Neurocomputing 174 (2016) 331–343

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Sparse Bayesian extreme learning committee machine for engine simultaneous fault diagnosis

Pak Kin Wong [a,*], Jianhua Zhong [a], Zhixin Yang [a], Chi Man Vong [b]

[a] Department of Electromechanical Engineering, University of Macau, Macao
[b] Department of Computer and Information Science, University of Macau, Macao

ARTICLE INFO

Article history:
Received 8 September 2014
Received in revised form
16 February 2015
Accepted 25 February 2015
Available online 20 August 2015

Keywords:
Automotive engine
Multi-signal fusion
Simultaneous-fault diagnosis
Sparse Bayesian extreme learning machine
Probabilistic committee machine

ABSTRACT

The automotive engine is prone to various faults due to its complex structure and running conditions. Development of a fast response and accurate intelligent system for fault diagnosis of automotive engines is therefore greatly urged. The engine fault diagnosis faces challenges because of the existence of simultaneous-faults which are multiple single-faults appear concurrently. Another challenge is the high cost in acquiring the exponentially increased simultaneous-fault signals. Traditional signal-based engine fault diagnostic systems may not give reliable diagnostic results because they usually rely on single classifier and one kind of engine signal. To enhance the reliability of fault diagnosis and the number of detectable faults, this research proposes a new diagnostic framework namely probabilistic committee machine (PCM). The framework combines feature extraction (empirical mode decomposition and sample entropy), a parameter optimization algorithm, and multiple sparse Bayesian extreme learning machines (SBELM) to form an intelligent diagnostic framework. Multiple SBELM networks are built to form different domain committee members. The members are individually trained using air ratio, ignition pattern and engine sound signals. The diagnostic result from each fault detection member is then combined by using a new probabilistic ensemble method, which can improve the overall diagnostic accuracy and increase the number of detectable faults as compared to individual classifier. Experimental results show the proposed framework is superior to the existing single probabilistic classifier. Moreover, the proposed system can diagnose both single and simultaneous-faults for automotive engines while the system is trained by single-fault patterns only.

1. Introduction

The automotive engine is the power source; its reliability directly affects the vehicle driving performance and safety. Due to the complex structure and running conditions of the engine, vehicle failure is usually caused by the occurrence of engine fault. Therefore, it is of great significance to develop a fast response and reliable intelligent system for fault diagnosis of automotive engines. One of the main challenges in the fault diagnosis of automotive engines is that, engine fault diagnosis is a simultaneous-fault problem, meaning the several independent single-faults may appear at the same time [1,2]. Another difficulty is a typical multi-signal fusion problem; it involves the use of multi-signal, such as vibration, sound, ignition and air ratio, to simultaneously detect and identify the engine faults [3–5]. Current available methods for engine fault diagnosis are mainly classified into model-based, knowledge-based and data-driven 6–8. Model-based diagnostic methods rely on analytical models, while an exact engine analytical model is difficult to be determined because the engine is a complicated thermo-fluid and mechanical system [8,9]. In addition, no single model-based approach can diagnose all the faults of the engine because of different engine designs and configurations. Thus, model-based methods are too hard in practical use. In terms of knowledge-based methods, decision tree is the early developed approach for fault diagnosis, which makes a diagnosis through a sequence of questions and answers [10]. However, due to the complexity of modern vehicles, the diagnostic knowledge is usually incomplete and vague. So, a reliable decision tree for modern vehicle diagnostic problems is very difficult to obtain. The data-driven methods generally rely on signal-based analysis techniques [6], which are very suitable for analyzing the system whose domain information is insufficient or unknown such as automotive engine [2,11]. Hence, the proposed diagnostic framework in this study is based on the data-driven methods. Among the data-driven methods, machine learning methods are commonly employed to build fault classifiers to deal with signal-based diagnostic problems [11–13].

In traditional machine learning methods for engine fault diagnosis, both single and simultaneous-faults are considered as independent labels. However, considering the simultaneous-faults as a single-label

* Corresponding author. Tel.: +853 88224956; fax: +853 88222426.
E-mail address: fstpkw@umac.mo (P.K. Wong).

http://dx.doi.org/10.1016/j.neucom.2015.02.097
0925-2312/

classification problem will result in high cost in acquiring exponentially increased simultaneous-fault signals. For example, with d single-faults (labels) and one normal condition, there are $2^d-(d+1)$ artificial simultaneous-fault labels. By reviewing the open literature [1,14,15], it is learnt that features of single-fault signal patterns can be found from the simultaneous-fault patterns in some applications, if a proper feature extraction technique is employed. In other words, a simultaneous-fault symptom can be identified by analyzing the single-fault patterns only when the machine learning method is trained by using proper feature extraction technique, so that it can save a lot of resource for collecting a large combination of simultaneous-fault training data. In this way, a proper feature extraction technique is studied in this research. To extract useful fault features, signal processing methods have been widely used. Fast Fourier transform (FFT) is a dominating analysis tool for signal processing, but FFT is not suitable for non-stationary signals. However, many signals for engine diagnosis, such as engine ignition signal, are non-stationary signals which require time-frequency analysis [1,10,16]. In the open literature, many time-frequency analysis methods have been proposed to process the non-stationary signal in both time and frequency domains. One of classical time-frequency analysis methods is short-time Fourier transform (STFT), but its main drawback is that the time and frequency resolutions are linked together. Therefore, STFT is impossible to achieve good time and frequency resolutions at the same time, resulting in information loss. Other methods, such as wavelet transform and empirical mode decomposition, are able to better deal with the problem. Nevertheless, wavelet transform also has its own shortcomings. For example, it suffers from the effect of the energy leakage because any signal character that does not well correlate with the shape of wavelet basis function will be masked or completely ignored [17]. Empirical mode decomposition (EMD), which was firstly proposed by Huang [18], is an effective tool for analyzing nonlinear and non-stationary signals, like the engine signal [19]. To extract the feature from the engine signal, EMD is employed to decompose the signal into a number of intrinsic mode functions (IMFs). By analyzing each IMF which includes the local characteristic features of the signal, the characteristic information of the original signal could be extracted more accurately and effectively. Even though the signal is effectively decomposed into some IMFs, the number of data points of each IMF is still large, resulting in high dimensional input to the fault classifier. In general, a high dimensional input will decrease the accuracy of the fault classifier, so a proper feature selection algorithm is proposed to reduce the input dimension of the fault classifier. An effective statistical algorithm called Sample Entropy (*SampEn*) has recently been introduced to provide one statistical feature to describe the signal regularity in different IMFs [20,21]. Besides, *SampEn* is less sensitive to noise and suitable for short-length time series data, which can fit the characteristics and patterns of engine signals in this application. These properties make *SampEn* an appealing tool for selecting one feature from each IMF. Hence, the input dimension of the fault classifier can be reduced. By reviewing the open literature, it is an original idea to apply EMD together with *SampEn* to extract representative features from simultaneous-fault signal patterns of engines.

As suggested by practicing engineers, a reliable engine diagnosis should be done by analyzing many kinds of engine signals, that is, engine diagnosis is a multi-signal fusion problem. In the open literature, the existing engine fault classifiers are usually constructed based on one kind of engine signal. However, one single classifier cannot cover all the engine faults and may result in a bias result [22,23]. It is learnt from a real world that a group decision is usually better than a single person decision. To let a fault classification system diagnose more faults and generate a more reliable diagnostic result, this study proposes a novel engine fault diagnostic framework namely probabilistic committee machine (PCM) which is created based on the concept of group decision. Committee machine (CM) consists of a group of diverse classifiers. With this structure, CM can combine results acquired by individual classifier so as to obtain a group decision that is superior to any individual classifier acting alone. This concept has been used to solve some decision and multi-signal fusion problems [24,25], except engine diagnosis.

Selection of fault classifiers for the committee machine is very important to multi-signal diagnostic problems. Moreover, a proper individual classifier must be able to offer the probabilities of all possible faults so that 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. Therefore, it is logical to employ probabilistic classifier for each member in the committee machine. Currently, there are two common probabilistic classifiers, probabilistic neural network (PNN) and relevance vector machine (RVM), available in the existing literature [1,14]. The main drawback of PNN is that the complexity of the network and the training time are heavily related to the number of inputs. In term of RVM, it is the Bayesian approach for support vector machines; the obstacle is that the complexity of Hessian matrix computing attains $O(N^2)$ where N is the number of training data, which hinders RVM to deal with large-scale data, like the problem of multiple-engine signal diagnosis presented in this paper. Recently, an improved statistical method based on extreme learning machine namely sparse Bayesian extreme learning machine (SBELM), which is a logistic regression algorithm in the neurons space, was developed by the authors to overcome the above problems in classification [26]. SBELM inherits the fast training time from extreme learning machine, which prunes the number of corresponding hidden neurons (L) to minimum, and the sparsity of weights from the sparse Bayesian learning approach. According to [26], SBELM usually trends to achieve the best accuracy with $L \approx 50$, which is much less than N being used in RVM. Noted that the complexity of Hessian matrix of SBELM is approximately $O(L^2)$, $L \ll N$. Hence, the execution time for SBELM is significantly reduced as compared with RVM. Therefore, it is believed that the fast training time and the property of sparsity can enable SBELM to effectively deal with large-scale data. Besides, SBELM can let the user easily define its architecture because the classification accuracy of SBELM is insensitive to its hyperparameter, number of hidden nodes, whereas PNN and RVM do not have this attractive feature. As a result, SBELM is selected as a training algorithm for building the classifier as each committee member in this study. Generally, the aforementioned probabilistic classifiers are designed to solve the binary classification. Moreover, most of the practical problems are multi-class classification. One-versus-all strategy is usually employed to fix the multi-class classification problems. However, this strategy does not consider the issue of pairwise correlation (i.e., the correlation between every pair of faults or labels), which was verified to produce a large region of indecision [27,28]. To solve the multi-class classification problem effectively and produce probability output in this study, a suitable pairwise coupling strategy is adopt for SBELM. To demonstrate the effectiveness of the proposed pairwise coupled SBLEM, PNN and RVM with the same pairwise coupling strategy are also studied and compared. The detail explanation of the advantage of pairwise coupling strategy is discussed in Section 2.2.

Although the probabilities of engine faults can be obtained via the pairwise probabilistic classifiers, these probabilities only indicate the chance of occurrence of the engine faults. To determine the occurrence of the engine faults, a decision threshold ε is necessary to apply to those probabilities (e.g., Fault A occurs if the probability of Fault A is greater than ε). The decision threshold can be subjectively determined by human experts and also objectively determined through an optimization algorithm over sample data. This step is called *decision threshold optimization*. It is also learnt from the real world that different committee members will have various reliability and sensitivities to engine faults, so a fair committee machine should assign different weights to their committee members. Due to the member's weight and decision threshold are the major factors

affecting the system accuracy, an efficient searching algorithm, such as particle swarm optimization (PSO) [29,30], for optimal setting of the member's weight and decision threshold is considered in the proposed framework, which is presented in Section 2.3. Finally, a fair measure is also employed to evaluate the performance of the proposed diagnostic framework.

This paper is organized as follows. Section 2 presents the proposed framework and the related techniques. The experimental setup and data per-processing are discussed in Section 3. Section 4 discusses the experimental result and a comparison with the latest approaches. Finally, a conclusion is given in Section 5.

2. Proposed PCM framework and related techniques

The proposed probabilistic committee machine framework for the automotive engine simultaneous-fault diagnosis and its evaluation approach and construction method are illustrated in Fig. 1. The novel framework consists of four sub-modules: (1) feature extraction; (2) probabilistic committee machine; (3) optimizer for committee member weights and decision threshold; and (4) performance evaluation.

In the feature extraction sub-module, the features are extracted from the three kinds of engine signals including the air ratio (or λ), ignition pattern, and acoustic wave, which are denoted as x_1, x_2, x_3 respectively. The three kinds of engine signals are used for demonstration only. Practically, the diagnostic system can include more engine signals for diagnosis. Taking λ signal as an example, the signal x_1 is divided into three independent groups including validation dataset, training dataset, and test dataset, all of which are collected from experiments. 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. Then EMD is used to decompose the signal of air ratio, in which the training feature set, $x_{1\text{-}EMD\text{-}Train}$, validation feature set, $x_{1\text{-}EMD\text{-}Vali}$, and the unseen signal for testing, $x_{1\text{-}EMD\text{-}Test}$, are generated. Considering the existence of irrelevant and redundant information in the extracted features, *SampEn* is then applied to select the features from $x_{1\text{-}EMD\text{-}Train}$, $x_{1\text{-}EMD\text{-}Vali}$, and $x_{1\text{-}EMD\text{-}Test}$. The results are saved as $x_{1\text{-}SE\text{-}Train}$, $x_{1\text{-}SE\text{-}Vali}$ and $x_{1\text{-}SE\text{-}Test}$ respectively. In order to guarantee that all the features have even contribution, every feature in $x_{1\text{-}SE\text{-}Train}$, $x_{1\text{-}SE\text{-}Vali}$ and $x_{1\text{-}SE\text{-}Test}$ is normalized to $[-1, 1]$. Finally, the processed training dataset, validation dataset and unseen signals for testing are named as $x_{1\text{-}Proc\text{-}Train}$, $x_{1\text{-}Proc\text{-}Vali}$ and $x_{1\text{-}Proc\text{-}Test}$ respectively. Apart from the signal features, 2, 3, and 10 domain knowledge (DK) [31] features are added to the signal features for λ, ignition and sound signals respectively.

The proposed framework is an ensemble by three SBELMs, which are trained using the three processed training datasets $x_{Proc\text{-}Train}$ respectively. The output of the trained PCM model together with processed validation dataset $x_{Proc\text{-}Vali}$ is used to perform parameter optimization including the weight of each committee member and the decision threshold. Note that the optimizer is not used in the procedure of diagnosis. Then the kth optimal weight is combined with the kth member's probabilistic output, where the unseen test dataset $x_{k\text{-}Proc\text{-}Test}$ is used, and their results are systematically combined to produce the final output. Moreover, a decision threshold, which is optimized by PSO, is utilized to identify simultaneous-faults under probabilistic committee machine output. The performance evaluation sub-module adopts F-measure (F_{me}). The details of the four sub-modules in the framework and the acquisition of experimental data are discussed in the following sub-sections.

2.1. Sparse Bayesian extreme learning machine for classification

Different from typical extreme learning machine in calculating $\mathbf{H}^{+}$ that is the inverse of matrix hidden layer output $\mathbf{H}$ [32,33]; SBELM employs the Bayesian mechanism to learn the output weights $\mathbf{w}$. The input for SBELM is the hidden layer outputs $\mathbf{H}$, in which $\mathbf{H}=[\mathbf{h}_1(\mathbf{x}_1), \ldots, \mathbf{h}_1(\mathbf{x}_N)]^{\mathrm{T}} \in \mathrm{R}^{N\times(L+1)}$ and $\mathbf{h}_i(\mathbf{x}_i)=[1,\ g_1(\theta_1\cdot\mathbf{x}_i+b_1), \ldots, g_L(\theta_L\cdot\mathbf{x}_i+b_L)]$, where $i=1,\ldots,N$, $g(.)$ is activation function of hidden layer, θ is weight vector connecting the hidden and input nodes, b is the threshold of the hidden node. For two-class classification, every training sample can be considered as an independent Bernoulli event $\mathrm{P}(t\,|\,\mathbf{x})$. The

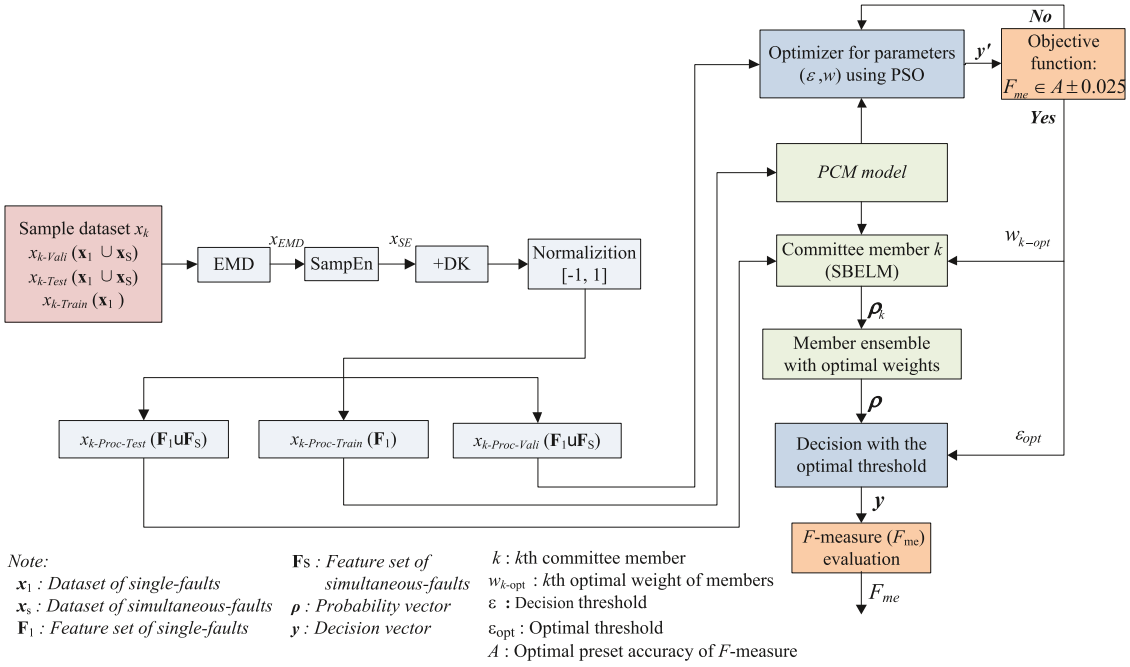


Fig. 1. Proposed framework of engine simultaneous-fault diagnosis using probabilistic committee machine (from training to testing).

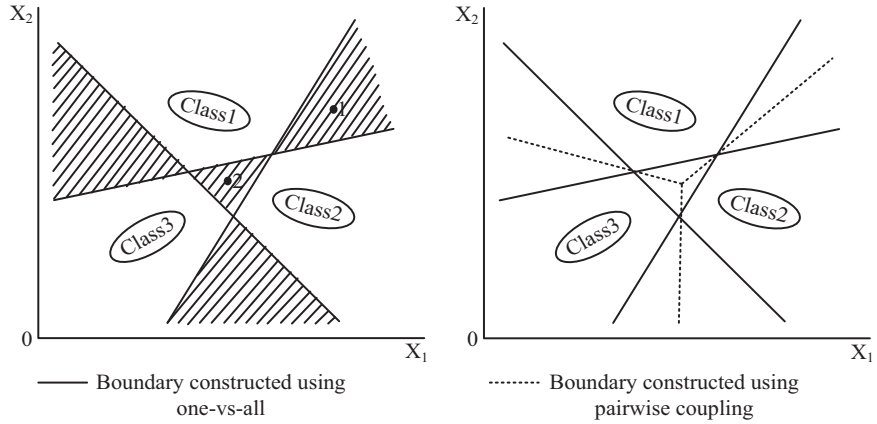


Fig. 2. Indecisive regions (shaded regions) using one-versus-all (left) and pairwise coupling (right).

likelihood is expressed as:

$$\mathrm{P}(\boldsymbol{t}\,|\,\mathbf{w},\ \mathbf{h}) = \prod_{i=1}^{N} \sigma\left[y(\mathbf{h}_i;\mathbf{w})\right]^{t_i}\left\{1-\sigma\left[y(\mathbf{h}_i;\mathbf{w})\right]\right\}^{1-t_i} \tag{1}$$

where $\sigma(.)$ is sigmoid function $\sigma\left[y(\mathbf{h};\mathbf{w})\right] = 1/(1+e^{-y(\mathbf{h};\mathbf{w})})$, $y(\mathbf{h};\mathbf{w}) = \mathbf{hw}$, $\mathbf{t} = (t_1 \ldots t_N)^T$, $t_i = \{0,1\}$ and $\mathbf{w} = (w_0, w_1, \ldots, w_L)^T$. A zero-mean Gaussian distribution over each parameter w_i conditions on an automatic relevance determination (ARD) of hyperparameter a_i [34,35] is expressed by

$$\mathrm{P}(w_i\,|\,a_i) = \mathcal{N}\left(w_i\,|\,0,\ a_i^{-1}\right) \tag{2}$$

$$a = [a_0, a_1, \ldots, a_L]^T$$

$$\mathrm{P}(\mathbf{w}\,|\,a) = \prod_{k=0}^{L} \frac{\alpha_i}{\sqrt{2\pi}} \exp\left(-\frac{\alpha_i w_i^2}{2}\right) \tag{3}$$

There always exists an independent a_i associated with each w_i; some values of w_i are to be zero when a_i tends to infinity. The values of the hyperparameters $\boldsymbol{a}$ are calculated by maximizing the marginal likelihood by integrating the weight parameters $\mathbf{w}$

$$\mathrm{P}(\mathbf{t}\,|\,\boldsymbol{a},\ \mathbf{H}) = \int \mathrm{P}(\mathbf{t}\,|\,\mathbf{w},\ \mathbf{H})\mathrm{P}(\mathbf{w}\,|\,\boldsymbol{a})d\mathbf{w} \tag{4}$$

However, Eq. (4) cannot be directly integrated out. To solve this problem, ARD approximates a Gauss for it with Laplace approximation approach, such that $\log\left\{P(\mathbf{t}\,|\,\mathbf{w},\mathbf{H})P(\boldsymbol{w}\,|\,\boldsymbol{a})\right\} \propto \mathcal{N}(\mathbf{w}_{MP}, \Sigma)$. Where $\mathbf{w}_{MP}$ and Σ are the center and covariance matrix of Gaussian distribution respectively. Generally, Newton–Raphson method – iterative reweighted least-squares algorithm (IRLS) is effectively applied to find $\mathbf{w}_{MP}$. To determine $\mathbf{w}_{MP}$, the gradient $\nabla\mathbf{E}$ and Hessian matrix Φ are necessary to be figured out by

$$\nabla\mathbf{E} = \nabla_w \log\left\{P(\mathbf{t}\,|\,\mathbf{w},\mathbf{H})P(\boldsymbol{w}\,|\,\boldsymbol{a})\right\} = \mathbf{H}^{\mathrm{T}}(\mathbf{t}-\mathbf{y}) - \mathbf{Aw} \tag{5}$$

$$\Phi = \nabla_w\nabla_w \log\left\{P(\mathbf{t}\,|\,\mathbf{w},\mathbf{H})P(\boldsymbol{w}\,|\,\boldsymbol{a})\right\}\big|_{w_{MP}} = -(\mathbf{H}^{\mathbf{T}}\mathbf{BH}+\mathbf{A}) \tag{6}$$

where $\mathbf{y} = [y_1, y_2, \ldots, y_N]^T$, $\mathbf{A} = \mathrm{diag}(\boldsymbol{a})$, $\mathbf{B} = \mathrm{diag}(\beta_1, \beta_2, \ldots, \beta_N)$ is a diagonal matrix with $\beta_i = y_i(1-y_i)$. Therefore, based on IRLS, $\mathbf{w}_{MP}$ is obtained by the following function:

$$\mathbf{w}_{MP} \leftarrow \mathbf{w}_{MP} - \Phi^{-1}\nabla\mathbf{E} = (\mathbf{H}^{\mathrm{T}}\mathbf{BH}+\mathbf{A})^{-1}\,\mathbf{H}^{\mathrm{T}}\mathbf{B}\hat{\boldsymbol{t}} \tag{7}$$

where $\hat{\boldsymbol{t}} = \mathbf{Hw} + \mathbf{B}^{-1}(\mathbf{t}-\mathrm{y})$. By referring to [26], the center $\mathbf{w}_{MP}$ and covariance matrix Σ of Gauss distribution over $\mathbf{w}$ by Laplace approximation are

$$\Sigma = (\mathbf{H}^{\mathrm{T}}\mathbf{BH}+\mathbf{A})^{-1} \tag{8}$$

$$\mathbf{w}_{MP} = \Sigma\mathbf{H}^{\mathrm{T}}\mathbf{B}\hat{\boldsymbol{t}} \tag{9}$$

Therefore, $\log\left\{P(\mathbf{t}\,|\,\mathbf{w},\mathbf{H})P(\boldsymbol{w}\,|\,\boldsymbol{a})\right\} \propto \mathcal{N}(\mathbf{w}_{MP}, \Sigma)$ is obtained. After gaining Gaussian approximation for $\mathbf{w}$, the integral of product of the two prior probability functions of Eq. (4) becomes tractable. Setting the differential of $\mathscr{L}(\alpha) = \mathrm{Log}\ P(\mathbf{t}\,|\,\boldsymbol{a},\mathbf{H})$ with respect to α to zero, it yields

$$\frac{\partial\mathscr{L}(\alpha)}{\partial\alpha_i} = \frac{1}{2\alpha_i} - \frac{1}{2}\Sigma_{ii} - \frac{1}{2}\mathrm{w}_{MPi}^2 = 0 \ \Rightarrow \alpha_i^{new} = \frac{1-\alpha_i\Sigma_{ii}}{\mathrm{w}_{MPi}^2} \tag{10}$$

After the maximum number of iterations through Eq. (10), most elements of $\boldsymbol{a}$ tend to infinity. According to the mechanism of ARD, ARD prior prunes the corresponding hidden neurons when the elements of $\mathbf{w}$ associated with $\boldsymbol{a}$ tend to zero. The final probability distribution $\mathrm{P}(t_{new}\,|\,\mathbf{x}_{new},\ \mathbf{w}_{MP})$ is predicted by using sparse weight based on $\mathcal{Y}(\mathbf{h};\mathbf{w}_{MP}) = \mathbf{hw}_{MP}$ and $\sigma\left[\mathcal{Y}(\mathbf{h};\mathbf{w}_{MP})\right] = \left(1+e^{-\mathcal{Y}(\mathbf{h};\mathbf{w}_{MP})}\right)^{-1}$.

2.2. Pairwise probabilistic multi-label classification

The traditional classifiers and SBELM are designed only for the issue of binary classification. However, most of the practical problems are multi-class classification as well as require probabilistic output. Usually, one-versus-all strategy is employed to deal with multi-class classification problems. One-versus-all strategy constructs a group of classifiers $\boldsymbol{l}_{class} = [C_1, C_2, \ldots, C_d]$ in a d-label classification problem. The one-versus-all strategy is simple and easy to implement. However, it generally gives a poor result [27,28] since one-versus-all does not consider the pairwise correlation and hence induces a much larger indecisive region than pairwise coupling strategy (using one-versus-one) as showed in Fig. 2. Pairwise coupling strategy also constructs a group of classifiers $\boldsymbol{l}_{class} = [\boldsymbol{C}_1, \boldsymbol{C}_2, \ldots, \boldsymbol{C}_d]$ in a d-label classification problem. However, each $\boldsymbol{C}_i = [C_{i1}, \ldots C_{ij}, \ldots, C_{id}]$ is composed of a set of $d-1$ different pairwise classifiers C_{ij}, $i \neq j$. Since C_{ij} and C_{ji} are complementary, there are totally $d(d-1)/2$ classifiers in $\boldsymbol{l}_{class}$ as shown in Fig. 3. To solve the multi-class classification as well as probabilistic output problem, pairwise coupling strategy is adopted for the aforementioned probabilistic classifiers, PNN, RVM, and SBELM.

For a new instance, the strategy combines all the outputs of every pair of classifier C_{ij} to re-estimate the overall probability ρ_i where $i=1$ to d. Note that each ρ_i is an independent probabilistic output for the ith single fault. The nature of simultaneous-fault diagnosis is that $\sum\rho_i$ is unnecessarily equal to 1. There are several available methods for pairwise coupling strategy [28], which are, however unsuitable for simultaneous-fault diagnosis because of their constraint $\sum\rho_i = 1$ [26]. Therefore, the following simple pairwise coupling strategy for simultaneous-fault diagnosis is proposed. Every ρ_i is calculated as

$$\rho_i = C_i(\boldsymbol{x}) = \frac{\sum_{i=1:i\neq j}^{d} n_{ij}C_{ij}(\boldsymbol{x})}{\sum_{j=1:i\neq j}^{d} n_{ij}} = \frac{\sum_{j=1:i\neq j}^{d} n_{ij}\rho_{ij}}{\sum_{j=1:i\neq j}^{d} n_{ij}} \tag{11}$$

where n_{ij} is the number of training feature vectors with either ith or jth labels. Hence, the probability can be more accurately estimated from $\rho_{ij}=C_{ij}(\mathbf{x})$ because the pairwise correlation between the labels is taken into account. To simplify the notation, the SBELM hereafter refers to the new version of pairwise-coupled SBELM, which differs from the SBELM presented in [26] in terms of pairwise coupling strategy adopted.

2.3. Optimization of weights of CM and decision threshold using PSO

After the committee machine is constructed and trained, the optimal weight w_{opt} of each committee member can be selected. In addition to that, the decision threshold ε is also the major factor affecting the classification accuracy. Hence, this research employs a well-known optimization algorithm, PSO, to determine the best weights and decision threshold ε_{opt}. PSO can produce results with less standard deviation whereas the results of typical method, genetic algorithms (GA), are unrepeatable under the same problem. PSO also has the same effectiveness as GA in finding the global optimal solution, but with significantly better computational efficiency. Therefore, PSO is adopted to obtain w_{opt} and ε_{opt} based on $x_{Proc\text{-}Vali}$ dataset.

2.4. Ensemble of committee members with optimal weights

One of most frequently used ensemble methods is weight averaging. In this method, every committee member has a suitable weight related to its ability to generalization. A Weight averaging method is very well to deal with the ensemble problem of balanced committee sensitivities to engine faults, which is demonstrated in Table 1. However, the weight averaging method cannot give a fair or predictable result when it deals with the issue of imbalanced committee member sensitivities to engine faults. In Table 1, when the committee member 2 is untrained by the dataset for fault d_2, its output probability ρ_{2-2} is usually unpredictable. However, the weight averaging method still uses the unpredictable output to calculate the overall average, resulting in an unfair or unpredictable result.

To overcome the above problem, a novel ensemble method using predefined null output and weight for untrained label is proposed which is given by Eq. (12). In Eq. (12), $w_{k\text{-}opt}$ and ρ_{k-i} are presented to be zero when the kth classifier cannot make a diagnosis for ith fault label (i.e. kth classifier is not trained by ith single-fault). In this way, the proposed method can overcome the problem of the traditional weight averaging method. The probability of the ith fault is expressed as:

$$P_i=\frac{\sum_{K=1}^{M} w_{k-opt}\rho_{k-i}}{\sum_{K=1}^{M} f(w_{k-opt})},\ \ i=1,2,\ldots,d \ \text{ and } \ k=1,2,\ldots M$$

$$\text{subject to } f(w_{k-opt})=\begin{cases} w_{k-opt} \\ 0 : \text{if } \rho_{k-i}=0 \end{cases} \tag{12}$$

where $w_{k\text{-}opt}$ is the optimal weights for kth committee member, $w_{k\text{-}opt} \in [0\ 1]$, $k=1$–M where M is the number of committee members, and sum of $w_{k\text{-}opt}$ is not equal to 1; $\rho_{k-i}\in[0\ 1]$ is the kth classifier output for the ith single-fault probability, $i=1$ to d where d is the total number of detectable single-faults. Finally, the probabilistic outputs of classifiers are combined with optimal weights to generate the probability vector $\boldsymbol{P}=[\mathrm{P}_1, \mathrm{P}_2,\ldots,\mathrm{P}_d]$.

2.5. Decision threshold (DT)

The probability vector $\boldsymbol{P}$ can be provided to the user as a quantitative measure for reference and further processing. Practically, it is desirable to identify the multi-label decision as a decision vector $\boldsymbol{y}=[y_1, y_2,\ldots,y_d]$ for all $y_i\in\{0, 1\}$[23,28], which is, however, not directly available from the probability vector $\boldsymbol{P}$. By applying a decision threshold ε_{opt}, $\boldsymbol{y}$ could be derived from $\boldsymbol{P}$, i.e., $y=\mathrm{DT}(\mathbf{P})=[\varepsilon(\mathrm{P}_1), \varepsilon(\mathrm{P}_2), \ldots, \varepsilon(\mathrm{P}_d)]$, where

$$y_i=\begin{cases} 0 \\ 1 \quad \text{if } \ P_i\geq\varepsilon \end{cases} \tag{13}$$

where $\varepsilon \in [0\ 1]$ is a decision threshold and y_i indicates unseen $\boldsymbol{x}$ belongs to the ith label or not, $i=1$ to d. For example, if $\varepsilon=0.5$ and $\boldsymbol{P}=l_{class}(\mathbf{x})=[0.72,\ 0.33,\ 0.58,\ 0.89,\ 0.45]$, then $\mathbf{y}=\mathrm{DT}(\mathbf{P})=[1,0,1,1,0]$. Therefore, $\boldsymbol{x}$ is diagnosed as a simultaneous-fault for the labels (1, 3, 4). When the result of simultaneous-faults is obtained, the F-measure is employed to evaluate the accuracy of the proposed framework.

2.6. F-measure

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-

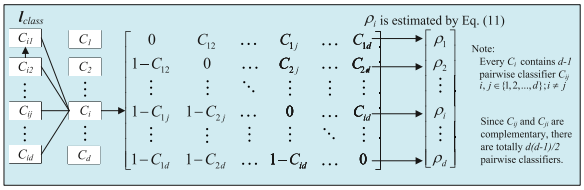


Fig. 3. Pairwise coupling strategy of probabilistic classification.

Table 1
Issue of the weight averaging method for balanced and unbalanced committee member sensitivities to engine faults.

Balanced member sensitivities to engine faults	Committee member 1	Committee member 2	Average output probability (P_1) for d_1
Fault d_1	Trained	Trained	$P_1=\frac{w_1\rho_{1-1}+w_2\rho_{2-1}}{w_1+w_2}\in[0\ \ 1]$
Output probability for d_1 for an unseen case	$\rho_{1-1}\in$ [0 1]	$\rho_{2-1}\in$ [0 1]	P_1 is a reasonable result
Unbalanced member sensitivities to engine faults	Committee member 1	Committee member 2	Average output probability (P_2) for d_2
Fault d_2	Trained	Untrained	$P_2=\frac{w_1\rho_{1-2}+w_2\rho_{2-2}}{w_1+w_2}$
Output probability for d_2 for an unseen case	$\rho_{1-2}\in$ [0 1]	ρ_{2-2} is unpredictable	P_2 is an unfair/unpredictable result

Remark: w_1 and w_2 are weights for Committee members 1 and 2 respectively; P_1 and P_2 are average output probabilities for d_1 and d_2 respectively.

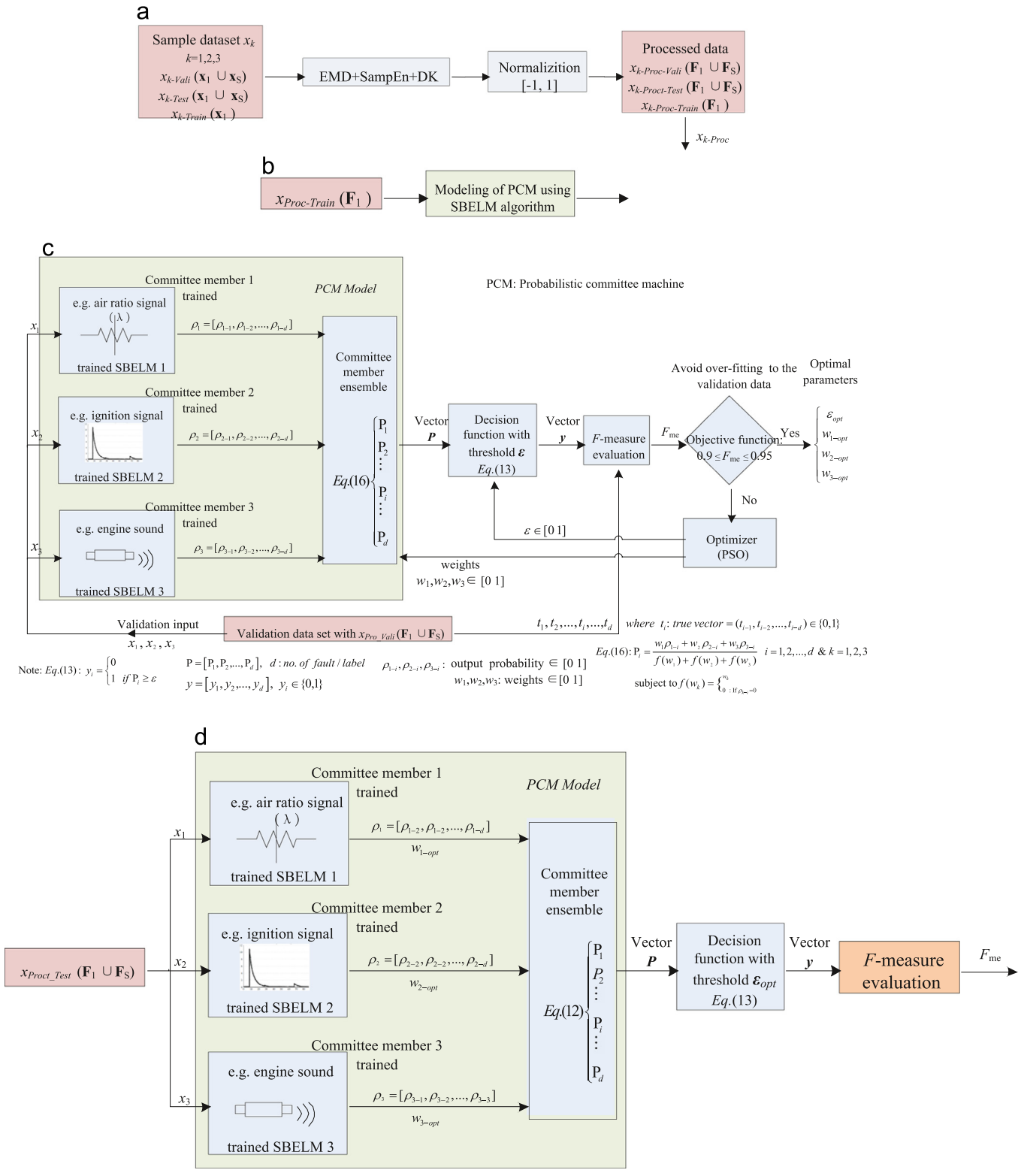


Fig. 4. Detailed workflow of the proposed diagnostic framework. (a) Feature extraction. (b) Model training. (c) Optimization framework for optimization of w_k and ε. (d) Trained diagnostic system and evaluation test.

measure [36] is employed. *F*-measure is currently the most accepted statistical test to evaluate the performance of simultaneous-fault diagnostic systems [1,2,14,37]. 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 by Eq. (14). The larger the *F*-measure value, the higher the diagnostic accuracy will be

$$F_{me} = \frac{2\sum_{j=1}^{d}\sum_{i=1}^{N_t} y_{ij}t_{ij}}{\sum_{j=1}^{d}\sum_{i=1}^{N_t} y_{ij} + \sum_{j=1}^{d}\sum_{i=1}^{N_t} t_{ij}} \in [0,\ 1] \tag{14}$$

where $\boldsymbol{y}_i=[y_{i1},\ldots y_{ij},\ldots,y_{id}]$ and $\mathbf{t}_i=[t_{i1},\ldots t_{ij},\ldots,t_{id}]$ are the predicted decision vector and the true decision vector, respectively, for $j=1$ to d and $i=1$ to N_t and $\forall y_{ij}$, $t_{ij}\in\{0, 1\}$. N_t is the number of single-fault and simultaneous-fault test patterns. For optimization of the weights and decision threshold, F_{me} also serves as an important parameter in an objective function. In order to avoid over-fitting on validation dataset, the objective function is specifically defined as

$$F_{me}\in A\pm 0.025 \tag{15}$$

where A is the preset optimal accuracy of F-measure and A lies between 0 and 1. In this study, A is set to be 0.925 as a trial. The optimization of A will be left as a future work.

2.7. Summary of the proposed framework

The proposed framework and techniques are summarized in Fig. 4, where three committee members ($M=3$) are adopted. Fig. 4 (a) shows the workflow of combining techniques, including EMD, *SampEn*, DK, as feature extraction sub-module. Every dataset x_k divided into training, validation and testing are required to go through the feature extraction process. Fig. 4(b) shows $x_{Pro\text{-}Train}$ ($\mathbf{F}_1$) is used to train the probabilistic committee machine, where the training dataset of simultaneous-fault patters is unnecessary and its architecture is depicted in the PCM model. In this research, the PCM model is an ensemble of three SBELMs. In order to diversify the classifiers, the three SBELMs are respectively trained by using three kinds of engine signals, including air ratio, ignition pattern and engine sound signals. Then, the PCM model is passed to PSO to search for the optimal weights of each committee member and the decision threshold based on the validation set $x_{Pro\text{-}Vali}$ ($\mathbf{F}_1\cup\mathbf{F}_S$) and F_{me} (Fig. 4(c)). To optimize the weights of the committee members and decision threshold, the fitness of each candidate parameter is obtained and evaluated based on F_{me} using $x_{Pro\text{-}Vali}$ ($\mathbf{F}_1\cup\mathbf{F}_S$). In Fig. 4(d), the kth committee member is tested by using test dataset $x_{k\text{-}Pro\text{-}Test}$ ($\mathbf{F}_1\cup\mathbf{F}_S$) that outputs its own probability vector $\rho_k=[\rho_{k-1},\rho_{k-2},\ldots,\rho_{k-d}]$, then the results are combined to produce the overall probability vector $\boldsymbol{P}=[\mathrm{P}_1, \mathrm{P}_2,\ldots, \mathrm{P}_d]$ based on the obtained optimal weights. After $\boldsymbol{P}=[\mathrm{P}_1, \mathrm{P}_2,\ldots,\mathrm{P}_d]$ is ready, the optimal threshold obtained is employed to identify the multi-label decision vector $\boldsymbol{y}=[y_1, y_2,\ldots,y_d]$. Finally, the accuracy of the proposed framework is evaluated by using F-measure.

3. Experiment setup and data preprocessing

To obtain representative sample data for classifier construction and verify the effectiveness of the proposed framework, experiments were carried out. The details of the experimental setup are discussed in the following subsections. All the proposed methods mentioned were implemented by using MatLab R2008a and executed on a computer with a Core 2 Duo E6750 @ 2.13 GHz with 4 GB RAM onboard.

3.1. Test rig and sample data acquisition

In this case study, a well-known inline 4-cylinder engine, HONDA B18C, was employed as the test platform. The proposed experimental setup is shown in Fig. 5, in which a computer linked inductive pickup clamp, five gas analyzer and microphone are used to acquire the air ratio, ignition and acoustic wave patterns respectively. According to the microphone orientation suggested by SAE J1492 standard, the microphone was installed at a distance of 50 cm away from the exhaust pipe and off the ground by 20 cm. In total, 15 kinds of single-faults as described in Tables 2 and 5 kinds of simultaneous-faults as described in Table 3 were imitated and

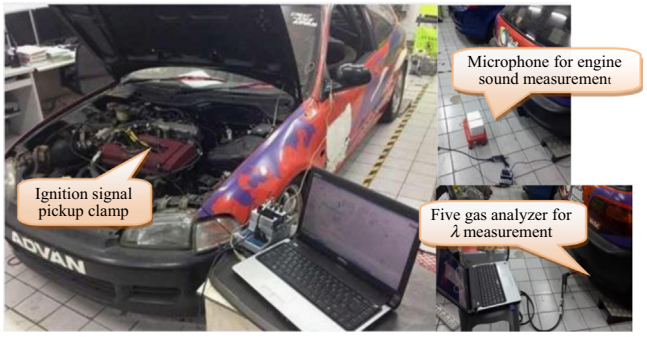


Fig. 5. Collection of faulty patterns from a HONDA B18C engine.

Table 2
Sample single-faults of engines.

Case no.	Single-faults
d_1	Retarded ignition timing
d_2	High resistance in secondary circuit
d_3	Partially broken spark-plug cable
d_4	Defective spark plug
d_5	Narrow spark plug gap
d_6	Misfire due to extremely lean mixture
d_7	Carbon fouled in spark plug
d_8	Engine knock
d_9	Rich mixture
d_{10}	Wide spark plug gap
d_{11}	Clogged air filter element
d_{12}	Defective fuel injector in one or more cylinder
d_{13}	Defective spark plug or ignition circuit in one or more cylinders
d_{14}	Leakage in air manifold
d_{15}	Lean mixture

selected as demonstration examples. Note that the combination of single-faults is not randomly generated; the imitation of the simultaneous-faults is based on reasonable combinations (e.g., it is impossible to have narrow spark plug gap and wide spark plug gap at the same time). The relationship between engine faults and signal types is presented in Table 4, in which three kinds of signals are acquired at the same time when a fault is simulated. Table 4 also reveals that the more the signal types, the larger the number of detectable faults is, so the concept of committee machine can enhance the number of detectable faults.

To construct and test the proposed diagnostic framework, the sample for each of the single-faults and simultaneous-faults, was repeated 100 times under two standard engine testing conditions, slow idle speed and fast idle speed (i.e. 2200 ± 300 rpm). Each time, 1 s of raw signal x_k, including λ, ignition and sound signals, is simultaneously recorded with a sampling rate of 8092 Hz. The sampling rate is set to a frequency higher than the test engine speeds, which can ensure no missing signal. In other words, one case of each signal type has 8092 sampling data points. Finally, for each signal type x_k, there are 1500 single-fault sample data (i.e. 15 kinds of single-faults $\times$ 100 samples) and 500 simultaneous-fault sample data (i.e. 5 kinds of simultaneous-faults $\times$ 100 samples). In order to evaluate the diagnostic performance for both single-faults and simultaneous-faults, each sample data x_k is divided into different subsets as shown in Table 5, where $x_{k\text{-}Valid\text{-}1}$ denotes the validation set of 400 single-fault patterns without feature extraction, $x_{k\text{-}Valid\text{-}S}$ denotes the validation set of the extracted features of 300 simultaneous-fault patterns. Fig. 6 shows some examples of faulty engine signals at slow idle speed condition. It can be observed from Fig. 6 that the engine signals are difficult to analyze manually, so the proposed intelligent diagnostic framework is very helpful to engine fault diagnosis.

Table 3
Sample simultaneous-faults of engines.

Case no.	Simultaneous-faults
si_1	High resistance in secondary circuit and misfire due to extremely lean mixture
si_2	Narrow spark plug gap and carbon fouled in spark plug
si_3	Partially broken spark-plug cable and wide spark plug gap
si_4	High resistance in secondary circuit and narrow spark plug gap and rich mixture
si_5	Partially broken spark-plug cable and engine knock and wide spark plug gap

Table 4
Relationship of single-faults and signal types.

Committee member no.	Signal type	d_1	d_2	d_3	d_4	d_5	d_6	d_7	d_8	d_9	d_{10}	d_{11}	d_{12}	d_{13}	d_{14}	d_{15}
1	λ						√			√		√			√	√
2	Ignition	√	√	√	√	√		√	√		√			√		
3	Sound	√			√		√	√	√			√	√	√	√	

Table 5
Division of sample dataset into different subsets.

	Type of dataset	Single-faults (1500)	Simultaneous-faults (500)
Raw sample data (x_k)	Validation dataset	$D_{k\text{-}Valid1}$ (400)	$D_{k\text{-}Valid\text{-}S}$ (300)
	Training dataset	$D_{k\text{-}Train\text{-}1}$ (700)	
	Test dataset	$D_{k\text{-}Test\text{-}1}$ (400)	$D_{k\text{-}Test\text{-}S}$ (200)
Feature extraction (EDM+*SampEn*)	Validation dataset	$D_{k\text{-}Proc\text{-}Valid\text{-}1}$ (400)	$D_{k\text{-}Proc\text{-}Valid\text{-}S}$ (300)
	Training dataset	$D_{k\text{-}Proc\text{-}Train\text{-}1}$ (700)	
	Test dataset	$D_{k\text{-}Proc\text{-}Test\text{-}1}$ (400)	$D_{k\text{-}Proc\text{-}Test\text{-}S}$ (200)

3.2. Feature extraction by EMD and SampEn

The procedure for the proposed feature extraction is illustrated in Fig. 7. The EMD decomposes the acquired signal into n IMFs and residual signal l. After separating the signal $\boldsymbol{x}$ into n IMFs, the dimension of each IMF remains unchanged. In other words, each IMF contains the same number of data points as the signal $\boldsymbol{x}$. So the input dimension of each classifier is very massive (i.e. $n \times N$ data points), resulting in poor fault classification accuracy. To resolve this problem, an effective feature selection method is considered to reduce the input dimension. In this study, an effective statistical algorithm namely Sample Entropy (*SampEn*) is employed to calculate the representative feature from each IMF, such that the input dimension of each classifier is reduced from $n \times N$ to n.

After taking many tests, it is found that the main features of the engine faults are closely related to the first six IMFs. Taking acoustic signal of fault d_1 as an example (Fig. 8), EMD decomposes the signal into nine IMFs and one residual signal. Fig. 8 shows that the *SampEn* values of IMF1 to IMF6 are much higher than those of IMF7 to IMF9. By theory, a large value of *SampEn* means much information content [20], so only the *SampEn* of IMF1 to IMF6 are retained to represent the signal features in this study, and the others are neglected in order to reduce the input dimension to each fault classifier.

3.3. Domain features

After feature extraction by EMD+*SampEn*, domain features are further extracted for the diagnosis as they are important information of a specific field. In the engine ignition signal, *firing voltage*, *burn time* and *average spark voltage* are employed as domain knowledge (DK) features [1,2]. Besides, ten statistical time-domain features are employed as DK features to analyze the engine sound signal based on [31]. Table 6 presents the ten statistical time-domain features. The domain knowledge features of λ are simple, they are $\lambda < 0.9$ implying rich fuel and $\lambda > 1.1$ indicating lean fuel at test engine speeds. Table 7 shows that the sizes of λ, ignition and sound signals are reduced from 8092 data to 16, 18 and 32 features respectively. The procedure of feature extraction greatly reduces the input complexity of the classifiers. To ensure that all the features have even contribution, all reduced features go through normalization within $[-1,1]$. After normalization, a processed dataset $x_{k\text{-}Proc}$ is obtained. The kth committee member can be trained by using $x_{k\text{-}Proc\text{-}TRAIN}$.

4. Experimental results and discussion

Since there are many combinations of the techniques for feature extraction and probabilistic classifier, a set of experiments was carried out to determine the best combination for the proposed framework.

4.1. Results and discussion of various combinations of feature extraction techniques

In the phase of feature extraction, different reasonable combinations of domain knowledge (DK), FFT, wavelet packet transform with principal component analysis (WPT+PCA) [1] and EMD+*SampEn* were tested. For those feature extraction methods, several simple settings are necessary. In the family of mother wavelets, Daubechies wavelet (Db) is the most popular one. Hence, it is employed to compare with other feature extraction techniques. After trying many combinations of the three Daubechies wavelets (Db3, Db4 and Db5) and three levels of decomposition (Levels 3, 4 and 5), Db4 with level 4 decomposition was finally employed to decompose the signals. With PCA, the dimensions of decomposed signals of λ, ignition and sound were further reduced

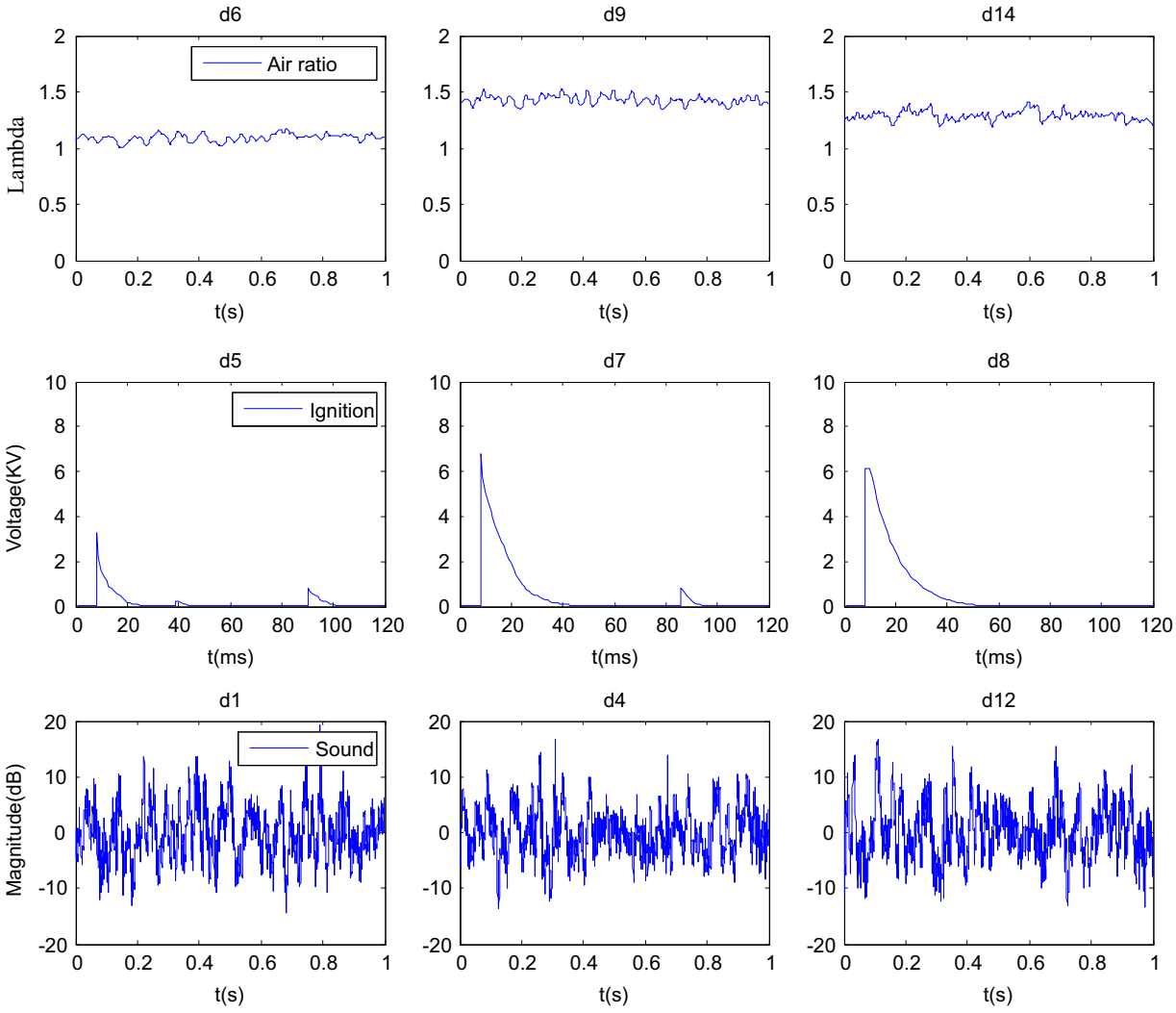


Fig. 6. Examples of faulty engine signals at slow idle speed.

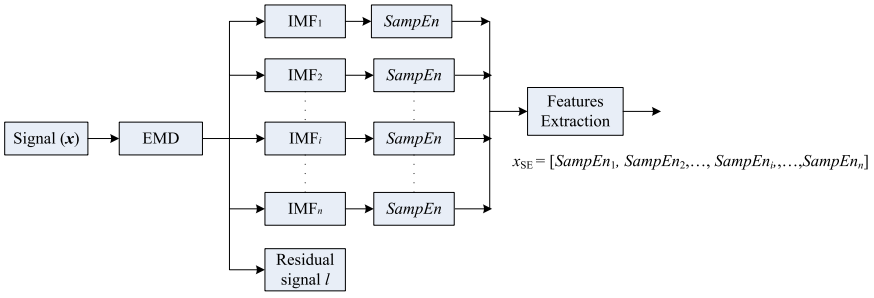


Fig. 7. Procedure for performing feature extraction.

from 8092 inputs to 26, 37 and 50 respectively. In terms of FFT and DK, the feature size of FFT is equal to 8092, and while the DK features of λ, ignition and sound are equal to 2, 3 and 10 respectively. In the EMD+*SampEn* method, two parameters m and r are necessary to be set in *SampEn*. Theoretically, the accuracy of the signal entropy can be improved when the number of matches of length m and $m+1$ increases. The number of matches can be increased by choosing small m and large r. However, if r is too large, some fluctuations of the signal cannot be detected. If r is too small, noise has an impact on the *SampEn* measure. In this study, the values of m and r are set to be 2 and 0.2 respectively based on the suggestion in [21]. The number of extracted features by using EMD+*SampEn* is presented in Table 7.

In this case study, the classification techniques for comparison include pairwise-coupled PNN (PCPNN), pairwise-coupled RVM (PCRVM), SBELM and PCM. In this paper, PCM employs three SBELM as committee members. Moreover, SBELM and each SBELM in PCM use the same number of hidden neurons L. On the contrary, there are two hyper-parameters, spread s and width w^*, which are necessary to be defined in PCPNN and PCRVM respectively. According to the experimental results by using a fivefold cross validation method, s, w^*, and L were set to be 0.4, 0.62, and 50

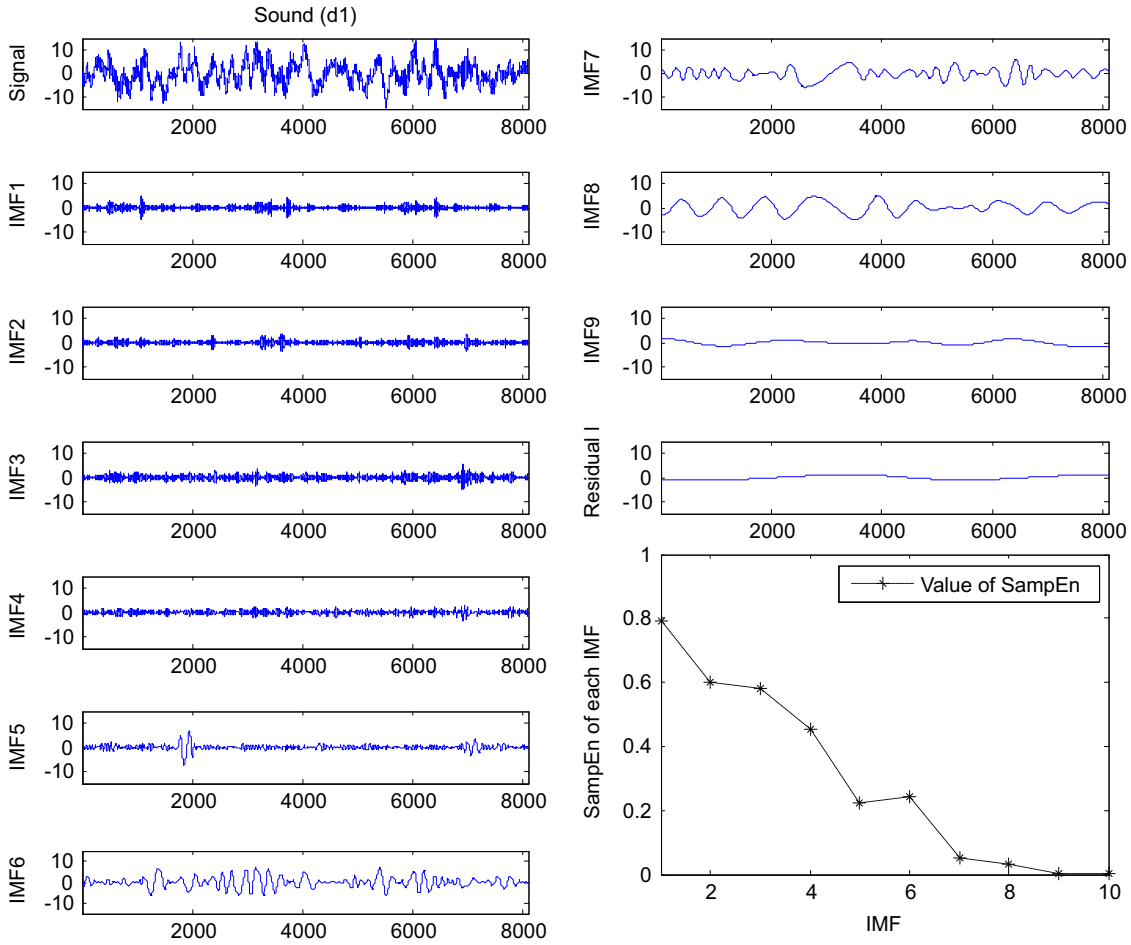


Fig. 8. Example of EMD for decomposing sound signal of d_1 and *SampEn* of each IMF.

Table 6
Definition of common statistical features in time-domain for acoustic signal.

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

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

respectively. Note that these parameters are fixed for the procedure of training and testing.

After determining the configurations of the feature extraction and classification techniques, the reasonable combinations of feature extraction techniques were tested and the results are shown in Table 8, in which the weight of each committee member and threshold ε are predefined as 1 ($w_1=w_2=w_3=1$) and 0.5 respectively. Note that PCPNN, PCRVM, and SBELM determine their *F*-measures by directly combining all features from the three engine signals, λ, ignition and sound as their inputs whereas PCM employs three SBELM committee members to analyze the respective features extracted and make decision by utilizing Eq. (12).

Table 8 illustrates that the feature extraction techniques are effective. Taking PCM as an example, FFT, WPT+PCA, and EMD+*SampEn* techniques respectively give 5.55%, 6.99%, and 8.37% improvement as compared with PCM without any feature extraction. When using the other classifiers, the feature extraction methods can also improve the *F*-measure from 5.39% to 8.58% as compared to those without feature extraction. Besides, the feature extraction techniques with considering the DK features (i.e. WPT+PCA+DK or EMD+*SampEn*+DK) can improve the

diagnostic accuracy by 1% as compared to those without considering the DK features. Table 8 also indicates that no matter which classification technique is, the integration of EMD+*SampEn* and DK as feature extraction approach gives the best accuracy. The reason is that EMD is a self-adaptive time-frequency analysis technique and it decomposes the signal into a serial of IMFs which not only relates to the sampling frequency but also changes with the signal itself. Therefore, it is more suitable for processing the non-linear and non-stationary signal as compared with FFT or WPT. This result verifies the proposed feature extraction technique, EMD+*SampEn*+DK, is the most effective.

4.2. Result and discussion of optimization framework

After selecting the feature extraction technique (EMD+*SampEn*+DK), the extracted features were employed to construct and train the committee machine for further evaluation. This research uses PSO to determine the best w_{opt} of each committee member and decision threshold ε_{opt}. There are five parameters in the PSO that are necessary to be defined. According to the literature [2,9], the number of generation, population size, inertial weight, cognitive parameter and social parameter of PSO were set to be 1000, 100, 0.9, 2, and 2 respectively. In the initial phase of PSO, the

Table 7
Number of extracted features in different signal types.

Engine signal type	EMD+*SampEn*	+DK	×Two testing conditions	=Total feature obtained
λ signal	6	2	2	2×(6+2)=16
Ignition signal	6	3	2	2×(6+3)=18
Sound signal	6	10	2	2×(6+10)=32

Table 8
Diagnostic accuracies of different combinations of techniques.

Feature extraction	Classification technique			
	PCPNN F_{me} (%) based on simultaneous-fault test set	PCRVM F_{me} (%) based on simultaneous -fault test set	SBELM F_{me} (%) based on simultaneous-fault test set	PCM F_{me} (%) based on simultaneous-fault test set
None	65.10	65.34	66.67	69.74
FFT	70.78	71.26	72.06	75.29
WPT+PCA	71.92	72.53	73.65	76.73
WPT+PCA+DK	72.35	73.41	74.79	77.87
EMD+*SampEn*	73.47	73.73	75.25	78.11
EMD+*SampEn*+DK	**74.53**	**74.65**	**76.38**	**79.64**

Table 9
Accuracy comparsion of optimized weights and threshold versus empirical threshold and equal weights.

Classifier	No. of features	Method for determining threshold and weights	Threshold	Weights	F_{me} (%) based on simultaneous-fault test set
PCM	λ=16 Ignition=18 Sound=32	Proposed optimization framework	0.7272	w_1=0.1867 w_2=**0.8649** w_3=0.2814	81.49
PCM	λ=16 Ignition=18 Sound=32	Empirical threshold and equal weights	0.5	w_1=1 w_2=1 w_3=1	79.64

Remark: Feature extraction method is based on the proposed method (EMD+SampEn+DK).

Table 10
Evaluation results of proposed PCM framework with PCPNN, PCRVM and SBELM.

Classifier	Feature number	Optimal decision threshold	Optimal weight	Accuracies for test cases F_{me} (%)		
				Single-faults	Simultane-ous faults	Overall-faults (i.e. Single+simultaneous faults)
PCPNN	16+18+32=66	0.6875	–	88.41	75.63	80.00
PCRVM	16+18+32=66	0.7056	–	90.56	77.54	82.21
SBELM	16+18+32=66	0.7102	–	90.87	78.77	83.23
PCM	λ=16 Ignition=18 Sound=32	0.7272	w_1=0.1867 w_2=0.8649 w_3=0.2814	92.00	81.49	86.17

Remark: Feature extraction method is based on the proposed method (EMD+SampEn+DK).

weight (w_1, w_2, w_3) and decision threshold were randomly selected from the interval [0 1].

In this study, the empirical constraint equation, Eq. (15), is employed as the objective function to avoid over-fitting on the validation dataset. After 1000 generations, the optimal weight and threshold are obtained based on Eq. (15) and the validation data set. The optimal weights and decision threshold obtained are then used to further exam the diagnostic accuracy of the proposed diagnostic framework. The optimized weights and threshold as well as F_{me} of the proposed diagnostic framework are shown in Table 9 in which the optimal weight for the second committee member w_2 (0.8649) is higher than w_1 and w_3. In other words, the committee member trained by ignition patterns shows a great influence on the simultaneous-fault diagnosis. The main reason can be explained in Table 3 in which many components of the simultaneous-faults are from the single-faults associated with the ignition signal. In this way, the second committee member is assigned a heavy weight by PSO in order to make the overall output easily satisfy the objective function. Table 9 also illustrates that the proposed optimization framework can improve the diagnostic accuracy by 2.3% as compared with the empirical decision threshold of 0.5 and equal weights ($w_1=w_2=w_3=1$), under the same feature extraction technique and simultaneous-fault test datasets. In other words, the proposed optimization framework is effective.

4.3. Evaluation of the proposed PCM framework

To verify the effectiveness of the proposed PCM diagnostic framework, the aforesaid three single probabilistic classifiers are introduced to compare with proposed framework based on their decision thresholds obtained by PSO. The experimental results of average *F*-measures are shown in Table 10. The overall accuracy of the proposed PCM framework outperforms PCPNN, PCRVM and SBELM by 6.17%, 3.96% and 2.94% respectively under the combination of single and simultaneous faults. Note that all the classifiers are only trained by single-fault patterns while the overall performance is evaluated over both single-fault and simultaneous-faults test patterns, so it is not surprised that the *F*-measures in all simultaneous fault detections are not very high. Table 10 also reveals that the proposed diagnostic framework achieves the best accuracies for both single fault (92%) and simultaneous faults (81.49%). The main reason is that each SBELM committee member in the proposed framework is trained with a different engine signal. In this way, each committee member becomes different from each other, which can improve the classification accuracy under a well-accepted concept, group decision. For example, considering an ensemble with M trained classifiers $\{C_1,\ldots,C_k,\ldots, C_M\}$, if the classifies are trained using different subsets and their errors are uncorrelated, even when C_k is wrong, the other classifiers may still be correct. Besides, Table 10 also indicates that the overall diagnostic accuracy of SBELM is superior to PCPNN and PCRVM. It is because SBELM is insensitive to its hyperparameter, number of hidden neurons, while PCPNN and PCRVM are very sensitive to their hyperparameters. In this case study, the hyperparameters for PCPNN and PCRVM may not be optimal. In fact, the optimal hyperparameters for both PCPNN and PCRVM are very difficult to be determined. Moreover, PNN cannot handle too many input data points, so it is not surprising to see that the accuracy of PCPNN is the worst.

In this study, the proposed PCM framework is a promising approach to detect simultaneous-faults without costly simultaneous-fault training patterns. Moreover, the proposed method employs three kinds of engine signals (λ, ignition and sound) to train the diverse committee members, which ensures the diagnostic result to be more reliable. Therefore, it can be concluded that the proposed framework is an effective technique to overcome the existing challenges of the automotive engine simultaneous-fault diagnosis.

5. Conclusions

In this paper, simultaneous-fault diagnosis of automotive engine based on three kinds of engine signals is studied. A novel framework combining feature extraction, probabilistic committee machine, and decision threshold optimization based on a fair assessment, *F*-measure, has successfully been developed to overcome the challenges of simultaneous-fault diagnosis and multi-signal analysis in automotive engines. In the proposed framework, the feature extraction technique is designed by combining the EMD+*SampEn*+DK to effectively capture single-fault components from the simultaneous-fault patterns. It implies that the acquisition of the large amount of simultaneous-fault signals can be avoided. Moreover, PSO is employed to optimize the two main kinds of parameters in the framework, weight of every committee member and decision threshold.

To verify the effectiveness of the proposed probabilistic committee machine, the single probabilistic classifiers, PCPNN, PCRVM, and SBELM, are also evaluated to diagnose the simultaneous-faults. Experimental results show that the proposed framework is superior to the three single classifiers, and is able to diagnose both single and simultaneous-faults while it is only trained with single fault data. Therefore, the proposed framework is suitable to analyze the simultaneous-fault and multi-signal problems. As the proposed framework is generic, it can be applied to different kinds of vehicle engines and similar engineering problems. The research originalities are listed as follows:

(1) A new PCM diagnostic framework is proposed.
(2) The proposed PCM is the first in the literature that simultaneously uses three kinds of engine signal patterns, including air ratio, ignition and engine sound signals, to train various strength of committee members, so that the diagnostic accuracy can be enhanced and more engine faults can be detected while the traditional engine diagnostic systems cannot achieve because they are single member systems and work with one type of engine signal only.
(3) Since the SBELM proposed by the authors in [26] cannot be directly applied to the PCM diagnostic framework, this project develops a new kind of classifier based on the previous version of SBELM, namely, pairwise-coupled SBELM. Its structure differs from the author's previous work in [26] in term of pairwise-coupling strategy. The proposed SBELM can train the classifier effectively, generate the probability of each fault accurately and let the users define its structure easily. These attractive features enable the proposed SBELM to be superior to the traditional classifiers, such as relevance vector machine and probabilistic neural network.
(4) It is the first attempt at applying probabilistic classifiers as committee members so that engineers can use it to trace the other possible faults according to the rank of their probabilities once the predicted fault(s) from the classifier is incorrect in the problem.
(5) A novel ensemble method based on optimal weights and predefined null outputs and weights is developed to deal with the problem of diverse committee members with different abilities (i.e. various engine signals will have different sensitivities to each kind of fault). As a result, a reliable diagnostic result can be obtained from the outputs of various committee members.

Acknowledgments

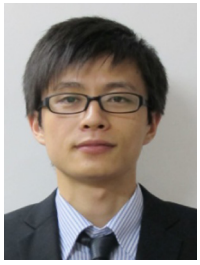
The authors would like to thank the financial support from the University of Macau, Grant numbers: MYRG2014-00178-FST and

MYRG075(Y1-L2)–FST12-VCM, and the Science and Technology Development Fund of Macau S.A.R., Grant number FDCT/075/2013/A. The authors would also like to thank the technical support from Hang Cheong Wong and Siyu Jia.

References

[1] C.M. Vong, P.K. Wong, W.F. Ip, A new framework of simultaneous-fault diagnosis using pairwise probabilistic multi-label classification for time-dependent patterns, IEEE Trans. Ind. Electron. 60 (2013) 3372–3385.
[2] C.M. Vong, P.K. Wong, W.F. Ip, C.C. Chiu, Simultaneous-fault diagnosis of automotive engine ignition systems using prior domain knowledge and relevance vector machine, Math. Probl. Eng. 2013 (2013), pp. 19, Article ID 974862.
[3] O. Basir, X. Yuan, Engine fault diagnosis based on multi-sensor information fusion using Dempster–Shafer evidence theory, Inf. Fusion 8 (2007) 379–386.
[4] P.L. Hsu, K.L. Lin, L.C. Shen, Diagnosis of multiple sensor and actuator failures in automotive engines, IEEE Trans. Veh. Technol. 44 (1995) 779–789.
[5] K. Choi, S. Singh, A. Kodali, K.R. Pattipati, J.W. Sheppard, S.M. Namburu, et al., Novel classifier fusion approahces for fault diagnosis in automotive systems, IEEE Trans. Instrum. Meas. 58 (2007) 602–611.
[6] J. Mohammadpour, M. Franchek, K. Grigoriadis, A survey on diagnostics methods for automotive engines, Int. J. Engine Res. 13 (2012) 41–46.
[7] A. Azarian, A. Siadat, A global modular framework for automotive diagnosis, Adv. Eng. Inform. 26 (2012) 131–144.
[8] M. Nyberg, Model-based diagnosis of an automotive engine using several types of fault models, IEEE Trans. Control Syst. Technol. 10 (2002) 679–689.
[9] P.K. Wong, L.M. Tam, K. Li, C.M. Vong, Engine idle-speed system modelling and control optimization using artificial intelligence, Proc. Inst. Mech. Eng. Part D: J. Automob. Eng. 224 (2010) 55–72.
[10] H.L. Gelgele, K. Wang, An expert system for engine fault diagnosis: development and application, J. Intell. Manuf. 9 (1998) 539–545.
[11] Y. He, C. Rutland, Application of artificial neural networks in engine modelling, Int. J. Engine Res. 5 (2004) 281–296.
[12] B. Samanta, Gear fault detection using artificial neural networks and support vector machines with genetic algorithms, Mech. Syst. Signal Process. 18 (2004) 625–644.
[13] A. Widodo, B.S. Yang, Application of nonlinear feature extraction and support vector machines for fault diagnosis of induction motors, Expert Syst. Appl. 33 (2007) 241–250.
[14] Z. Yang, P.K. Wong, C.M. Vong, J. Zhong, J. Liang, Simultaneous-fault diagnosis of gas turbine generator systems using a pairwise-coupled probabilistic classifier, Math. Probl. Eng. 2013 (2013), pp. 14, Article ID 827128.
[15] R. Eslamloueyan, M. Shahrokhi, R. Bozorgmehri, Multiple simultaneous fault diagnosis via hierarchical and single artificial neural networks, Sci. Iran. 10 (2003) 300–310.
[16] G. Zheng, A. Leung, Internal combustion engine noise analysis with time–frequency distribution, J. Eng. Gas Turbines Power 124 (2002) 645–649.
[17] S.J. Loutridis, Damage detection in gear systems using empirical mode decomposition, Eng. Struct. 26 (2004) 1833–1841.
[18] N.E. Huang, Z. Shen, S.R. Long, M.C. Wu, H.H. Shih, Q. Zheng, et al., The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis, Proc. R. Soc. London Ser. A: Math. Phys. Eng. Sci. 454 (1998) 903–995.
[19] Y. Li, P.W. Tse, X. Yang, J. Yang, EMD-based fault diagnosis for abnormal clearance between contacting components in a diesel engine, Mech. Syst. Signal Process. 24 (2010) 193–210.
[20] J.S. Richman, J.R. Moorman, Physiological time-series analysis using approximate entropy and sample entropy, Am. J. Physiol. – Heart Circ. Physiol. 278 (2000) 2039–2049.
[21] Y. Song, P. Liò, A new approach for epileptic seizure detection: sample entropy based feature extraction and extreme learning machine, J. Biomed. Sci. Eng. 3 (2010) 556–567.
[22] J.D. Wu, C.Q. Chuang, Fault diagnosis of internal combustion engines using visual dot patterns of acoustic and vibration signals, NDT & E Int. 38 (2005) 605–614.
[23] J.D. Wu, C.H. Liu, An expert system for fault diagnosis in internal combustion engines using wavelet packet transform and neural network, Expert Syst. Appl. 36 (2009) 4278–4286.
[24] S. Chen, W. Wang, H. Van Zuylen, Construct support vector machine ensemble to detect traffic incident, Expert Syst. Appl. 36 (2009) 10976–10986.
[25] V. Tresp, A Bayesian committee machine, Neural Comput. 12 (2000) 2719–2741.
[26] J. Luo, C.M. Vong, P.K. Wong, Sparse Bayesian extreme learning machine for multi-classification, IEEE Trans. Neural Netw. Learn. Syst. 25 (2014) 836–843.
[27] S. Abe, Support Vector Machines for Pattern Classification, 2nd ed., Springer, London, UK, 2010.
[28] T.F. Wu, C.J. Lin, R.C. Weng, Probability estimates for multi-class classification by pairwise coupling, J. Mach. Learn. Res. 5 (2004) 975–1005.
[29] J. Kennedy, Particle Swarm Optimization, Springer-Verlag, USA, 2006.
[30] J. Robinson, Y. Rahmat-Samii, Particle swarm optimization in electromagnetics, IEEE Trans. Antennas Propag. 52 (2004) 397–407.
[31] A. Widodo, E.Y. Kim, J.-D. Son, B.-S. Yang, A.C. Tan, D.-S. Gu, et al., Fault diagnosis of low speed bearing based on relevance vector machine and support vector machine,, Expert Syst. Appl. 36 (2009) 7252–7261.
[32] G.B. Huang, X. Ding, H. Zhou, Optimization method based extreme learning machine for classification, Neurocomputing 74 (2010) 155–163.
[33] G.B. Huang, H. Zhou, X. Ding, R. Zhang, Extreme learning machine for regression and multiclass classification,, IEEE Trans. Syst. Man Cybern. Part B: Cybern. 42 (2012) 513–529.
[34] C.M. Bishop, Pattern Recognition and Machine Learning, springer, New York, 2006.
[35] D.J. MacKay, Bayesian Methods for Backpropagation Networks, Springer, USA, 1996.
[36] R. Baeza-Yates, B. Ribeiro-Neto, Modern information retrieval, ACM press, New York, 1999.
[37] I. Yélamos, M. Graells, L. Puigjaner, G. Escudero, Simultaneous fault diagnosis in chemical plants using a multilabel approach, Am. Inst. Chem. Eng. J. 53 (2007) 2871–2884.

Pak-Kin Wong received the Ph.D. degree in Mechanical Engineering from The Hong Kong Polytechnic University, Hong Kong, in 1997. He is currently the Head of Department of Electromechanical Engineering and Associate Dean (Academic Affairs), Faculty of Science and Technology, University of Macau. His research interests include automotive engineering, fluid transmission and control, engineering applications of artificial intelligence, and mechanical vibration. He has published over 166 scientific papers in refereed journals, book chapters, and conference proceedings.



Jianhua Zhong received the M.S. degree in Electromechanical Engineering from the University of Macau, Macao, in 2011.

He is currently pursuing a Ph.D. degree in Electromechanical Engineering at the University of Macau. His research interests include condition monitoring and fault diagnosis using machine learning methods.

Zhixin Yang obtained his BEng in Mechanical Engineering from the Huazhong University of Science and Technology, China in 1992, and Ph.D. in Industrial Engineering and Engineering Management from the Hong Kong University of Science and Technology in 2000.

He is currently an assistant professor in the Department of Electromechanical Engineering, Faculty of Science and Technology, University of Macau. His research areas include design reuse, engineering applications of artificial intelligence, and manufacturing exection system.

Chi-Man Vong received the M.S. and Ph.D. degrees in Software Engineering from the University of Macau in 2000 and 2005, respectively.

He is currently an Associate Professor with the Department of Computer and Information Science, Faculty of Science and Technology, University of Macau. His research interests include machine learning methods and intelligent systems.