$  and adopting the Bernoulli distribution for  $P(\mathbf{t} | \mathbf{F})$ . The likelihood of the data is written as follows [27]:

$$P(\mathbf{t} | \mathbf{F}, \mathbf{w}) = \prod_{n=1}^N \sigma\{z(\mathbf{f}_n; \mathbf{w})^{t_n}\} [1 - \sigma\{z(\mathbf{f}_n; \mathbf{w})\}]^{1-t_n} \quad (5)$$

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

where  $\mathbf{w} = (w_0, w_2, \dots, w_N)^T$  are the adjustable parameters, and a radial basis function (RBF) is typically chosen for  $K(\cdot)$ .

The current objective is to find the optimal weight vector  $\mathbf{w}$  in (5) for the given dataset  $\mathbf{F}$ , which is equivalent to find  $\mathbf{w}$  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, it is impossible to determine the weights analytically. Hence, closed-form expressions for either the marginal likelihood  $P(\mathbf{w} | \boldsymbol{\alpha})$  or equivalently the weight posterior  $P(\mathbf{w} | \mathbf{t}, \mathbf{F}, \boldsymbol{\alpha})$  are denied. Thus, the following approximation procedure is chosen [32], which is based on Laplace's method.

- (a) For the current fixed values of  $\boldsymbol{\alpha}$ , the most probable weights  $\mathbf{w}_{\text{MP}}$  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:

$$\begin{aligned} \mathbf{w}_{\text{MP}} &= \underset{\mathbf{w}}{\operatorname{argmax}} \log P(\mathbf{w} | \mathbf{t}, \mathbf{F}, \boldsymbol{\alpha}) \\ &= \underset{\mathbf{w}}{\operatorname{argmax}} \log \{P(\mathbf{t} | \mathbf{F}, \mathbf{w}) P(\mathbf{w} | \boldsymbol{\alpha})\} \\ &= \underset{\mathbf{w}}{\operatorname{argmax}} \left\{ \sum_{n=1}^N [t_n \log s_n \right. \\ &\quad \left. + (1 - t_n)(1 - \log s_n)] \right. \\ &\quad \left. - \frac{1}{2} \mathbf{w}^T \mathbf{A} \mathbf{w} \right\} \end{aligned} \quad (6)$$

$$\text{with } s_n = \sigma \{z(\mathbf{f}_n; \mathbf{w})\},$$

$$\mathbf{A} = \operatorname{diag}(\alpha_0, \alpha_1, \dots, \alpha_N).$$

- (b) Laplace's method is simply a Gaussian approximation to the log-posterior around the mode of the weights  $\mathbf{w}_{\text{MP}}$ . Equation (6) is differentiated twice to give

$$\nabla_{\mathbf{w}} \nabla_{\mathbf{w}} \log P(\mathbf{w} | \mathbf{t}, \mathbf{F}, \boldsymbol{\alpha}) \Big|_{\mathbf{w}_{\text{MP}}} = -(\boldsymbol{\Phi}^T \mathbf{B} \boldsymbol{\Phi} + \mathbf{A}), \quad (7)$$

where  $\mathbf{B} = \operatorname{diag}(\beta_1, \dots, \beta_N)$  is a diagonal matrix with  $\beta_n = \sigma\{z(\mathbf{f}_n; \mathbf{w})\} [1 - \sigma\{z(\mathbf{f}_n; \mathbf{w})\}]$ , and  $\boldsymbol{\Phi}$  is a  $N \times (N+1)$  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 (7), the covariance matrix  $\boldsymbol{\Sigma} = -(\nabla_{\mathbf{w}} \nabla_{\mathbf{w}} \log P(\mathbf{w} | \mathbf{t}, \mathbf{F}, \boldsymbol{\alpha})|_{\mathbf{w}_{\text{MP}}})^{-1} = (\boldsymbol{\Phi}^T \mathbf{B} \boldsymbol{\Phi} + \mathbf{A})^{-1}$  can be obtained.

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

$$\alpha_i^{\text{new}} = \frac{\gamma_i}{\mathbf{u}_i^2}, \quad (8)$$

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

**2.5. Pairwise Coupled RVM.** The traditional RVM formulation is designed only for binary classification; that is, the output is either positive (+1) or negative (−1). In order to resolve the current simultaneous-fault problem, multiclass strategies of one-versus-all (1vA) and one-versus-one (1v1, or specifically named as pairwise coupling) [28] can be employed. Traditionally 1vA strategy constructs a group of classifiers  $f_{\text{class}} = [C_1, \dots, C_d]$  in a  $d$ -label classification problem. For any unknown input  $\mathbf{f}$ , the classification vector  $\mathbf{y} = [y_1, y_2, \dots, y_d]$ , where  $y_i = 1$  if  $C_i(\mathbf{f}) = +1$  or  $y_i = 0$  if  $C_i(\mathbf{f}) = -1$ . The 1vA strategy is simple and easy to implement. However, it generally gives a poor result [29, 33, 34] since 1vA does not consider the pairwise correlation and hence induces a much larger indecisive region than 1v1 as shown in Figure 5.

On the other hand, pairwise coupling (1v1) also constructs a group of classifiers  $f_{\text{class}} = [C_1, \dots, C_d]$  in a  $d$ -label classification problem. However, each  $C_i = [C_{i1}, \dots, C_{ij}, \dots, C_{id}]$  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  $d(d - 1)/2$  pairwise classifiers in  $f_{\text{class}}$  (Figure 6(b)).

In this study, each  $C_{ij}$  can be an RVM classifier which estimates the pairwise probability that an unknown instance  $\mathbf{f}$  belongs to the  $i$ th label against the  $j$ th label, that is,  $C_{ij}(\mathbf{f}) = P(l_i | \mathbf{f}, l_i \text{ or } l_j)$ . There are several methods for pairwise coupling strategy [28], which are, however, suitable for multiclass diagnosis only because of the constraint  $\sum \rho_i = 1$ . Note that the nature of simultaneous-fault diagnosis is that  $\sum \rho_i$  is not necessarily equal to 1. Therefore, the following simple pairwise coupling strategy for simultaneous-fault diagnosis is proposed.

Every  $C_{ij}$  is trained only by the training data with the  $i$ th and  $j$ th labels. Let  $\rho_{ij} = C_{ij}(\mathbf{f}) = P(l_i | \mathbf{f}, l_i \text{ or } l_j)$  be the pairwise probability of the  $i$ th label against the  $j$ th label for an unknown instance  $\mathbf{f}$ , where  $C_{ij}(\mathbf{f})$  is estimated using RVM. Then,  $\rho_i$  is calculated as

$$\begin{aligned} \rho_i = C_i(\mathbf{f}) &= \frac{\sum_{j=1: j \neq i}^d n_{ij} C_{ij}(\mathbf{f})}{\sum_{j=1: j \neq i}^d n_{ij}} \\ &= \frac{\sum_{j=1: j \neq i}^d n_{ij} \rho_{ij}}{\sum_{j=1: j \neq i}^d n_{ij}}, \end{aligned} \quad (9)$$

where  $n_{ij}$  is the number of training data with the  $i$ th and  $j$ th labels. Hence, the probability  $\rho_i$  can be more accurately estimated from  $\rho_{ij} = C_{ij}(\mathbf{f})$  because the pairwise correlation between the labels are taken into account. With the above pairwise coupling strategy, PCRVM can more accurately estimate the probability vector  $\boldsymbol{\rho}$  and hence generate a higher classification accuracy for simultaneous-fault diagnosis.

**2.6. Decision Threshold Optimization and F-Measure.** PCRVM can only provide the probability vector

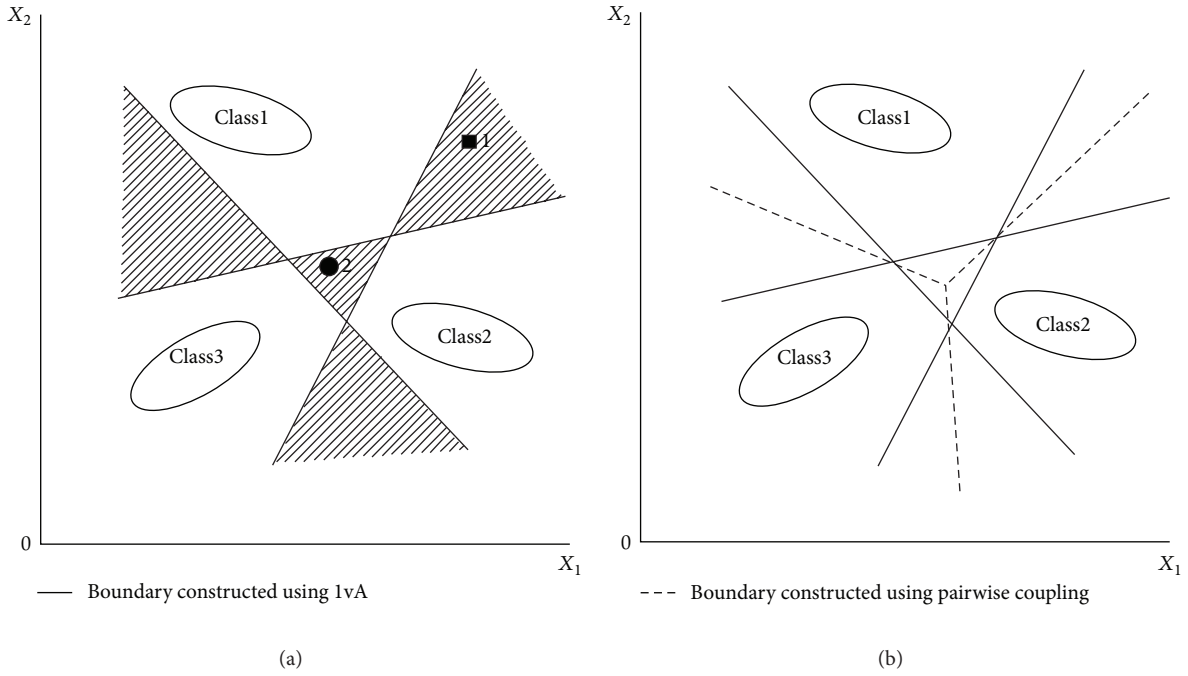


FIGURE 5: Indecisive regions (shaded area) using 1vA (a) and pairwise coupling (1v1) (b) [29].

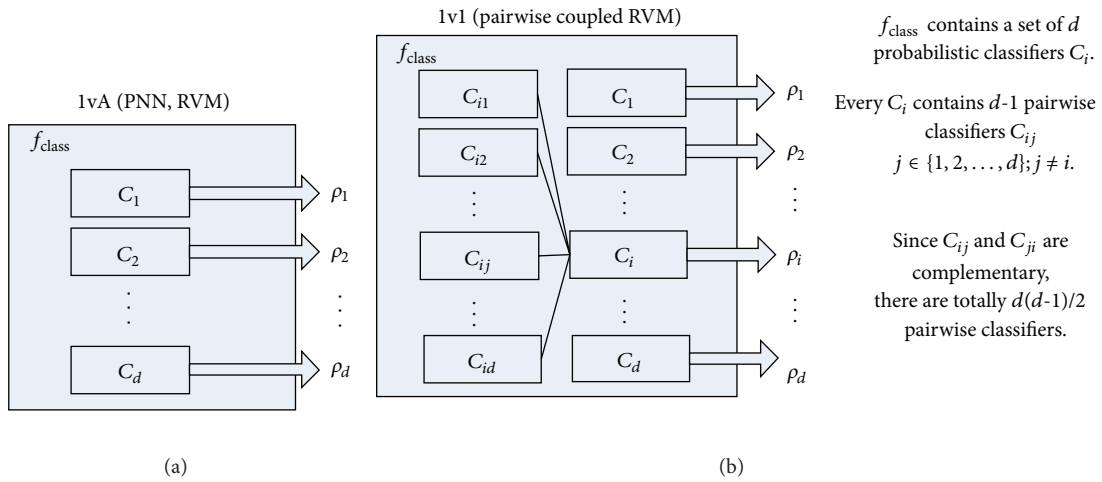


FIGURE 6: Architecture of 1vA classifier (a) and 1v1 classifier (b).

$\rho = [\rho_1, \rho_2, \dots, \rho_d]$  of the single-fault labels but the desired result is the classification vector  $\mathbf{y} = [y_1, y_2, \dots, y_d] = [\varepsilon(\rho_1), \varepsilon(\rho_2), \dots, \varepsilon(\rho_d)]$ . It is obvious that the value of decision threshold  $\varepsilon$  will greatly affect the classification accuracy. For a situation without any prior information, the best estimate of  $\varepsilon$  may be simply set to 0.5, that is, the presence of a fault is considered if its probability is at least 0.5. However, the value of  $\varepsilon$  should be optimized according to the classification accuracy. In other words, the value  $\varepsilon$  should be chosen to produce the highest classification accuracy over a validation dataset.

Besides, the traditional evaluation of classification accuracy only considers exact matching of the predicted label vector  $\mathbf{y}$  against the true label vector  $\mathbf{l}$ . This evaluation is

however not suitable for simultaneous-fault diagnosis where partial matching is preferred. Therefore, a common evaluation called  $F$ -measure is employed.

$F$ -measure [35] is commonly used as performance evaluation for information retrieval systems where a document may belong to a single or multiple tags simultaneously. This is very similar to the current application that contains a mixture of single-fault and simultaneous-fault patterns. With  $F$ -measure, the evaluation of single-fault and simultaneous-fault test patterns can be appropriately done at one time. To define  $F$ -measure  $F_{\text{me}}$ , two concepts of precision ( $\pi$ ) and recall ( $\tau$ ) are used so that

$$F_m = \frac{2\pi\tau}{\pi + \tau}, \quad (10)$$

where  $\pi$  and  $\tau$  are originally designed for single-fault patterns only but can be extended to handle simultaneous-fault patterns. For  $N_T$  single-fault and simultaneous-fault test data,

$$\pi = \frac{\sum_{i=1}^d \sum_{j=1}^{N_T} y_i^j l_i^j}{\sum_{i=1}^d \sum_{j=1}^{N_T} y_i^j}, \quad \tau = \frac{\sum_{i=1}^d \sum_{j=1}^{N_T} y_i^j l_i^j}{\sum_{i=1}^d \sum_{j=1}^{N_T} l_i^j}, \quad (11)$$

where  $y_i^j$  and  $l_i^j$  are, respectively, the  $i$ th predicted label and the  $i$ th true label in the  $j$ th test data,  $y_i^j$  and  $l_i^j \in \{0, 1\}$ . Substituting (11) into (10), the final  $F$ -measure equation is given in (12). The larger the  $F$ -measure value, the higher the diagnostic accuracy is

$$F_m = \frac{2 \times \sum_{i=1}^d \sum_{j=1}^{N_T} y_i^j l_i^j}{\sum_{i=1}^d \sum_{j=1}^{N_T} y_i^j + \sum_{i=1}^d \sum_{j=1}^{N_T} l_i^j} \in [0, 1]. \quad (12)$$

With  $F$ -measure, the value  $\varepsilon$  can be optimized using typical direct search techniques such as *Genetic Algorithms* (GA) and *Particle Swarm Optimization* (PSO) [36].

**2.7. Principle of Detection of Single Faults and Simultaneous Faults.** After an unknown instance  $\mathbf{f}$  is passed to the above system, a probability vector  $\boldsymbol{\rho}$  is produced. If  $\mathbf{f}$  is caused by a single fault (e.g., the  $j$ th fault),  $\mathbf{f}$  contains only the symptoms of the  $j$ th fault. Then, in  $\boldsymbol{\rho}$ , the corresponding probability  $\rho_j \geq \varepsilon^*$  so that  $y_j = 1$  in the decision vector  $\mathbf{y}$  while all other  $y_k = 0$ ,  $k \neq j$ . In other words,  $\sum y_j = 1$  and hence a single fault is detected.

For the case that  $\mathbf{f}$  is caused by two simultaneous faults (e.g., the  $j$ th and  $k$ th faults),  $\mathbf{f}$  is constituted by the symptoms of the  $j$ th and  $k$ th faults. These symptoms may be overlapping or interdistorted. In the current diagnostic system, probabilities are employed to give the similarity of  $f$  against the  $j$ th and  $k$ th faults by  $C_j$  and  $C_k$ , respectively. If their symptoms are not highly overlapping or interdistorted, there is a high chance that the corresponding probabilities  $\rho_j, \rho_k \geq \varepsilon^*$ . Under this circumstance,  $y_j = 1$  and  $y_k = 1$ , making  $\sum y_j \geq 1$  so that a simultaneous fault can be detected. The mechanism is similar for three or more simultaneous faults. By combining these cases, the proposed system can diagnose both single fault and simultaneous faults using classifiers trained with single faults only.

**2.8. Summary of Proposed Framework and Techniques.** The previous framework and techniques are summarized in Algorithm 1. Figure 7(a) shows the workflow of using DK and WPT+PCA as feature extraction. Every dataset for training, validation and test requires going through the step of feature extraction. Figure 7(b) shows the construction of the classifier  $f_{\text{class}}$ . The classifier has the architecture of pairwise coupling as depicted in Figure 6(b). Then the classifier is passed to an optimizer to search for the optimal decision threshold based on a validation set  $VALID\_F$  and  $F$ -measure  $F_{\text{me}}$  as shown in Figure 7(c), where  $f_{\text{class}}$  outputs the probability vector  $\boldsymbol{\rho} = [\rho_1, \rho_2, \dots, \rho_d]$  for each case in  $VALID\_F$ . To optimize the threshold, the  $F$ -measure  $F_{\text{me}}$  over  $VALID\_F$  can be

evaluated as the fitness value. Since the direct search technique is easily stuck by local minima, it is necessary to run  $M$  different times of the optimization step in Figure 7(c) to avoid this issue. For testing and running, the step in Figure 7(d) is very similar to Figure 7(c) except the optimal threshold  $\varepsilon_{\text{opt}}$  that has been determined. The choice of parameters of the feature extraction, classification, and direct search techniques are discussed in Section 4.

### 3. Experimental Setup

To verify the effectiveness of the proposed methodology, an experiment was set up for sample data acquisition and evaluation tests. The details of the experimental setup and preparation of datasets are presented in the following subsections.

**3.1. Data Sampling.** In total, a set of single faults and simultaneous faults were imitated and selected as demonstration examples. There are 10 kinds of single faults as described in Tables 1 and 4 kinds of simultaneous faults as described in Table 2. However, there is an issue that the simultaneous-fault patterns are not caused by a random combination of single faults but some reasonable combinations (e.g., it is impossible to have wide spark-plug gap and narrow spark-plug gap at the same time). Moreover, the experimental data show that a simultaneous-fault ignition pattern is caused by a combination of at most three single faults. Beyond these constraints, the ignition patterns cannot be captured due to engine stall. Some sample ignition patterns of these single faults and reasonable simultaneous faults are shown in Figures 8 and 9, respectively.

In this study, five well-known inline 4-cylinder electronic ignition engines, namely, HONDA B18C, HONDA D15B, HONDA K20A, TOYOTA 2NZ-FE, and MITSUBISHI 4G15, were employed as the experimental platforms, and a computer-linked automotive scope meter was used (Figure 10) to capture raw ignition patterns. Different models of engines were used for training in order to enhance the generalization of the classifier. To capture ignition patterns, the sampling frequency of the scope meter was set to a high rate of 100 kHz, that is, 100,000 sampling points per second. Under the software provided by the scope meter, ignition patterns were recorded in a PC and converted into a file of excel format for processing and analysis.

For each case (single fault or simultaneous faults in Tables 1 and 2) in every test engine, sixteen ignition patterns (four patterns for each cylinder) were captured over two different engine testing conditions according to the standard procedure in [3] (1200 rpm and 2000 rpm). As the pattern obtained in each cylinder per engine cycle is somewhat unrepeatable, four patterns per cylinder are required. The reason for causing unrepeatable patterns is that a constant engine speed is difficult to hold during sampling. Furthermore, each cylinder has its own manufacturing error, different inlet and exhaust flow characteristics, and so forth. Finally there were 1600 ignition patterns of single faults (i.e., 10 labels  $\times$  4 patterns  $\times$  4 cylinders  $\times$  2 testing conditions  $\times$  5 engines) and 800 ignition



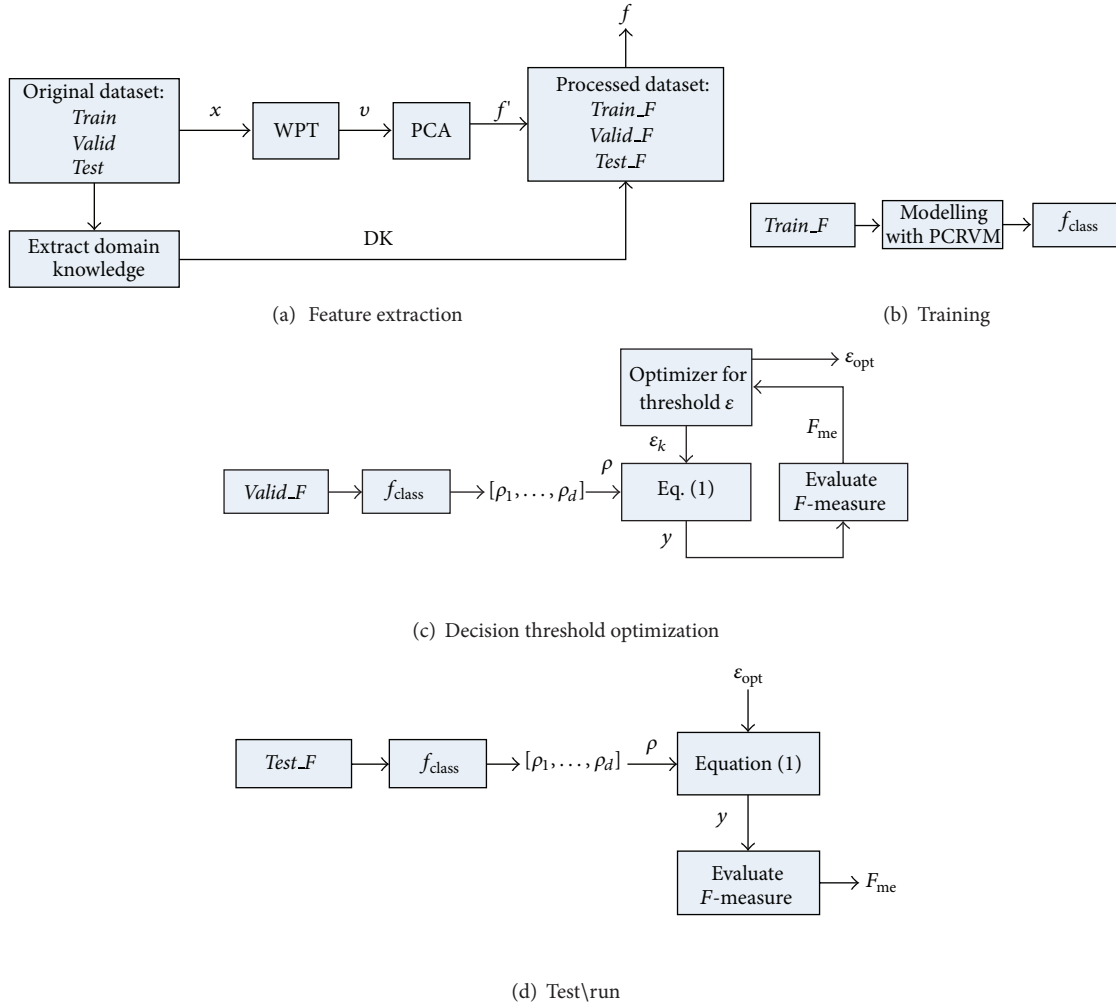


FIGURE 7: Workflow of feature extraction, decision threshold optimization, diagnostic system training, and testing.

TABLE 1: Sample single faults of engine trouble reflected by ignition patterns.

| Case number | Symptom or possible cause             |
|-------------|---------------------------------------|
| 1           | Retarded ignition timing              |
| 2           | High resistance in secondary circuit  |
| 3           | Partially broken spark-plug cable     |
| 4           | Defective spark plug                  |
| 5           | Narrow spark-plug gap                 |
| 6           | Misfire due to extremely lean mixture |
| 7           | Carbon fouled in spark plug           |
| 8           | Engine knock                          |
| 9           | Rich mixture                          |
| 10          | Wide spark-plug gap                   |

TABLE 2: Sample possible simultaneous faults of engine trouble reflected by ignition patterns.

| Case number | Symptom or possible cause   |
|-------------|---|
| 1           | High resistance in secondary circuit and misfire due to extremely lean mixture        |
| 2           | Narrow spark-plug gap and carbon fouled in spark plug                                 |
| 3           | Partially broken spark-plug cable and wide spark-plug gap                             |
| 4           | High resistance in secondary circuit cable and narrow spark-plug gap and rich mixture |
| 5           | Partially broken spark-plug cable and engine knock and wide spark-plug gap            |

patterns of simultaneous faults (i.e., 5 labels  $\times$  4 patterns  $\times$  4 cylinders  $\times$  2 testing conditions  $\times$  5 engines).

3.2. *Data Normalization.* As the number of sampling points of every captured pattern is not exactly the same due to engine

speed fluctuation and various testing conditions, all patterns were normalized within the same range in order to match the number of inputs of the classifier. Normalization of the ignition patterns was done in terms of duration. In this study, the number of sampling points for every pattern was less than 17,000. For the sake of conservation, a standard number of

Given a training dataset  $TRAIN\_F$  of single-fault patterns only, a validation dataset  $VALID\_F$  and a test dataset  $TEST\_F$  of single-fault and simultaneous-fault patterns (all datasets have been preprocessed by the combination of DK and WPT and PCA, as presented in Figure 7(a)):

- (i) Train the probabilistic classifier  $f_{class}$   
 $f_{class}$  includes  $d(d-1)/2$  pairwise classifiers  $C_{ij}$  as shown in Figure 6(b).
- (ii)  $\forall \mathbf{f} \in VALID\_F$ , prepare the probability vector  $\boldsymbol{\rho}$   
 Calculate  $f_{class}(\mathbf{f}) = \boldsymbol{\rho} = [\rho_1, \rho_2, \dots, \rho_d]$  (Figure 7(c)).
- (iii) For  $k = 1$  to  $M$ // Run a direct search technique, such as GA or PSO,  $M$  times  
 Produce an initial population  $\epsilon$  for the decision threshold  $\epsilon$ 
  - (a)  $\forall \epsilon \in \epsilon$ , find the classification vector  $\mathbf{y}(\mathbf{f}) = \mathbf{y} = [y_1, y_2, \dots, y_d] = [\epsilon(\rho_1), \epsilon(\rho_2), \dots, \epsilon(\rho_d)]$  according to (1).
  - (b) Calculate the  $F$ -measure  $F_{me}$  with  $\mathbf{y}(\mathbf{f})$  and  $\mathbf{l}(\mathbf{f})$  using (12), that is, find  $F_{me}$  over  $VALID\_F$ , where  $\mathbf{l}(\mathbf{f}) = [l_1, l_2, \dots, l_d]$  is the true classification vector for input  $\mathbf{f}$  provided from  $VALID\_F$ .
  - (c) Produce next generation of  $\epsilon$   
 Until convergence or matching stopping criteria, return the best solution  $\epsilon$  as the  $\epsilon_k$ .
- (iv) Among all  $\epsilon_k, k = 1$  to  $M$ , choose the one producing the highest  $F$ -measure  $F_{me}$  as the optimal decision threshold  $\epsilon_{opt}$ .
- (v) Return the trained probabilistic classifier  $f_{class}$  and the optimized decision threshold  $\epsilon_{opt}$  as the main components of the intelligent diagnostic system.
- (vi) The performance of  $f_{class}$  and  $\epsilon_{opt}$  can be evaluated with  $TEST\_F$  and  $F_{me}$  as illustrated in Figure 7(d).

ALGORITHM 1: Algorithm of the proposed framework for simultaneous-fault diagnosis of time-dependent ignition patterns.

TABLE 3: Evaluation of different combinations of techniques.

| $F$ -measure over $TEST$ with<br>$\epsilon = 0.5$ | Classification technique |         |                |
|---|--------------------------|---------|----------------|
|   | PNN                      | RVM     | PCRVM          |
| Feature extraction                                |                          |         |                |
| None  | 0.73214                  | 0.75063 | 0.80442        |
| DK  | 0.73581                  | 0.76023 | 0.81202        |
| FFT   | 0.78362                  | 0.78162 | 0.82162        |
| WPT + PCA<br>(Haar, level $J = 9$ )               | 0.78221                  | 0.79171 | 0.82178        |
| DK + FFT  | 0.78333                  | 0.80162 | 0.82242        |
| DK + WPT + PCA<br>(Haar, level $J = 9$ )          | 0.78548                  | 0.81245 | <b>0.84225</b> |

TABLE 4: GA operators and parameters.

|                      |   |
|----------------------|---|
| Number of generation | 1000  |
| Population size      | 50  |
| Selection method     | Standard proportional selection   |
| Crossover method     | Simple crossover with probability = 80%   |
| Mutation method      | Hybrid static Gaussian and uniform mutation with probability = 40% and standard deviation = 0.2 |

sampling points for all patterns was set to 18,000 in order not to lose any exceptional information. To standardize the duration of all patterns, steady-state values can be appended to the rear part of the patterns if necessary. Normally, the steady-state value for the ignition pattern is equal to zero (0 V). For those patterns having fewer than 18,000 data points, zeros can be appended. Therefore, the durations of all sample patterns were normalized before feature extraction using WPT + PCA.

3.3. Allocation of Datasets. In order to test the diagnostic performance for both single faults and simultaneous faults,

about 3/4 of the single-fault patterns were taken as training data  $TRAIN$ . There were 1/16 of the single-fault patterns and 1/5 simultaneous-fault patterns in the validation dataset  $VALID$ , while the remaining 3/16 of the single-fault patterns and 4/5 simultaneous-fault patterns were used as test dataset  $TEST$ .

## 4. Experimental Results

To select the best combination of the techniques for feature extraction, classification, and threshold optimization, many experiments based on the sample dataset were conducted. The sample dataset was separated into 3 groups:  $TRAIN$  for training the classifier,  $VALID$  for the threshold optimization and selection of direct search techniques, and  $TEST$  for evaluating the performance of different combinations of the feature extraction, classification, and threshold optimization techniques. The performance evaluation over  $TEST$  is based on  $F$ -measure that can evaluate single-fault and simultaneous-fault patterns at one time according to partial matching criterion. All experiments were carried out under a PC with Core i5 @ 3.20GHz and 4GB RAM. All the proposed techniques mentioned were implemented using Matlab R2008a.

4.1. Results of Various Combinations of Feature Extraction and Classification Techniques. The reasonable combinations of DK, FFT, and WPT+PCA for feature extraction were tested as shown in Table 3 along with the corresponding evaluation. The classification techniques used in the experiment include PNN, RVM, and PCRVM. PNN [24] was selected for comparison because it is a traditional probabilistic classification using radial basis (Gaussian) kernel. The input dimension  $s$  of the classifiers for evaluation is subject to the feature extraction technique. In terms of WPT, PCA, and DK, WPT transforms the original patterns of 18000 points into different packets at level  $J$ . The value of  $J$  can be determined using

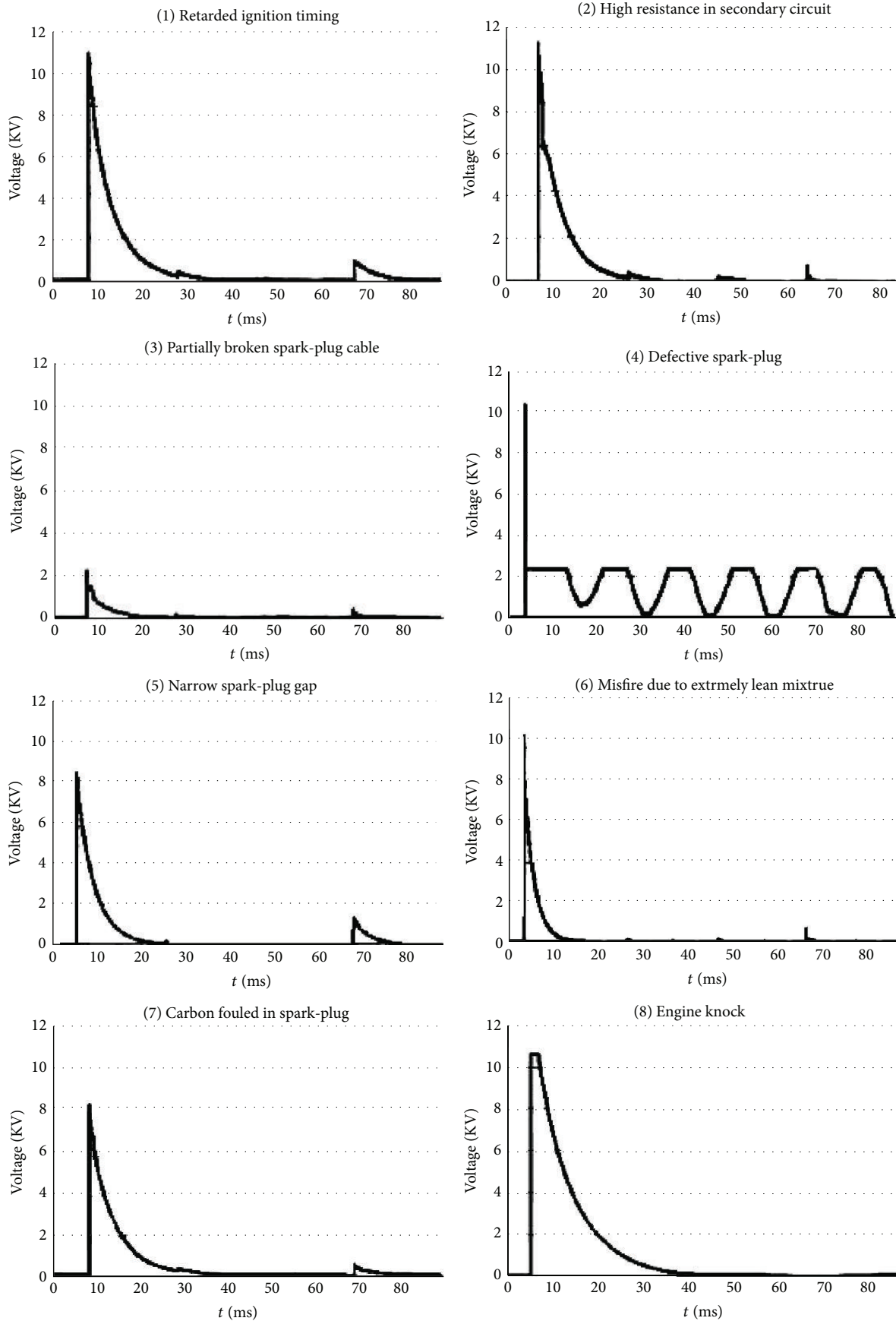


FIGURE 8: Continued.

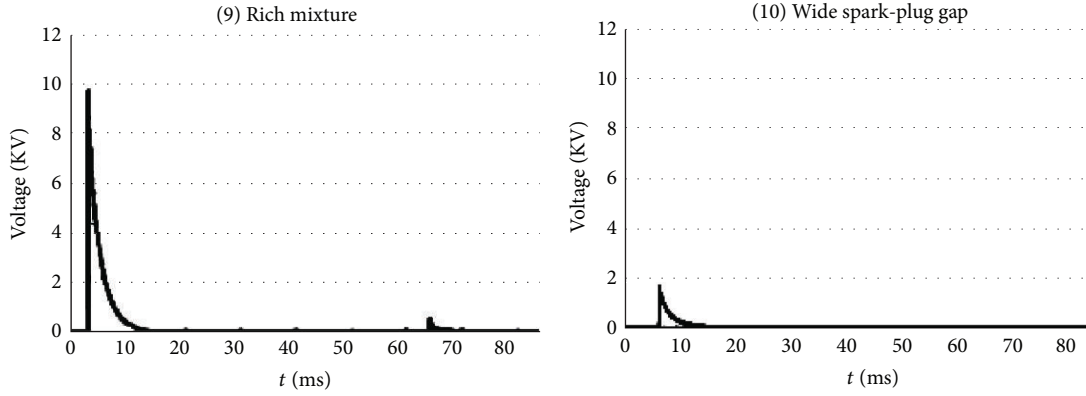


FIGURE 8: Sample ignition patterns and their corresponding single engine faults.

entropy information. A built-in function *bestlev* (meaning best level) is available in Matlab wavelet toolbox for this purpose. After carrying out many experiments using the function *bestlev*,  $J$  was tested to be 9 for the sample dataset of ignition patterns. In this study, the common mother wavelet, *Haar wavelet*, was selected for the purposes of illustration and comparison of different feature extraction techniques. For better performance, different types of mother wavelets could be evaluated in the future. After PCA, the most 22 important dimensions were selected as described in Section 2.3. Therefore, the size of  $s$  is equal to 22 plus the three domain features, that is, 25 totally. For FFT and DK, the sizes of  $s$  are equal to 18000 and 3 features, respectively.

In the construction of the intelligent engine diagnostic systems with different techniques for comparison, each feature extraction technique was firstly employed to preprocess the training dataset *TRAIN*, and then different classification techniques were applied. The performance of every combination was evaluated over *TEST* using  $F$ -measure. In order to reflect the effectiveness of the feature extraction, the classification techniques under *TRAIN* without any preprocessing were also examined. Therefore, there were totally 18 combinations of feature extraction and classification techniques as shown in Table 3.

For classification techniques of PNN, RVM, and PCRVM, several simple settings are necessary. PNN requires a hyperparameter called *smoothing factor* or *spread*, which is equivalent to the width of the *Gaussian* kernel within PNN. If the value of spread is set too high, the trained classifier may easily overfit the training patterns and hence a lower generalization. In the case study, the value of spread for PNN was simply set to be 0.2 according to rule of thumb [37]. The RVM and PCRVM employ different classification strategies (1vA versus 1v1) but they share the same set of hyperparameters, namely, type of kernel functions and the corresponding kernel parameters. For illustration purpose, *Gaussian* was selected as the kernel function  $K(\cdot)$  and its kernel width was set to be 1.0 in order to calculate the design matrix  $\Phi$  in (7). The experimental results of various combinations of feature extraction and classification techniques are shown in Table 3. In order to evaluate the  $F$ -measures under different combinations of preprocessing and classifications, the decision threshold was

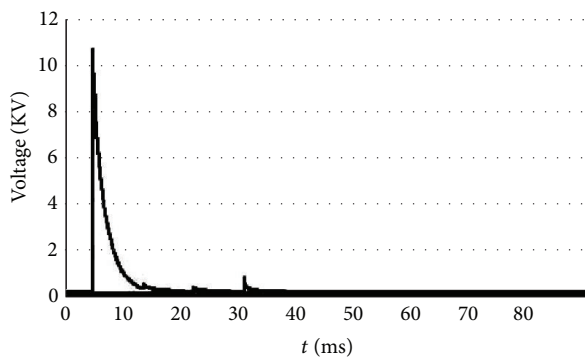
simply set to 0.5 for a simple and fair comparison in this phase.

4.2. *Results with Threshold Optimization.* Genetic Algorithms (GA) are the most classical direct search technique, while *Particle Swarm Optimization* (PSO) is another popular choice. Both of them were tested for the optimization of the decision threshold and they share the same objective function. Since the  $F$ -measure  $F_{me} \in (0, 1)$ , the objective function of optimization can be simply set as follows:

$$\min(1 - F_{me}). \quad (13)$$

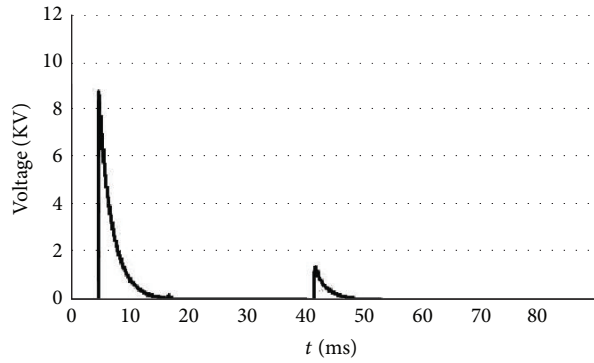
The higher the  $F_{me}$ , the better the optimization result will be. The optimization procedure follows the proposed algorithm in Algorithm 1, where the number of runs  $M$  was set to be 20. Tables 4 and 5 show the detail settings of the GA and PSO operators and parameters, respectively, according to the literature [36]. Therefore, among 20 runs of the proposed algorithm for every combination of feature extraction and classification techniques under GA and PSO optimization, the optimized threshold  $\epsilon_{opt}$  and its corresponding  $F_{me}$  value of different combinations of techniques are shown in Tables 6 and 7, respectively.

4.3. *Individual Result of Single- and Simultaneous-Fault Diagnosis.* The objective of this research is to train a probabilistic classifier using single-fault patterns and then predict both single and simultaneous faults. However, it is unclear whether the performance of the trained probabilistic classifier on simultaneous faults in Section 4.2 is correct or not because the classification results of different combinations of techniques were all evaluated over the whole test dataset *TEST*, which contains single-fault and simultaneous-fault patterns. To better illustrate the performance of the proposed method, *TEST* was further separated into two groups, one for purely single-faults  $TEST_1$ , and another for purely simultaneous-faults  $TEST_2$ . All evaluation tests were done using the combination of DK+WPT+PCA as feature extraction and the PSO-optimized threshold of 0.7147 because Tables 6 and 7 show that this combination produces the best  $F$ -measure. The  $F$ -measures of purely single faults and purely simultaneous



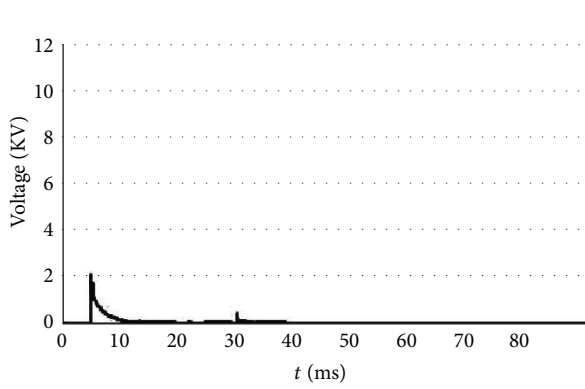
Simultaneous fault 1  
 (2) High resistance in spark-plug cable +  
 (6) Misfire due to extremely lean mixture

(a)



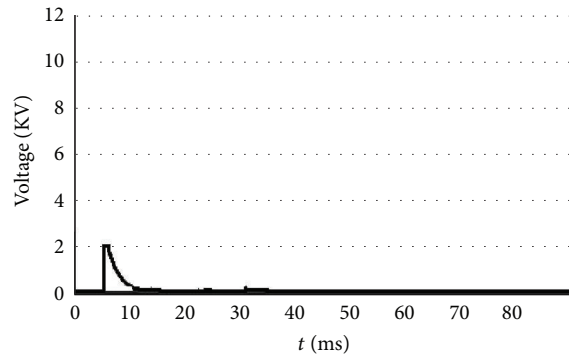
Simultaneous fault 2  
 (5) Narrow spark-plug gap +  
 (7) Carbon fouled in spark plug

(b)



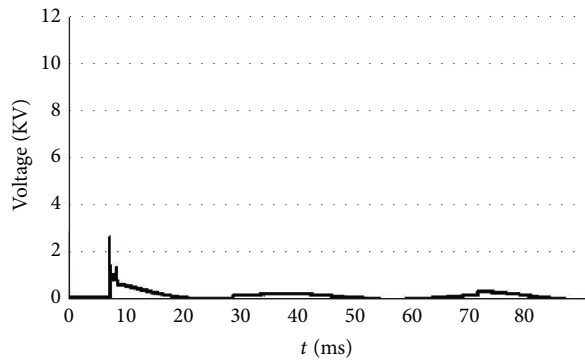
Simultaneous fault 3  
 (3) Partially broken spark-plug cable +  
 (10) Wide spark-plug gap

(c)



Simultaneous fault 4  
 (2) High resistance in spark-plug cable +  
 (5) Narrow spark-plug gap +  
 (9) Rich mixture

(d)



Simultaneous fault 5  
 (3) Partially broken spark-plug cable +  
 (8) Engine knock +  
 (10) Wide spark-plug gap

(e)

FIGURE 9: Sample ignition patterns and their corresponding simultaneous engine faults.



FIGURE 10: Collection of ignition patterns from a test engine using a computer-linked automotive scope meter.

TABLE 5: PSO operators and parameters.

|                      |      |
|----------------------|------|
| Number of generation | 1000 |
| Population size      | 50   |
| $w_c$                | 0.9  |
| $c_1$                | 2    |
| $c_2$                | 2    |

faults are shown in Tables 8 and 9, respectively, which were calculated using (12) with the related faults. For example, for Fault 1,  $F_{me}$  is evaluated on the test cases of Fault 1 only. For simultaneous faults of the combination (5, 7), after prediction there is a classification vector  $\mathbf{y} = [y_1, \dots, y_{10}]$ , and a true vector  $\mathbf{l} = [l_1, \dots, l_{10}]$ , then  $y_5$  and  $y_7$  with the true values  $l_5$  and  $l_7$  from the test cases are employed to compute the two separate  $F_{me}$  values for detail analysis.

**4.4. Results Comparison with the Latest Technique.** To further verify the effectiveness of the presented framework, the existing binarization approach using SVM [23] was applied to the ignition system diagnosis for comparison. The binarization approach builds classifiers directly based on raw ignition patterns, so there is no feature extraction step. In this approach, a number of binary classifiers  $Bf_j(\cdot)$  were constructed, respectively, using support vector machines (SVM) with one-versus-all splitting strategy where  $j = 1$  to  $d$ ,  $d$  is the number of single faults again. A decision vector  $\mathbf{y} = [\theta(Bf_1(\mathbf{x})) \cdots \theta(Bf_d(\mathbf{x}))]$  can be obtained for an unknown pattern  $\mathbf{x}$ , where  $Bf_j(\mathbf{x}) \in R$  is the raw output value of the  $j$ th SVM classifier, and  $\theta(Bf_j(\mathbf{x})) = 1$  if  $Bf_j(\mathbf{x}) \geq 0$  and  $\theta(Bf_j(\mathbf{x})) = 0$  Otherwise. From this framework, only single-fault patterns were used for training the binary classifiers while simultaneous-fault patterns are also not necessary. Since there is no probabilistic output but only a binary decision vector is generated in the binarization approach, no decision threshold optimization is necessary in this experiment. The results using the binarization approach is shown in Table 10.

## 5. Discussion of Results

**5.1. Effect of Feature Extraction and Pairwise Probabilistic Classification.** The experimental results presented in Section 4 are discussed in this section. Table 3 illustrates that the step of feature extraction is effective. DK is the time-related features of an ignition pattern but only improves the overall classification accuracy about 1% as compared with the methods without any feature extraction, while FFT and WPT+PCA give about 4.4% and 4.8% improvement, respectively. When combining both time-related and frequency-related features by DK and WPT+PCA, the overall classification accuracy is about 7% higher than that without any feature extraction. Table 3 also indicates that no matter which classification technique is employed, the integration of DK and WPT+PCA as feature extraction gives the best accuracy. In addition, the three classification techniques are compared by using  $F$ -measure as well. Both PNN and RVM employ 1vA strategy for probabilistic classification. In other words, only  $d$  binary classifiers were constructed for  $d$  labels so that there are large indecision regions between pairs of classes. Therefore, when a test case lies on these regions, PNN and RVM mostly fail to classify the faults correctly. However, PCRVM employs 1v1 strategy, which minimizes those indecision regions. Table 3 verifies the effectiveness of the 1v1 strategy because PCRVM outperforms the other two classification techniques. This situation is almost the same as the tests with optimized decision threshold as shown in Tables 6 and 7. Therefore, the proposed PCRVM is a very effective and promising classification technique.

**5.2. Effect of Decision Threshold Optimization.** Tables 3, 6, and 7 illustrate that the GA and PSO can improve the overall accuracy by 3.48% and 3.5% as compared with the fixed decision threshold of 0.5, but these two techniques give nearly the same threshold and  $F_{me}$ . The reason is that the experiment was run for 20 times for both the GA and PSO, and then the pair of results with the highest  $F_{me}$  was returned. However, it is found that the standard deviations of the 20 results for the GA and PSO are 1.02E-3 and 3.23E-4, respectively. For the GA, the standard deviation is larger than PSO in this case study. This result indicates that PSO is more stable than the GA and theoretically requires a fewer number of runs to obtain a suboptimal result than the GA. This is because PSO is somehow insensitive to the initial values, whereas the GA is initialized with random start points within the search space and the search result is very sensitive to the initial values [36]. Consequently, PSO is recommended for this application.

**5.3. Diagnosis of Simultaneous Faults.** Table 8 reveals that the trained classifiers using PNN, RVM, and PCRVM perform well because the test cases contain single-fault patterns only. Due to the advantage of pairwise coupling, PCRVM performs the best among the three classification techniques.

For the test cases of simultaneous-fault patterns, there are only five reasonable combinations of simultaneous faults because not every combination is possible. Since a simultaneous-fault pattern is caused by different single faults, some

TABLE 6: Evaluation of different combinations of techniques using GA-optimized threshold.

| <i>F</i> -measure over <i>VALID</i><br>with GA-optimized $\epsilon_{\text{opt}}$ | Classification technique |                 |                         |                 |                         |                 |
|--|--------------------------|-----------------|-------------------------|-----------------|-------------------------|-----------------|
|  | PNN                      |                 | RVM                     |                 | PCRVM                   |                 |
|  | $\epsilon_{\text{opt}}$  | $F_{\text{me}}$ | $\epsilon_{\text{opt}}$ | $F_{\text{me}}$ | $\epsilon_{\text{opt}}$ | $F_{\text{me}}$ |
| Feature Extraction   |                          |                 |                         |                 |                         |                 |
| None   | 0.6875                   | 0.76114         | 0.6757                  | 0.79604         | 0.7005                  | 0.81244         |
| DK   | 0.7193                   | 0.79819         | 0.6813                  | 0.80325         | 0.7167                  | 0.83023         |
| FFT  | 0.6791                   | 0.79236         | 0.7401                  | 0.82129         | 0.7353                  | 0.82621         |
| WPT + PCA<br>(Haar, level $J = 9$ )  | 0.6998                   | 0.79125         | 0.7161                  | 0.83118         | 0.7234                  | 0.85816         |
| DK + FFT   | 0.7245                   | 0.79123         | 0.6755                  | 0.82212         | 0.7291                  | 0.85414         |
| DK + WPT + PCA<br>(Haar, level $J = 9$ )   | 0.7313                   | 0.81431         | 0.7023                  | 0.84524         | <b>0.7124</b>           | <b>0.87113</b>  |

\*The experiment was run for 20 times, and the best  $\epsilon_{\text{opt}}$  and  $F_{\text{me}}$  were returned.

TABLE 7: Evaluation of different combinations of techniques using PSO-optimized threshold.

| <i>F</i> -measure over <i>VALID</i><br>with PSO-optimized $\epsilon_{\text{opt}}$ | Classification technique |                 |                         |                 |                         |                 |
|---|--------------------------|-----------------|-------------------------|-----------------|-------------------------|-----------------|
|   | PNN                      |                 | RVM                     |                 | PCRVM                   |                 |
|   | $\epsilon_{\text{opt}}$  | $F_{\text{me}}$ | $\epsilon_{\text{opt}}$ | $F_{\text{me}}$ | $\epsilon_{\text{opt}}$ | $F_{\text{me}}$ |
| Feature extraction  |                          |                 |                         |                 |                         |                 |
| None  | 0.6934                   | 0.7711          | 0.6877                  | 0.78014         | 0.6993                  | 0.80563         |
| DK  | 0.6983                   | 0.77132         | 0.6919                  | 0.80532         | 0.7154                  | 0.83011         |
| FFT   | 0.6916                   | 0.78932         | 0.6997                  | 0.81829         | 0.7278                  | 0.82622         |
| WPT + PCA<br>(Haar, level $J = 9$ )   | 0.7024                   | 0.79451         | 0.7068                  | 0.82788         | 0.7177                  | 0.87112         |
| DK + FFT  | 0.7161                   | 0.78398         | 0.6949                  | 0.83129         | 0.7269                  | 0.84941         |
| DK + WPT + PCA<br>(Haar, level $J = 9$ )  | 0.7116                   | 0.82561         | 0.7114                  | 0.85445         | <b>0.7147</b>           | <b>0.88911</b>  |

\*The experiment was run for 20 times and the best  $\epsilon_{\text{opt}}$  and  $F_{\text{me}}$  were returned.

TABLE 8: Single-fault diagnosis under *F*-measure (with DK+WPT+PCA,  $\epsilon_{\text{opt}} = 0.7147$ ).

| Test dataset $TEST_1$<br>(single faults) | PNN    | RVM    | PCRVM  |
|--|--------|--------|--------|
| Fault 1                                  | 0.9185 | 0.9342 | 0.9781 |
| Fault 2                                  | 0.9094 | 0.9118 | 0.9224 |
| Fault 3                                  | 0.9773 | 0.9357 | 0.9911 |
| Fault 4                                  | 0.9087 | 0.9272 | 0.9287 |
| Fault 5                                  | 0.9368 | 0.9458 | 0.9489 |
| Fault 6                                  | 0.8902 | 0.9010 | 0.9020 |
| Fault 7                                  | 0.9722 | 0.9812 | 0.9923 |
| Fault 8                                  | 0.9453 | 0.9477 | 0.9513 |
| Fault 9                                  | 0.9042 | 0.9126 | 0.9242 |
| Fault 10                                 | 0.9035 | 0.9133 | 0.9454 |

of the time-related and frequency-related features may be distorted or even vanished. Therefore, the feature extraction using DK and WPT+PCA cannot work very well and hence the values of  $F_{\text{me}}$  in Table 9 drop a little bit as compared with the values in Table 8, but they can still provide an accuracy

ranging from 0.49 to 0.8. Once again, PCRVM outperforms the other classified techniques because of pairwise coupling strategy. Within the simultaneous-fault diagnosis, the most misclassified fault is Fault 10, because the ignition pattern of Fault 10 is almost distorted by Fault 3. Nevertheless, the experimental results can still verify the following:

- (1) the proposed framework can alleviate the problem of exponential growth of training dataset for simultaneous-fault ignition patterns by training the probabilistic classifier using single-fault patterns only. This evidence can be found in Tables 8 and 9 that the single-fault patterns can be almost correctly classified, while the overall classification accuracy for simultaneous-fault ignition patterns is still satisfactory;
- (2) the feature extraction techniques of DK combined with WPT+PCA can effectively capture the time-related and frequency-related features from single-fault and simultaneous-fault ignition patterns;
- (3) the features of single-fault ignition patterns can really be detected in some feasible simultaneous-fault ignition patterns; this feasibility will create a new research direction for automotive engine diagnosis;

TABLE 9: Simultaneous fault diagnosis under  $F$ -measure with classifiers trained with single faults (with DK + WPT + PCA and  $\epsilon_{\text{opt}} = 0.7147$ ).

| Test dataset $TEST_s$<br>(simultaneous faults only) | Classification technique | Fault 2 | Fault 3 | Fault 5 | Fault 6 | Fault 7 | Fault 8 | Fault 9 | Fault 10 |
|---|--------------------------|---------|---------|---------|---------|---------|---------|---------|----------|
| Faults (2, 6)                                       | PNN                      | 0.6917  | —       | —       | 0.6848  | —       | —       | —       | —        |
|   | RVM                      | 0.7128  | —       | —       | 0.7067  | —       | —       | —       | —        |
|   | PCRVM                    | 0.7435  | —       | —       | 0.7392  | —       | —       | —       | —        |
| Faults (5, 7)                                       | PNN                      | —       | —       | 0.7321  | —       | 0.7054  | —       | —       | —        |
|   | RVM                      | —       | —       | 0.7522  | —       | 0.7214  | —       | —       | —        |
|   | PCRVM                    | —       | —       | 0.7926  | —       | 0.7557  | —       | —       | —        |
| Faults (3, 10)                                      | PNN                      | —       | 0.8003  | —       | —       | —       | —       | —       | 0.4854   |
|   | RVM                      | —       | 0.7943  | —       | —       | —       | —       | —       | 0.6159   |
|   | PCRVM                    | —       | 0.7844  | —       | —       | —       | —       | —       | 0.6995   |
| Faults (2, 5, 9)                                    | PNN                      | 0.7434  | —       | 0.7457  | —       | —       | —       | 0.6809  | —        |
|   | RVM                      | 0.7123  | —       | 0.7359  | —       | —       | —       | 0.7256  | —        |
|   | PCRVM                    | 0.7847  | —       | 0.7858  | —       | —       | —       | 0.7704  | —        |
| Faults (3, 8, 10)                                   | PNN                      | —       | 0.6842  | —       | —       | —       | 0.7118  | —       | 0.6803   |
|   | RVM                      | —       | 0.6724  | —       | —       | —       | 0.7319  | —       | 0.6931   |
|   | PCRVM                    | —       | 0.7349  | —       | —       | —       | 0.7828  | —       | 0.6993   |

—: Indicates that the probability is lower than  $\epsilon_{\text{opt}}$  and hence ignored.

TABLE 10: Evaluation results of the binarization approach and the proposed framework for fault diagnosis.

|                          | Binarization | Proposed framework               |
|--------------------------|--------------|----------------------------------|
| Feature Extraction       | None         | WPT + PCA + DK                   |
| Classification           | SVM          | SVM                              |
| threshold                | None         | None                             |
| $F$ -measure over $TEST$ |              | PSO-optimized threshold = 0.7147 |
| Overall cases            | 0.4518       | 0.6792                           |
| Single fault cases       | 0.4500       | 0.7404                           |
| Simultaneous fault cases | 0.4528       | 0.6455                           |
|                          |              | 0.8891                           |
|                          |              | 0.9567                           |
|                          |              | 0.7714                           |

- (4) RVM is more robust than PNN for probabilistic classification;
- (5) the pairwise coupling (lv1) strategy can improve the accuracy for common probabilistic classification techniques.

**5.4. Comparison with the Latest Approach.** Table 10 reveals that the binarization approach works badly on ignition pattern classification. In other words, the binarization method does not work for engine ignition-system diagnosis. In addition, after feature extraction, the performance of the binarization can be generally raised about 50% as well. Therefore, the effectiveness of feature extraction is verified under all frameworks and techniques tested in this paper. It is highly believed that this feature extraction can also work well in many other practical applications.

## 6. Conclusions

One of the challenges in ignition system diagnosis is that more than one single fault may appear at a time. Another challenge is the acquisition of large amount of costly

simultaneous-fault ignition patterns for constructing the diagnostic system because the number of the training patterns depends on the combination of different single faults. In this paper, simultaneous-fault diagnosis for automotive engine ignition patterns was studied and a new framework combining feature extraction, probabilistic classification, and decision threshold optimization based on a fair multilabel assessment,  $F$ -measure, has successfully been developed. With the proposed diagnosis framework, the acquisition of large amount of simultaneous-fault patterns can be avoided.

In this study, the combination of feature extraction techniques of DK, FFT, and WPT+PCA have been tried along with the classification techniques of PNN, RVM, and PCRVM to tackle the simultaneous-fault diagnosis. The experimental results reveal that PCRVM combined with WPT+PCA and DK performs the best for both single-fault and simultaneous-fault diagnoses. Its average accuracy for single-fault diagnosis is about 0.95 while the average accuracy for simultaneous faults is only about 0.76. It implies that the feature extraction technique based on DK and WPT+PCA for simultaneous-fault detection may not be perfect. Alternative approach, such as the integration of feature extraction, classification, and multiexpert reasoning, could be studied in the future.



This study also shows that the decision threshold for identifying the number of simultaneous faults can be optimized over  $F$ -measure using direct search techniques, such as GA and PSO. Both the GA and PSO generate almost the same decision threshold but PSO requires less computational time and is more stable because of its lower standard deviation in multiple runs. Moreover, PSO has fewer operators and hence fewer adjustable parameters that can further reduce the user burden. Overall speaking, PSO should be the first choice of the threshold optimization technique in the current application.

To further verify the effectiveness of the proposed framework, the latest method, binarization method using SVM, was also employed to diagnose the simultaneous faults. The results show that the diagnosis accuracy of the binarization method is worse than that of the proposed framework. Therefore, the proposed framework is very suitable for engine ignition-system fault diagnosis. Since the proposed framework for simultaneous-fault diagnosis is general, it can be adapted to other similar applications. Finally, the original contributions of the research are summarized as follows.

- (1) The research is a first attempt at integrating DK+WPT+PCA, PCRV, and direct search techniques into a general framework for simultaneous-fault diagnosis of automotive ignition systems.
- (2) The proposed diagnostic system is the first in the literature that can be trained with single-fault signal patterns (i.e., single-fault time-dependent patterns) only, while it can diagnose simultaneous-fault signal patterns too.
- (3) This paper is also the first in the literature that reports that the features of single-fault ignition patterns can be detected in some feasible simultaneous-fault ignition patterns. This fact is an important contribution to automotive engine diagnosis.
- (4) The integration of the pairwise coupling (1v1) strategy into RVM is original, and the 1v1 strategy can really improve the classification accuracy of RVM.

## Notation

**A:** Diagonal matrix of hyperparameters  
**a:** End point of burning time  
**B:** Diagonal matrix in RVM  
 $Bf_j(\cdot)$ :  $j$ th binary classifier  
 $C_i$ :  $i$ th probabilistic classifier  
 $C_i(\mathbf{f})$ : Probability of  $\mathbf{f}$  belonging to the  $i$ th label  
 $C_{ij}$ : Pairwise classifier  
 $C_{ij}(\mathbf{f})$ : Pairwise probability of  $\mathbf{f}$  belonging to the  $i$ th label against the  $j$ th label  
 $c_1$ : Cognitive parameter of PSO  
 $c_2$ : Social parameter of PSO  
**D:** Sample dataset  
 $d$ : Number of labels (faults)  
 $e_j$ :  $j$ th eigen value  
 $e'_j$ :  $j$ th normalized eigen value  
**F:** Set of feature vectors

$F'$ : Set of feature vectors created by WPT and PCA  
 $F_1$ : Firing voltage  
 $F_2$ : Burn time  
 $F_3$ : Average spark voltage of spark line  
 $F_{me}$ :  $F$ -measure  
 $\mathbf{f}$ : Feature vector  
 $\mathbf{f}'$ : Feature vector created by WPT and PCA  
 $\mathbf{f}_h$ :  $h$ th feature vector  
 $f_{class}$ : Probabilistic classifier  
**H:** PCA transformation matrix  
 $\mathbf{h}_j$ :  $j$ th eigen vector  
**J:** Decomposition level of WPT  
 $K(\cdot)$ : Kernel function in RVM  
 $L_p$ : Length of ignition pattern (i.e., number of data point in ignition pattern)  
**l:** True label vector  
 $\mathbf{l}_i$ :  $i$ th true label vector  
 $l_i$ :  $i$ th label in **l**  
 $l_{ig}$ :  $g$ th label in  $\mathbf{l}_i$   
 $l_i^j$ :  $i$ th label in the  $j$ th test data  
**N:** Number of training data  
 $N_D$ : Number of cases in sample dataset  
 $N_T$ : Number of test data  
 $n_{ij}$ : Number of training data with the  $i$ th and  $j$ th labels  
 $P(l_j | \mathbf{f})$ : Probability of  $\mathbf{f}$  belonging to  $l_j$   
 $P(\cdot)$ : Probability  
 $s$ : Input dimension of classifier to be evaluated  
**t:** Set of faulty labels in training dataset  
 $t_n$ : Faulty label of the  $n$ th training case  
**TEST:** Original test dataset  
 $TEST_1$ : Single-fault cases in test dataset  
 $TEST_s$ : Simultaneous-fault cases in test dataset  
 $TEST_F$ : Test dataset after feature extraction  
**TRAIN:** Original training dataset  
 $TRAIN_F$ : Training dataset after feature extraction  
**V:** Set of coefficient vectors  
**VALID:** Original validation dataset  
 $VALID_F$ : Validation dataset after feature extraction  
**v:** Coefficient vector  
**w:** Optimal vector in RVM  
 $w_n$ :  $n$ th optimal parameter in RVM  
 $\mathbf{w}_{MP}$ : Most probable weight vector in RVM  
 $w_c$ : Inertial weight of PSO  
**WPT( $\cdot$ ):** Wavelet packet transform function  
**X:** Set of ignition pattern vectors  
**x:** Unseen ignition pattern  
 $x_i$ :  $i$ th data point in **x**  
**y:** Predicted label vector  
 $y_i$ :  $i$ th predicted label  
 $y_i^j$ :  $i$ th predicted label in the  $j$ th test data  
 $z(\mathbf{f})$ : Predicted decision  
 $\alpha$ : Hyperparameter vector of RVM  
 $\alpha_n$ :  $n$ th hyperparameter of RVM  
 $\varepsilon$ : Decision threshold  
 $\varepsilon_k$ :  $k$ th tentative threshold produced in optimization process

$\varepsilon_{\text{opt}}$ : Optimized threshold  
 $\theta(\cdot)$ : Decision function of binarization approach  
 $\pi$ : Precision  
 $\rho$ : Probability vector  
 $\rho_i$ : Probability of the  $i$ th label  
 $\rho_{ij}$ : Pairwise probability of the  $i$ th label against the  $j$ th label  
 $\Sigma$ : Covariance matrix in RVM  
 $\Sigma_{ii}$ :  $i$ th diagonal element of covariance matrix  $\Sigma$   
 $\sigma(\cdot)$ : Logistic sigmoid function  
 $\tau$ : Recall  
 $\epsilon$ : Initial population  
 $\Phi$ : Design matrix in RVM.

## Acknowledgment

The research is supported by the University of Macau Research Grant nos. MYRG075(Y2-L2)-FST12-VCMM, MYRG141(Y2-L2)-FST11-IWF, and MYRG149(Y2-L2)-FST11-WPK.

## References

- [1] 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.
- [2] W. H. Crouse and D. L. Anglin, *Automotive Mechanics*, McGraw-Hill, Columbus, Ohio, USA, 1993.
- [3] C. C. Liu and J. Chu, *Wave Analysis and Inspection Technology for Computer Controlled Automotive Engine*, Sunlex, Guangzhou, China, 2005.
- [4] C. M. Vong, P. K. Wong, L. M. Tam, and Z. Zhang, "Ignition pattern analysis for automotive engine trouble diagnosis using wavelet packet transform and support vector machines," *Chinese Journal of Mechanical Engineering*, vol. 24, no. 5, p. 870, 2011.
- [5] Y. Wang, Y. Xing, and H. He, "An intelligent approach for engine fault diagnosis based on wavelet pre-processing neural network model," in *Proceedings of IEEE International Conference on Information and Automation (ICIA '10)*, pp. 576–581, June 2010.
- [6] 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, part 2, pp. 7252–7261, 2009.
- [7] 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.
- [8] N. Saravanan, V. N. S. K. Siddabattuni, and K. I. Ramachandran, "Fault diagnosis of spur bevel gear box using artificial neural network (ANN), and proximal support vector machine (PSVM)," *Applied Soft Computing Journal*, vol. 10, no. 1, pp. 344–360, 2010.
- [9] C. M. Vong, P. K. Wong, and W. F. Ip, "Support vector classification using domain knowledge and extracted pattern features for diagnosis of engine ignition systems," *Journal of the Chinese Society of Mechanical Engineers*, vol. 31, no. 5, pp. 363–374, 2010.
- [10] V. K. Rai and A. R. Mohanty, "Bearing fault diagnosis using FFT of intrinsic mode functions in Hilbert-Huang transform," *Mechanical Systems and Signal Processing*, vol. 21, no. 6, pp. 2607–2615, 2007.
- [11] G. Betta, C. Liguori, A. Paolillo, and A. Pietrosanto, "A DSP-based FFT-analyzer for the fault diagnosis of rotating machine based on vibration analysis," *IEEE Transactions on Instrumentation and Measurement*, vol. 51, no. 6, pp. 1316–1321, 2002.
- [12] Y. Liu, L. Guo, Q. Wang, G. An, M. Guo, and H. Lian, "Application to induction motor faults diagnosis of the amplitude recovery method combined with FFT," *Mechanical Systems and Signal Processing*, vol. 24, no. 8, pp. 2961–2971, 2010.
- [13] J. F. Li and C. W. Wu, "Efficient FFT network testing and diagnosis schemes," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 10, no. 3, pp. 267–277, 2002.
- [14] J. D. Wu and C. H. Liu, "An expert system for fault diagnosis in internal combustion engines using wavelet packet transform and neural network," *Expert Systems with Applications*, vol. 36, no. 3, pp. 4278–4286, 2009.
- [15] R. Yan and R. X. Gao, "An efficient approach to machine health diagnosis based on harmonic wavelet packet transform," *Robotics and Computer-Integrated Manufacturing*, vol. 21, no. 4–5, pp. 291–301, 2005.
- [16] R. Zhou, W. Bao, N. Li, X. Huang, and D. R. Yu, "Mechanical equipment fault diagnosis based on redundant second generation wavelet packet transform," *Digital Signal Processing*, vol. 20, no. 1, pp. 276–288, 2010.
- [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] C. M. Vong, H. Huang, and P. K. Wong, "Engine spark ignition diagnosis with wavelet packet transform and case-based reasoning," in *Proceedings of IEEE International Conference on Information and Automation (ICIA '10)*, pp. 565–570, June 2010.
- [19] 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.
- [20] L. H. Chiang, E. L. Russell, and R. D. Braatz, "Fault diagnosis in chemical processes using Fisher discriminant analysis, discriminant partial least squares, and principal component analysis," *Chemometrics and Intelligent Laboratory Systems*, vol. 50, no. 2, pp. 243–252, 2000.
- [21] S. Wang and F. Xiao, "AHU sensor fault diagnosis using principal component analysis method," *Energy and Buildings*, vol. 36, no. 2, pp. 147–160, 2004.
- [22] K. Polat and S. Güneş, "An expert system approach based on principal component analysis and adaptive neuro-fuzzy inference system to diagnosis of diabetes disease," *Digital Signal Processing*, vol. 17, no. 4, pp. 702–710, 2007.
- [23] 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.
- [24] P. M. Ciarelli, E. Oliveira, C. Badue, and A. F. de Souza, "Multilabel text categorization using a probabilistic neural network," *International Journal of Computer Information Systems and Industrial Management Applications*, vol. 1, pp. 133–144, 2009.
- [25] P. Refregier and F. Vallet, "Probabilistic approach for multiclass classification with neural networks," in *Proceedings of International Conference on Artificial Networks*, pp. 1003–1007, 1991.

- [26] P. K. Wong, Q. S. Xu, C. M. Vong, and H. C. Wong, "Rate-dependent hysteresis modeling and control of a piezostage using online support vector machine and relevance vector machine," *IEEE Transactions on Industrial Electronics*, vol. 59, no. 4, pp. 1988–2001, 2012.
- [27] M. E. Tipping, "Sparse Bayesian learning and the relevance vector machine," *Journal of Machine Learning Research*, vol. 1, no. 3, pp. 211–244, 2001.
- [28] 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.
- [29] S. Abe, *Support Vector Machines for Pattern Classification*, Advances in Pattern Recognition, Springer, London, UK, 2nd edition, 2010.
- [30] E. O. Brigham and R. E. Morrow, "The fast Fourier transform," *IEEE Spectrum*, vol. 4, no. 12, pp. 63–70, 1967.
- [31] Z. Sun and C. Chang, "Structural damage assessment based on wavelet packet transform," *Journal of Structural Engineering*, vol. 128, no. 10, pp. 1354–1361, 2002.
- [32] D. J. C. MacKay, "The evidence framework applied to classification networks," *Neural Computation*, vol. 4, no. 5, pp. 720–736, 1992.
- [33] K.-B. Duan and S. S. Keerthi, "Which is the best multiclass SVM method? An empirical study," in *Multiple Classifier Systems*, N. Oza, R. Polikar, J. Kittler, and F. Roli, Eds., vol. 3541, pp. 278–285, Springer, Berlin, Germany, 2005.
- [34] M. Hülsmann and C. Friedrich, "Comparison of a novel combined ECOC strategy with different multiclass algorithms together with parameter optimization methods," in *Machine Learning and Data Mining in Pattern Recognition*, P. Perner, Ed., vol. 4571, pp. 17–31, Springer, Berlin, Germany, 2007.
- [35] R. Alguliev and R. Aliguliyev, "Experimental investigating the F-measure as similarity measure for automatic text summarization," *Applied and Computational Mathematics*, vol. 6, no. 2, pp. 278–287, 2007.
- [36] P. K. Wong, L. M. Tam, K. Li, and C. M. Vong, "Engine idle-speed system modelling and control optimization using artificial intelligence," *Proceedings of the Institution of Mechanical Engineers D: Journal of Automobile Engineering*, vol. 224, no. 1, pp. 55–72, 2010.
- [37] S. W. Lin, T. Y. Tseng, S. Y. Chou, and S. C. Chen, "A simulated-annealing-based approach for simultaneous parameter optimization and feature selection of back-propagation networks," *Expert Systems with Applications*, vol. 34, no. 2, pp. 1491–1499, 2008.



# Hindawi

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

