Gaussian Mixture Model Using Semisupervised Learning for Probabilistic Fault Diagnosis Under New Data Categories

Heng-Chao Yan, *Student Member, IEEE*, Jun-Hong Zhou, *Member, IEEE*, and Chee Khiang Pang, *Senior Member, IEEE*

***Abstract*— Fault diagnosis has played a vital role in industry to prevent operation hazards and failures. To overcome the limitation of conventional diagnosis approaches, which misclassify new types of faults into existing categories from training, a novel probabilistic diagnosis framework will be proposed in this paper for effective detection on new data categories. Gaussian mixture model (GMM) is applied for the pattern recognition, while its training procedure is improved from conventional unsupervised learning to novel semisupervised learning. Even with unlabeled training data, component number in our GMM can be autoselected instead of predetermined. For online testing, the probabilistic classification results from GMM's soft assignment assist to improve overall diagnosis framework, which is able to first detect whether new types of faults occur and further categorize them in detail via the GMM update. The effectiveness of our fault diagnosis framework is testified on an industrial fault simulator of rotary machine and the partial discharge measurement of various high-voltage electronic equipment components. Compared with existing approaches, our probabilistic diagnosis framework is able to achieve an average diagnosis accuracy of 97.9% without new data categories and it can also classify new data categories with diagnosis accuracy of at least 86.3% if occurred.**

***Index Terms*— Akaike information criterion, fault diagnosis, Gaussian mixture model, new data category, semisupervised learning, soft assignment.**

I. Introduction

EFFECTIVE fault diagnosis on realistic engineering systems can assist to timely schedule maintenance and guarantee continuous manufacturing in industry [1]–[5]. In general, fault diagnosis includes fault detection, fault location, and fault interpretation [6]–[8]. With the support of senor monitoring and signal processing technologies, sensor-based fault diagnosis has received much attention nowadays [9]–[11], which depends on condition monitoring (CM) information to evaluate the real-time equipment health condition. Methodologies for fault diagnosis can be divided into two main categories, namely, physics-based and data-driven methodologies [6]. Physics-based methodologies analyze equipment structures for the explicit mathematic modeling to diagnose the corresponding faults. Due to structure complexities in reality, they may be too complicated to be widely applied in practice. With fewer requirements on sophisticate expertise, data-driven methodologies are explored, developed, and implemented currently. Data-driven methodologies aim at the pattern recognition from the high-dimensional CM space to the low-dimensional fault detecting space.

Data-driven methodology includes statistical techniques like Gaussian mixture model (GMM) [12]–[14], K-means [15], and Bayesian framework [16]–[18] as well as machine learning techniques like neural network (NN), support vector machine, and fuzzy logic [19]–[24]. Most of the above diagnosis approaches execute the hard assignment, which classifies testing data sample into a certain category and outputs the scalar classification result. In complicated systems especially with multiple components, one logged data sample may share characteristics among several fault categories. In this case, hard assignment is not sufficient to fully reflect realistic situations. To overcome limitations of hard assignment, soft assignment calculates the classification probability for each existing category and uses above confidence scores in driving the probabilistic result. It could provide more classification information for maintenance engineers to investigate all possible faulty components and reduce misclassification consequences. As such, GMM as a popular soft assignment approach is motivated for the probabilistic fault diagnosis in this paper.

It is worth noting that conventional GMM still exist three main limitations for practical applications, namely, inflexible unsupervised learning, component number predetermined for training, and inability to detect new types of faults. According to expectation maximization (EM) algorithm [25]–[29] and unlabeled training data, the unsupervised learning is commonly chosen to train GMM. In reality, a limited number of labeled data samples could be also obtained from historical operation. However, conventional unsupervised learning has to discard their observation information about fault categories. As such, to fully utilize all training information from both labeled and unlabeled data, the semisupervised learning has been

Manuscript received March 13, 2016; revised July 28, 2016; accepted October 24, 2016. Date of publication February 8, 2017; date of current version March 8, 2017. The Associate Editor coordinating the review process was Dr. John Sheppard.

H.-C. Yan and C. K. Pang are with the Department of Electrical and Computer Engineering, National University of Singapore, Singapore 117583 (e-mail: justinpang@nus.edu.sg).

J.-H. Zhou is with the Manufacturing Execution and Control Group, A*STAR Singapore Institute of Manufacturing Technology, Singapore 638075.

C. K. Pang is also with the Engineering Cluster, Singapore Institute of Technology, Singapore 138682

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TIM.2017.2654552

explored in [30]–[33]. For instance, semisupervised learning was transferred into a convex optimization problem about the likelihood of training data and cluster purity in [34]. However, most of them focus on semisupervised learning for machine learning techniques. One semisupervised learning specific for GMM should be discussed and proposed.

Conventionally, the number of Gaussian components in GMM has to be predetermined based on the experience and expertise before training. An inappropriate selection may weaken or even block the training procedure. Due to the existence of unlabeled training data, it may be difficult to determine a proper value of component number in reality. As such, an effective autoselection strategy on component number requires to be applied. Some research literature has discussed this issue. The concept of hierarchical clustering was applied in [35] to reduce a large Gaussian mixture into smaller mixture. Component number selection and online parameter estimation were also discussed in [36]–[38]. However, requirements on large-scale data and heavy computation cost are two common weaknesses. An easy but effective strategy is still desired during training. In addition, although many diagnosis approaches are able to achieve acceptable classification performance on existing data categories, new types of faults may occur in future. In this case, most of the current diagnosis approaches including conventional GMM cannot know the total number of all fault types in advance for training, which may misclassify data samples from new categories into existing types of faults defined from the training data. Actually, there exist a few pioneering works to handle new types of faults detection in current literature [39]–[42], while most of them are from the view of signal processing to find out signal differences for new category detection. As such, each of them was only limited to the corresponding specific field such as a dc motor. To the best of our understanding, a general framework for new types of faults diagnosis is still a challenge that requires attention.

To mainly address above three issues from conventional GMM, a novel probabilistic fault diagnosis framework is proposed in this paper. GMM is applied as a pattern recognition algorithm, while the semisupervised learning is mathematically derived to estimate its parameters based on both labeled and unlabeled data. During training, Akaike information criterion (AIC) is calculated to evaluate the modeling performance and autoselect the component number of GMM. For testing, the probabilistic classification result from GMM's soft assignment is applied to improve overall framework for new types of faults diagnosis via two steps. The practicality of the considered problem and the feasibility of our proposed diagnosis framework are verified on an industrial fault simulator of rotary machine as well as the partial discharge (PD) measurement of various high-voltage electronic and power equipment components, respectively.

The rest of this paper is organized as follows. Section II reviews the GMM preliminaries, where one novel semisupervised learning specific for GMM training is deduced. Next, our probabilistic fault diagnosis framework under new data categories is presented in Section III. Section IV displays two industrial case studies on an industrial fault simulator of rotary machine as well as the PD measurement of high-voltage electronic equipment. Our conclusion and future work directions are presented in Section V.

II. Semisupervised Learning for Gaussian Mixture Model

GMM is briefly reviewed in this section. Next, its semisupervised learning is mathematically deduced from the conventional supervised and unsupervised learning.

A. Gaussian Mixture Model

GMM is a weighted linear sum of K component Gaussian densities to provide a richer class of density models [27]. For any D-dimensional input $\mathbf{x} \in \mathbb{R}^D$, the output of probability density functions from one GMM is calculated as

$$\Pr(\mathbf{x}) = \sum_{k=1}^{K} \pi_k \mathcal{N}(\mathbf{x}; \boldsymbol{\mu}_k, \Sigma_k) \tag{1}$$

where π_k is the weight of the k^{th} Gaussian component subject to $\sum_{k=1}^{K} \pi_k = 1$ and $0 \leq \pi_k \leq 1$ for $k = 1, 2, \ldots, K$. $\mathcal{N}(\mathbf{x}; \boldsymbol{\mu}_k, \Sigma_k)$ in (1) is defined as

$$\mathcal{N}(\mathbf{x}; \boldsymbol{\mu}_k, \Sigma_k) = \frac{1}{(2\pi)^{D/2} |\Sigma_k|^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_k)^T \Sigma_k^{-1} (\mathbf{x} - \boldsymbol{\mu}_k)\right] \tag{2}$$

where $\boldsymbol{\mu}_k \in \mathbb{R}^D$ is the mean vector of the k^{th} Gaussian component and $\Sigma_k \in \mathbb{R}^{D \times D}$ is its corresponding covariance matrix with the invertible constraint.

Since GMM executes the soft assignment to handle classification issues, for a GMM with K Gaussian components, the original confidence score of input $\mathbf{x}$ classified into Category k is defined as

$$\Pr(y_k = 1|\mathbf{x}) = \pi_k \mathcal{N}(\mathbf{x}; \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k). \tag{3}$$

After the probability normalization, the final confidence score, namely, the normalized classification probability, into Category k will be $(\Pr(y_k = 1|\mathbf{x})/\Pr(\mathbf{x}))$ for $k = 1, 2, \ldots, K$.

GMM uses above final confidence scores in deriving probabilistic classification result of $\mathbf{y} = [y_1, y_2, \ldots, y_K]^T \in \mathbb{R}^K$, in which only $y_k = 1$ is with $k = \arg\max \Pr(y_k = 1|\mathbf{x})$ and the classification probability of $((\Pr(y_k = 1|\mathbf{x}))/\Pr(\mathbf{x}))$. For simplicity without loss of rationality, the resting terms are equal to zero.

B. Proposal of Novel Semisupervised Learning

In general, unsupervised learning is applied to train GMM [26]–[29]. To fully utilize all training information from both labeled and unlabeled data, one novel semisupervised learning via EM algorithm is proposed in this section, which is specific for GMM training. For one unlabeled training data of $\mathbf{U} = [\mathbf{u}^1, \mathbf{u}^2, \ldots, \mathbf{u}^N] \in \mathbb{R}^{D \times N}$ and one labeled training data of $\mathbf{X} = [\mathbf{x}^1, \mathbf{x}^2, \ldots, \mathbf{x}^M] \in \mathbb{R}^{D \times M}$ with its corresponding observation $\mathbf{Y} = [\mathbf{y}^1, \mathbf{y}^2, \ldots, \mathbf{y}^M] \in \mathbb{R}^{K \times M}$, a semisupervised coefficient β is introduced under $0 \leq \beta \leq 1$ [30]. It targets to weighted integrate the supervised and unsupervised learning for parameter estimation.

In our proposed semisupervised learning, GMM parameters will be obtained via an EM algorithm in four steps.

1) *Step 1:* Initialize π_k, $\boldsymbol{\mu}_k$, and Σ_k in the corresponding Gaussian densities for $k = 1, 2, \ldots, K$.
2) *Step 2:* Calculate $\gamma(n,k)$ via

$$\gamma(n,k) = \frac{\pi_k \mathcal{N}(\mathbf{u}^n; \boldsymbol{\mu}_k, \Sigma_k)}{\sum_{i=1}^{K} \pi_i \mathcal{N}(\mathbf{u}^n; \boldsymbol{\mu}_i, \Sigma_i)} \tag{4}$$

for $n = 1, 2, \ldots, N$.
3) *Step 3:* With $\gamma(n,k)$ from Step 2, update π_k, $\boldsymbol{\mu}_k$, and Σ_k via

$$\pi_k = \frac{(1-\beta)\sum_{n=1}^{N}\gamma(n,k) + \beta\sum_{m=1}^{M} y_k^m}{(1-\beta)N + \beta M} \tag{5}$$

$$\boldsymbol{\mu}_k = \frac{(1-\beta)\sum_{n=1}^{N}\gamma(n,k)\mathbf{u}^n + \beta\sum_{m=1}^{M} y_k^m \mathbf{x}^m}{(1-\beta)\sum_{n=1}^{N}\gamma(n,k) + \beta\sum_{m=1}^{M} y_k^m} \tag{6}$$

and

$$\Sigma_k = \frac{(1-\beta)\sum_{n=1}^{N}\gamma(n,k)(\mathbf{u}^n - \boldsymbol{\mu}_k)(\mathbf{u}^n - \boldsymbol{\mu}_k)^T}{(1-\beta)\sum_{n=1}^{N}\gamma(n,k) + \beta\sum_{m=1}^{M} y_k^m} + \frac{\beta\sum_{m=1}^{M} y_k^m (\mathbf{x}^m - \boldsymbol{\mu}_k)(\mathbf{x}^m - \boldsymbol{\mu}_k)^T}{(1-\beta)\sum_{n=1}^{N}\gamma(n,k) + \beta\sum_{m=1}^{M} y_k^m} \tag{7}$$

for $m = 1, 2, \ldots, M$, in which each observation is $\mathbf{y}^m = [y_1^m, y_2^m, \ldots, y_K^m]^T \in \mathbb{R}^K$.
4) *Step 4:* Repeat Step 2 and Step 3 until all π_k, $\boldsymbol{\mu}_k$, and Σ_k are converged for $k = 1, 2, \ldots, K$, or the maximum iteration times is met if applicable.

The following are mathematical derivations of (5)–(7). With support of a semisupervised coefficient β, the corresponding likelihood function on $\mathbf{U}$, $\mathbf{X}$, and $\mathbf{Y}$ is defined as

$$L(\boldsymbol{\pi}, \boldsymbol{\mu}, \Sigma) = L(\mathbf{U}; \boldsymbol{\pi}, \boldsymbol{\mu}, \Sigma)^{1-\beta} L(\mathbf{X}, \mathbf{Y}; \boldsymbol{\pi}, \boldsymbol{\mu}, \Sigma)^{\beta} \tag{8}$$

where $L(\mathbf{U}; \boldsymbol{\pi}, \boldsymbol{\mu}, \Sigma) = \prod_{n=1}^{N}\sum_{k=1}^{K} \pi_k \mathcal{N}(\mathbf{u}^n; \boldsymbol{\mu}_k, \Sigma_k)$ [25] and $L(\mathbf{X}, \mathbf{Y}; \boldsymbol{\pi}, \boldsymbol{\mu}, \Sigma) = \prod_{m=1}^{M}\prod_{k=1}^{K} [\pi_k \mathcal{N}(\mathbf{x}^m; \boldsymbol{\mu}_k, \Sigma_k)]^{y_k^m}$ [26]. It is worth noting that (8) is the unsupervised learning when $\beta = 0$, while (8) is supervised learning when $\beta = 1$.

The log-likelihood function of (8) is rewritten as

$$\begin{aligned}\ln L(\boldsymbol{\pi}, \boldsymbol{\mu}, \Sigma) &= (1-\beta)\sum_{n=1}^{N} \ln\left[\sum_{k=1}^{K} \pi_k \mathcal{N}(\mathbf{u}^n; \boldsymbol{\mu}_k, \Sigma_k)\right] \\ &\quad + \beta\sum_{m=1}^{M}\sum_{k=1}^{K} y_k^m [\ln \pi_k + \ln \mathcal{N}(\mathbf{x}^m; \boldsymbol{\mu}_k, \Sigma_k)].\end{aligned} \tag{9}$$

To guarantee $\sum_{k=1}^{K} \pi_k = 1$ in GMM, two Lagrange multipliers λ and η [25], [26] are introduced into (9). According to the maximum likelihood estimation (MLE) algorithm for parameter estimation, the first partial derivative of (9) with λ and η respect to π_k should be equal to zero, which is obtained as

$$\begin{aligned}&\frac{\partial}{\partial \pi_k}\left\{(1-\beta)\left[\ln L(\mathbf{U}; \boldsymbol{\pi}, \boldsymbol{\mu}, \Sigma) + \lambda\left(\sum_{k=1}^{K}\pi_k - 1\right)\right]\right. \\ &\quad \left. + \beta\left[\ln L(\mathbf{X}, \mathbf{Y}; \boldsymbol{\pi}, \boldsymbol{\mu}, \Sigma) + \eta\left(\sum_{k=1}^{K}\pi_k - 1\right)\right]\right\} \\ &= (1-\beta)\left[\sum_{n=1}^{N}\frac{\gamma(n,k)}{\pi_k} + \lambda\right] + \beta\left[\sum_{m=1}^{M}\frac{y_k^m}{\pi_k} + \eta\right] = 0\end{aligned} \tag{10}$$

for $k = 1, 2, \ldots, K$.

We multiply both sides of (10) by π_k and sum over k to obtain $(1-\beta)(N + \sum_{k=1}^{K}\pi_k\lambda) + \beta(M + \sum_{k=1}^{K}\pi_k\eta) = 0$. To make use of constraint $\sum_{k=1}^{K}\pi_k = 1$ under both the supervised learning of $\beta = 1$ and the unsupervised learning of $\beta = 0$, it can be found that $\lambda = -N$ and $\eta = -M$.

Next, according to the MLE algorithm for parameter estimation, π_k can be solved from (10) such that $(1-\beta)[\sum_{n=1}^{N}(\gamma(n,k)/\pi_k) + N] + \beta[\sum_{m=1}^{M}(y_k^m/\pi_k) + M] = 0$ as

$$\pi_k = \frac{(1-\beta)\sum_{n=1}^{N}\gamma(n,k) + \beta\sum_{m=1}^{M} y_k^m}{(1-\beta)N + \beta M}. \tag{11}$$

Due to $(\partial \mathbf{b}^T \mathbf{A}\mathbf{b}/\partial \mathbf{b}) = \mathbf{A}\mathbf{b} + \mathbf{A}^T\mathbf{b}$ with one square matrix $\mathbf{A}$ and vector $\mathbf{b}$, the first partial derivative of (9) respect to $\boldsymbol{\mu}_k$ is deduced to solve $\boldsymbol{\mu}_k$ as

$$(1-\beta)\sum_{n=1}^{N}\gamma(n,k)\Sigma_k^{-1}(\mathbf{u}^n - \boldsymbol{\mu}_k) + \beta\sum_{m=1}^{M} y_k^m \Sigma_k^{-1}(\mathbf{x}^m - \boldsymbol{\mu}_k) = 0. \tag{12}$$

Next, $\boldsymbol{\mu}_k$ is solved from (12) as

$$\boldsymbol{\mu}_k = \frac{(1-\beta)\sum_{n=1}^{N}\gamma(n,k)\mathbf{u}^n + \beta\sum_{m=1}^{M} y_k^m \mathbf{x}^m}{(1-\beta)\sum_{n=1}^{N}\gamma(n,k) + \beta\sum_{m=1}^{M} y_k^m}. \tag{13}$$

In addition, for $(\partial \mathbf{b}^T \mathbf{A}\,\mathbf{b}/\partial \mathbf{A}) = \mathbf{b}\mathbf{b}^T$ with constant vector $\mathbf{b}$, $\det(\mathbf{A}) = (1/\det(\mathbf{A}^{-1}))$, and $(\partial \det(\mathbf{A})/\partial \mathbf{A}) = (\mathbf{A}^T)^{-1}\det(\mathbf{A})$ with an invertible square matrix $\mathbf{A}$ [25], we can obtain preliminaries of $\det(\boldsymbol{\Sigma}_k^{-1})^{1/2} = (1/\det(\boldsymbol{\Sigma}_k)^{1/2})$ and $(\partial \det(\boldsymbol{\Sigma}_k^{-1})^{1/2}/\partial \boldsymbol{\Sigma}_k^{-1}) = (1/2)\det(\boldsymbol{\Sigma}_k^{-1})^{1/2}\boldsymbol{\Sigma}_k$.

According to the MLE algorithm for parameter estimation to solve Σ_k, the first partial derivative of (9) respect to Σ_k^{-1} is deduced as

$$\frac{(1-\beta)}{2}\sum_{n=1}^{N}\gamma(n,k)[\Sigma_k - (\mathbf{u}^n - \boldsymbol{\mu}_k)(\mathbf{u}^n - \boldsymbol{\mu}_k)^T] + \frac{\beta}{2}\sum_{m=1}^{M} y_k^m [\Sigma_k - (\mathbf{x}^m - \boldsymbol{\mu}_k)(\mathbf{x}^m - \boldsymbol{\mu}_k)^T] = 0 \tag{14}$$

and Σ_k will be solved from (14) as

$$\Sigma_k = \frac{(1-\beta)\sum_{n=1}^{N}\gamma(n,k)(\mathbf{u}^n - \boldsymbol{\mu}_k)(\mathbf{u}^n - \boldsymbol{\mu}_k)^T}{(1-\beta)\sum_{n=1}^{N}\gamma(n,k) + \beta\sum_{m=1}^{M} y_k^m} + \frac{\beta\sum_{m=1}^{M} y_k^m (\mathbf{x}^m - \boldsymbol{\mu}_k)(\mathbf{x}^m - \boldsymbol{\mu}_k)^T}{(1-\beta)\sum_{n=1}^{N}\gamma(n,k) + \beta\sum_{m=1}^{M} y_k^m}. \tag{15}$$

As such, this completes the derivation of (5)–(7) for the semisupervised learning proposed in this section.

III. Probabilistic Fault Diagnosis Framework Under New Data Categories

In this section, an autoselection strategy on GMM's component number is applied during its training. For testing, the overall diagnosis framework is improved based on original confidence scores from GMM to detect and diagnose new data categories via two steps.

A. Autoselection Strategy on GMM's Component Number

In general, the number of Gaussian components, namely, total fault category K, should be predetermined before training. For the supervised learning, component number is easily decided from the labeled training data. However, for either unsupervised or semisupervised learning, engineers usually have to specify the component number based on expertise and trials before fitting to training data. It may be difficult in reality to know or find out an appropriate number of components due to existence of the unlabeled data.

Although $\mathbf{Y}$ provides a rough range on component number, it is still misleading to set component number to be the same as $\mathbf{Y}$s. $\mathbf{U}$ may include data samples from another new data categories. As such, an autoselection strategy on GMM's component number should be applied to relax the constraint of prior settings on K. Our motivation is to seek out an easy but effective way, which can evaluate the modeling fitness for multiple GMMs with varying components and output the best one as well-trained GMM via the semisupervised learning. In this paper, the AIC is used as our evaluation criterion for GMM, which is defined as

$$\mathrm{AIC} = 2N \log L + 2V \tag{16}$$

where $N \log L$ is the negative loglikelihood based on the training data and V is the total number of parameters to be estimated.

With the AIC implementation, our autoselection strategy on GMM component number is summarized in Fig. 1.

In Fig. 1, the semisupervised learning coefficient β is calculated by the number of data samples in $\mathbf{X}$ over the total number of training data samples. The maximum component number J is determined according to the category information from $\mathbf{Y}$, while it must be larger than total category number in $\mathbf{Y}$. An undue large J may incur the heavy computation cost and even singular covariance matrix in GMM. Next, a While loop is conducted for $j = 1, 2, \ldots, J$, in which multiple GMMs under j Gaussian components should be individually trained via (4)–(7) and their AICs are calculated correspondingly. When j is equal to J, the training procedure is completed and all multiple GMMs are checked. Finally, GMM with minimum AIC_j is output as the well-trained model, where the component number is $K = \arg\min \mathrm{AIC}_j$ for $j = 1, 2, \ldots, J$.

During training, our autoselection strategy uses the above AIC statistic to choose the best fitted GMM over varying component numbers. Next, the GMM with minimum AIC is used for the online testing of probabilistic fault diagnosis.

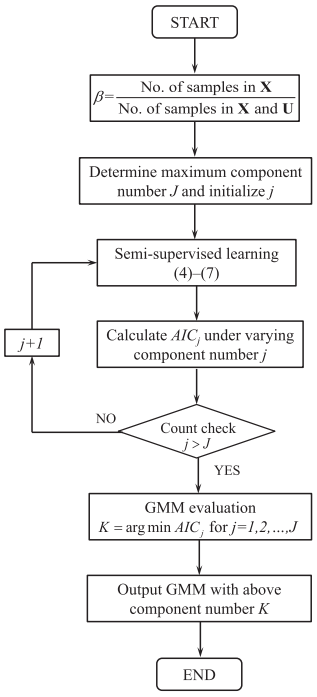


Fig. 1. Autoselection strategy on GMM component number during training.

B. Flowchart of Proposed Diagnosis Framework

GMM depends on feature inputs of each testing sample for probabilistic fault diagnosis. It is worth noting during the online testing that new types of faults may occur and their data samples are generated in reality. If so, conventional GMM is incapable of detecting new data categories. To overcome this limitation, the diagnosis framework should be improved, which targets to detect new types of faults and further categorize them in detail. Fig. 2 presents the overall flowchart of our proposed framework.

After the semisupervised learning with autoselection strategy on component number in Fig. 1, GMM with K Gaussian components is applied for testing. For one testing data sample $\mathbf{x}$, its feature inputs are fed into GMM, which is able to calculate the original confidence score into all existing K categories, namely, $\Pr(y_k = 1|\mathbf{x}) = \pi_k \mathcal{N}(\mathbf{x}; \boldsymbol{\mu}_k, \Sigma_k)$ for $k = 1, 2, \ldots, K$.

Instead of normalizing original confidence score to $((\Pr(y_k = 1|\mathbf{x}))/\Pr(\mathbf{x}))$ for conventional soft assignment, two decision blocks are introduced in our framework for new types of faults classification. As such, the proposed probabilistic diagnosis framework consists of two main steps, where the first step is to detect if data sample $\mathbf{x}$ is from new category or not and the second step is to categorize its details.

In the first decision block, the most likely category for input $\mathbf{x}$ is found out with the original confidence score of $\Pr(y_k = 1|\mathbf{x}) = \pi_k \mathcal{N}(\mathbf{x}; \boldsymbol{\mu}_k, \Sigma_k)$. Next, $\max \Pr(y_k = 1|\mathbf{x})$ is compared with a predetermined probability threshold ω. If $\max \Pr(y_k = 1|\mathbf{x})$ is smaller than ω, $\mathbf{x}$ should be classified as from another new category Category $K + 1$ and vice versa. The underlying motivation of this decision block is if $\mathbf{x}$ does not share enough statistical characteristics close

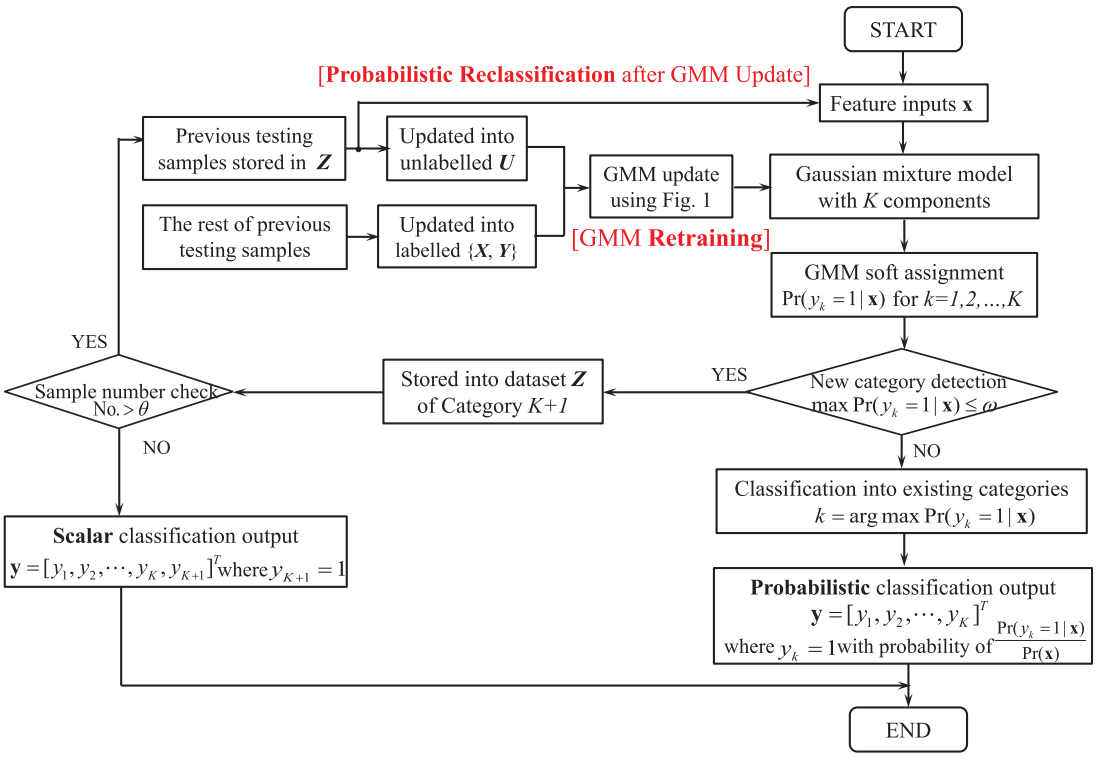


Fig. 2. Probabilistic diagnosis framework proposed to classify new data categories.

to its most likely category, it should be considered as from another new category and required to be further investigated. An appropriate selection of ω is crucial, and the labeled training data of $\{\mathbf{X}, \mathbf{Y}\}$ is used again for ω calculation. Data samples in $\{\mathbf{X}, \mathbf{Y}\}$ are fed into previously trained GMM, where the original confidence score of $\mathbf{x}^m$ to its corresponding category of $\mathbf{y}^m$ is calculated and recorded for $m = 1, 2, \ldots, M$. Next, ω is obtained as the mean of all above original confidence scores recorded.

After the first decision block, if $\mathbf{x}$ is not classified into new data category, namely, $\max \Pr(y_k = 1|\mathbf{x}) > \omega$, the trained GMM normalizes all $\Pr(y_k = 1|\mathbf{x})$ for $k = 1, 2, \ldots, K$ to sum up as one and outputs the probabilistic classification result of

$$\mathbf{y} = [y_1, y_2, \ldots, y_K]^T \tag{17}$$

where $y_k = 1$ is with $k = \arg\max \Pr(y_k = 1|\mathbf{x})$ and the corresponding classification probability of $((\Pr(y_k = 1|\mathbf{x}))/\Pr(\mathbf{x}))$.

Otherwise, $\mathbf{x}$ as one sample from new category is temporarily stored into dataset $\mathbf{Z}$ of new Category $K+1$. Since testing samples in $\mathbf{Z}$ may come from different new categories, it is not precise enough to classify all of them into one new category Category $K + 1$. By integrating all training and previous testing data, the proposed framework aims at updating GMM to classify the category details of testing samples in $\mathbf{Z}$.

Each component in GMM is one Gaussian distribution, whose mean vector and covariance matrix cannot be calculated unless enough corresponding data samples exist. From the statistic view, if the number of data samples is less than the number of feature inputs, the covariance matrix will be singular. As such, the second decision block is introduced to check whether the number of data samples in $\mathbf{Z}$ is enough for the GMM update. If number of data samples is not as many as θ, data samples stored in $\mathbf{Z}$ have to be labeled as one new category Category $K + 1$ only with the scalar result of

$$\mathbf{y} = [y_1, y_2, \ldots, y_K, y_{K+1}]^T \tag{18}$$

in which $y_{K+1} = 1$ is the hard assignment and the rest of terms are equal to zero.

On the other hand, when the number of data samples in $\mathbf{Z}$ is larger than θ, GMM is going to be retrained and updated accordingly. As shown in Fig. 2, all previous testing samples should be integrated with training data samples for the model update. Since testing samples stored in $\mathbf{Z}$ may come from various new categories, they are still regarded as the unlabeled data to update into $\mathbf{U}$. The rest of previous testing samples classified into existing data categories can be regarded as the labeled data to update into $\{\mathbf{X}, \mathbf{Y}\}$ correspondingly. Next, GMM is updated using Fig. 1, where j could be initialized as previous $K+1$ to avoid unnecessary computation trials. An autoselection strategy is executed to determine the component number during training. It is worth noting that θ is predetermined case by case according to expertise and realistic situations.

After GMM training with the updated number of Gaussian components, namely K, testing samples stored in $\mathbf{Z}$ are fed into model again to reclassify their category details with the probabilistic outputs. In all, for online testing, the proposed probabilistic framework is able to detect whether new data category occurs or not in the first step and further diagnose them via GMM update in the second step, which is more flexible and tailored in industry.

IV. Experimental Applications

To testify the effectiveness of our probabilistic diagnosis framework, two experimental setups on an industrial fault

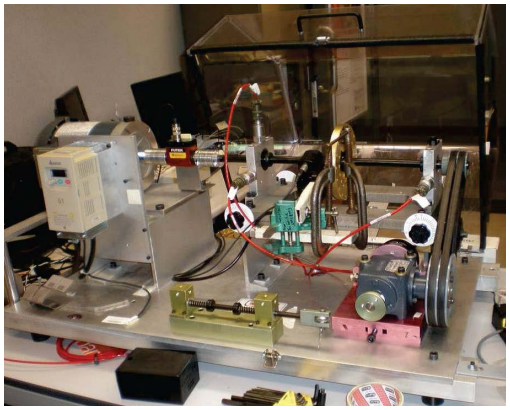


Fig. 3. Experimental setup of fault simulator.

TABLE I
SUMMARY ON ROTARY MACHINERY FAULT

Category No.	Rotary Machinery Fault
EHI1	Normal
EHI2	Bearing ball fault
EHI3	Imbalance
EHI4	Bearing outer race fault
EHI5	Bearing outer race with loose belt
EHI6	Imbalance with loose belt

simulator of rotary machine as well as the PD measurement of high-voltage electronic equipment are individually investigated in this section.

A. Industrial Fault Simulator of Rotary Machine

Fig. 3 displays the experimental setup on our fault simulator, in which the rotary machine consists of the rotor, bearing ball, belt drive, and motor. It can simulate various machinery faults such as the bearing fall fault, imbalance, bearing outer race fault, and two simultaneous faults with the loose belt.

During our experiment, 272 data samples from six machinery faults are balanced generated. Table I details all fault categories and digitalizes them from Equipment Health Index One (EHI1) to EHI6. Each EHI has 272 data samples.

In this experimental setup, eight sensors are installed inside the rotary machine for CM, namely, three current clamp meters for the current measurement of motor line, one Futek torque sensor placed on the shaft between the motor and first bearing to measure the torque generated, and four Kistler vibration sensors are placed near the two bearings to measure the vibration levels generated. Each sensor has a corresponding channel on the data acquisition module. In our experiment, the sensitivity of the accelerometer is 100 mV/g and the clamp meter's sensitivity is 100 mV/A.

Based on the logged sensor signals, statistical features calculated from time domain are implemented. Feature calculation is also conducted in the frequency domain. The first, second, and third harmonics of different frequencies are extracted to analyze the corresponding faults such as rotating imbalance, ball defect, and outer race defect. In total, there are 120 original features extracted from all eight sensors. The dominant feature identification approach [43] selects thirteen dominant features from four sensors only.

TABLE II
SAMPLES OF TRAINING AND TESTING UNDER TWO NEW MACHINERY CATEGORIES

Category No.	No. of Testing Samples	No. of Training Samples	
		Labelled	Unlabelled
EHI1	75	143	54
EHI2	62	131	79
EHI3	67	134	71
EHI4	68	-	204
EHI5	272	-	-
EHI6	272	-	-

After the feature extraction and selection, the experimental dataset from this machinery fault simulator is a 1632×14 matrix, whose last column is information about the digitalized fault category. The whole dataset is randomly split into training and testing partitions. To testify the effectiveness of autoselection strategy and improvement on new types of faults classification, both labeled and unlabeled data samples from different categories are combined in one arbitrary permutation for training; meanwhile, the category number from unlabeled data is purposely selected larger than labeled data's. For brevity, two representative diagnosis results for two and three new data categories are presented as follows. It is worth noting that our proposed framework does not know the total number of new categories in advance and the category corresponding to the maximum classification probability is regarded as the scalar classification result for accuracy calculation.

1) Diagnosis for Two New Machinery Categories: EHI5 and EHI6 are selected as two new data categories, which are totally unknown for the proposed framework. Table II displays the number of samples from each EHI into training and testing, where labeled training data consists of data samples arbitrarily selected from EHI1 to EHI3, while unlabeled training data contains samples arbitrarily selected from EHI1 to EHI4.

It can be calculated from Table II that the percentage of training data over the whole dataset is 50.0% and the semisupervised learning ratio is also $\beta = 50.0\%$. $J = 5$ and $\theta = 200$ are properly selected from observations on both the number of total data samples and the category number of whole training data. Being consistent with the category number from the labeled data in Table II, $j = 3$ is initialized instead of $j = 1$ to avoid redundant computation. Fig. 4(a) presents the AIC outputs of multiple GMMs with different component numbers during training. Finally, GMM with four Gaussian components is output as the best fitted model, which is consistent with the total number of training categories in Table II.

For online testing, when the number of data samples in $\mathbf{Z}$ surpasses θ, the former trained GMM should be updated. During update, $j = 4$ and $J = 7$ are chosen again without loss of generality. Fig. 4(b) presents the AIC outputs of the updated GMMs under different component numbers during update. The GMM component number is updated to six and the corresponding Gaussian components are recalculated.

Table III summarizes the classification results of testing samples from all six EHIs.

Besides individual accuracies in Table III, the overall diagnosis accuracy is calculated as 94.0%. Samples from

TABLE III
CLASSIFICATION RESULTS UNDER TWO NEW MACHINERY CATEGORIES OF EHI5 AND EHI6

Category No.	Testing Samples	Classification Details							
		As EHI1	As EHI2	As EHI3	As EHI4	As EHI5	As EHI6	As EHI7	Accuracy
EHI1	75	74	1	-	-	-	-	-	98.8%
EHI2	62	-	61	-	-	-	-	1	98.4%
EHI3	67	-	-	67	-	-	-	-	100.0%
EHI4	68	-	-	-	68	-	-	-	100.0%
EHI5	272	-	-	-	-	270	2	-	99.3%
EHI6	272	-	-	-	-	44	227	1	83.5%

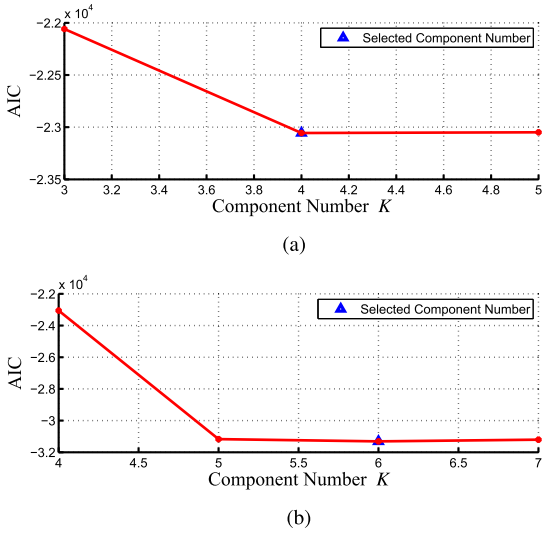


Fig. 4. AIC outputs of multiple GMMs with different component number under two new machinery categories. (a) AIC output during GMM training. (b) AIC output during GMM update.

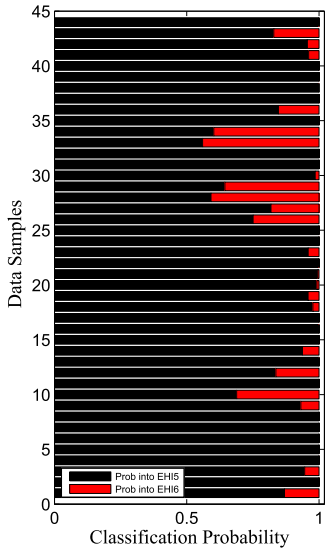


Fig. 5. Probabilities of misclassified data samples under two new machinery categories.

existing data categories like EHI1–EHI4 can be diagnosed correctly, and most of the data samples from new categories like EHI5 and EHI6 can be also classified well. However, it is observed that 44 data samples from EHI6 are misclassified into EHI5. Since GMM as soft assignment is able to provide the confidence scores for probabilistic fault classification, their classification probabilities into EHI5 and EHI6 are displayed as representative examples in Fig. 5.

It can be observed from Fig. 5 that some of data samples misclassified from EHI6 into EHI5 still hold a certain amount of classification probability into EHI6. As such, the probabilistic fault diagnosis provides more diagnosis information and assists to comprehensively locate, investigate, and repair possible faulty components.

2) Diagnosis for Three New Machinery Categories: In addition, EHI4, EHI5, and EHI6 are selected as three new data categories for testing. Similar to Table II, Table IV summarizes the number of samples from each EHI randomly split into training and testing.

As calculated from Table IV, the percentage of training data over the whole dataset is 33.3% and the semisupervised learning ratio is $\beta = 33.3\%$. Similar to above, $J = 5$ and $\theta = 200$ is still selected during training. $j = 2$ is initialized based on the labeling training data, and the AIC outputs are plotted in Fig. 6(a). As such, GMM with three

TABLE IV
SAMPLES OF TRAINING AND TESTING UNDER THREE NEW MACHINERY CATEGORIES

Category No.	No. of Testing Samples	No. of Training Samples	
		Labelled	Unlabelled
EHI1	96	89	87
EHI2	85	92	95
EHI3	91	-	181
EHI4	272	-	-
EHI5	272	-	-
EHI6	272	-	-

Gaussian components is output as the best model, which is also consistent with the category number of training data in Table IV.

For testing, GMM will be also updated again once the number of data samples in $\mathbf{Z}$ surpasses θ. During update, $j = 3$ and $J = 7$ are chosen, and Fig. 6(b) presents AIC outputs. Finally, GMM with six Gaussian components is retained as model output.

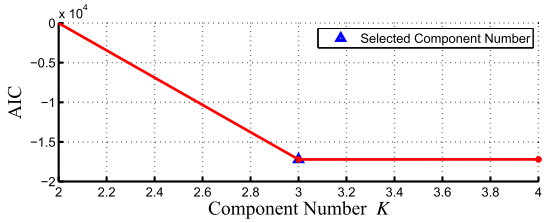
Table V summarizes the classification results of testing data samples from all six EHIs.

It is calculated from Table V that the overall diagnosis accuracy is 95.8%. Most of the testing data samples are classified correctly, while 39 data samples from EHI6 are still misclassified into EHI5. Owing to advantages of soft assignment, their classification probabilities into EHI5 and into EHI6 are also summarized in Fig. 7. Similar to Fig. 5, some misclassified data samples still keep

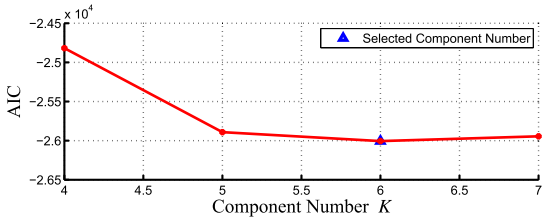
TABLE V

CLASSIFICATION RESULTS UNDER THREE NEW MACHINERY CATEGORIES OF EHI4–EHI6

Category No.	Testing Samples	Classification Details							
		As EHI1	As EHI2	As EHI3	As EHI4	As EHI5	As EHI6	As EHI7	Accuracy
EHI1	96	95	1	-	-	-	-	-	99.0%
EHI2	85	-	84	-	-	-	-	1	98.8%
EHI3	91	-	-	90	-	-	-	1	98.9%
EHI4	272	-	-	-	271	-	-	1	99.6%
EHI5	272	-	-	-	-	271	1	-	99.6%
EHI6	272	-	-	-	-	39	232	1	85.3%



(a)



(b)

Fig. 6. AIC outputs of multiple GMMs with different component number under three new machinery categories. (a) AIC output during GMM training. (b) AIC output during GMM update.

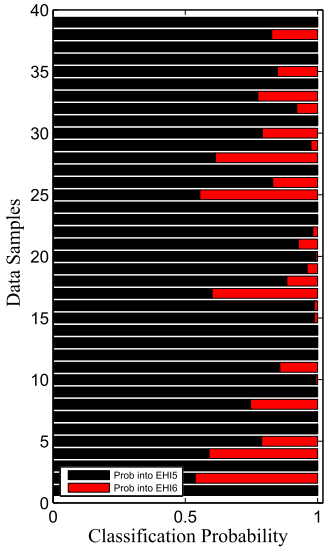


Fig. 7. Probabilities of misclassified data samples under three new machinery categories.

corresponding classification probability into the correct category and engineers may focus on all possible faulty components inside for further investigation.

To sum up, observations from above two diagnosis results are as follows.

1) Our autoselection strategy on GMM's component number works well in both the GMM training and update, where appropriate K can be autoselected. As observed from Figs. 4 and 6, an AIC drop exists at a crucial value. After it, AIC value does not change much even along with component number increases. In general, this crucial value should be empirically selected, since the proportions of additional components are almost close to zero even if the component number further increases.
2) Less information is provided under three new categories than two new categories, which may increase the training difficulty. However, it can be observed from Tables III and V that diagnosis results under three new categories are slightly better than them under two new categories, which may result from the arbitrary training data split and random initialization of GMM parameters.
3) As observed from Tables III and V, a few data samples from existing EHIs are misclassified into EHI7, which does not exist in reality. Our framework does not have the prior knowledge about the total number of data categories. When checking original confidence scores in the first decision block, the proposed framework misclassifies the corresponding data samples from realistic EHIs into a nonexisting new category, namely, EHI7 in Tables III and V.
4) Most of the testing data samples from new categories are diagnosed well. However, from above two diagnosis

TABLE VI

COMPARISON OF DIAGNOSIS ACCURACY WITH EXISTING APPROACHES

Approaches	Diagnosis Accuracy	
	Without new categories	With new categories
SOM	79.8%	32.2%
BPNN	96.5%	46.9%
ELM	97.6%	47.2%
Proposed Framework	97.9%	93.5%

TABLE VII

SUMMARY OF PD HEALTH CONDITION

Category No.	PD Health Condition	Inspection and Repair Advices
EHI1	Normal	1 year inspection
EHI2	Borderline	3-4 months inspection
EHI3	Borderline with Trending	7 days continuous inspection
EHI4	Minor	Service soon
EHI5	Significant	Service immediately

TABLE VIII

SAMPLES OF TRAINING AND TESTING UNDER TWO NEW PD CATEGORIES

Category No.	No. of Testing Samples	No. of Training Samples	
		Labelled	Unlabelled
EHI1	120	260	120
EHI2	125	-	375
EHI3	130	240	130
EHI4	500	-	-
EHI5	500	-	-

TABLE IX
CLASSIFICATION RESULTS UNDER TWO NEW PD CATEGORIES OF EHI4 AND EHI5

Category No.	Testing Samples	Classification Details					
		As EHI1	As EHI2	As EHI3	As EHI4	As EHI5	Accuracy
EHI1	120	100	11	8	1	-	83.3%
EHI2	125	3	32	90	-	-	72.0%
EHI3	130	2	117	11	-	-	90.0%
EHI4	500	101	1	-	381	17	76.2%
EHI5	500	2	-	-	-	498	99.6%

results, around 40 samples from EHI6 are misclassified into EHI5. One main reason is that EHI5 and EHI6 are multiple faults with loose belt, which share some similar characteristics. Unlike the hard assignment, probabilistic fault diagnosis provides the classification probabilities, which support further investigation and inspection for engineers.

3) Comparison With Existing Approaches: Our proposed framework is also compared with other existing approaches to validate the performance improvement. Since conventional diagnosis approaches are incapable of detecting new types of faults, which only classify all testing data samples into existing categories from training, two additional diagnosis studies with and without new data categories are carried out individually.

Self-organizing map (SOM) [19], backpropagation NN (BPNN) [22], and extreme learning machine (ELM) [20] are chosen as three existing approaches. Their model structures are tuned via trials and observations to achieve an acceptable training accuracy. Similar to Table II but with the different training percentage of 36.7%, EHI5 and EHI6 are still selected as new data categories for testing. Their diagnosis accuracies are summarized in Table VI.

It can be observed from Table VI that our proposed framework is able to achieve as good performance as existing approaches for conventional diagnosis work without new data categories. Owing to the good initialization of GMM parameters, which is obtained by simulation trials and observations to achieve an acceptable training accuracy, our proposed framework achieves the best diagnosis accuracy in this case study. For diagnosis work with new data categories, our proposed framework outperforms than other three existing approaches, which is able to classify most of the data samples from new data categories correctly via the GMM update.

B. Partial Discharge Measurement of Electronic Equipment

Besides the above industrial fault simulator of rotary machine, another dataset from the PD measurement of various high-voltage electronic and power equipment components is applied to further testify the effectiveness and applicability of our proposed framework in practice.

Material erosion, production defects, and overvoltage result in the insulation damage of the electronic and power equipment, which leads to the PD occurrence in reality. Severe equipment PDs may cause the self-breakdown, damages to the neighboring equipment, and even failures of the whole plant. PD engineers can use the PD portable analyzer console to measure the average PD magnitude and the number of PD pulses for the high voltage equipment in the field such as transformer, generator, motor, and cable. The corresponding operating temperature, humidity, and loading are also logged as the associated CM information. As such, each logged data sample has five CM feature inputs in total.

With support of expert knowledge and experience, all PD measurement data are divided and labeled into five EHIs, as detailed in Table VII. It is worth noting that the original PD measurement dataset is imbalanced. SMOTE [44] has oversampled the minority data samples to balance the dataset. In this paper, 500 data samples in each EHI are arbitrarily picked up from the oversampled dataset and combined together. The PD measurement dataset here is a 2500×6 matrix, whose last column is also the corresponding EHI information.

Similar to Table II of the machinery fault simulator, EHI4 and EHI5 are selected as two new PD categories in this case study. Also similar to above case studies, our proposed framework does not know them in advance. Table VIII displays the number of data samples from each EHI into training and testing, which are arbitrarily selected and combined.

As calculated from Table VIII, the percentage of training data over the whole dataset is 45.0% and β is equal to 44.5%. Table IX summarizes the classification results of testing samples from all five PD EHIs. In addition, the overall diagnosis accuracy is calculated as 86.3%.

It can be also observed from Table IX that most of the testing samples from new data categories, namely, EHI4 and EHI5, are classified well with an average accuracy of 87.9%, while their overall accuracy is not as high as the first case study of machinery fault simulator. We investigate the feature signals among all five EHIs and find out some heavy overlaps exist in PD magnitude and PD pulse count between EHI1 and EHI2, EHI1 and EHI4, as well as EHI2 and EHI3. Data samples from different EHIs still share some similar characteristics, which may result in the above misclassification cases of both the known and new data categories in Table IX. However, it is worth noting from view of PD engineers that the overall accuracy is still acceptable for this specific industry field.

It is concluded based on the above two industrial applications that GMM using novel semisupervised learning is a good candidate to improve probabilistic diagnosis framework especially under new data categories. It not only has as good diagnosis performance as existing diagnosis approaches in Table VI, but also is good at classifying new data categories if occurred in Tables III, V, and IX.

V. CONCLUSION

In this paper, a novel framework was proposed for the probabilistic fault diagnosis under new data categories. GMM was applied as the pattern recognition algorithm, while its training was improved from conventional unsupervised

learning to novel semisupervised learning. Since unlabeled training data existed, an autoselection strategy on GMM's component number was implemented. For testing, our proposed framework depended on the probabilistic classification results from GMM to first detect whether new categories occur and further categorize them in detail. The effectiveness of our diagnosis framework was testified on an industrial fault simulator of rotary machine and the PD measurement of high-voltage electronic equipment. It was compared with the three existing approaches, where the proposed framework not only could achieve overall accuracy of 97.9% for conventional fault diagnosis without new data categories, but also was able to effectively classify new types of faults with overall accuracy of at least 86.3%. Our future work includes the training improvement to overcome the negative effects of random parameter initialization, the framework improvement to effectively address the imbalanced data, as well as the overall robustness analysis.

REFERENCES

[1] S. Nandi, H. A. Toliyat, and X. Li, "Condition monitoring and fault diagnosis of electrical motors—A review," *IEEE Trans. Energy Convers.*, vol. 20, no. 4, pp. 719–729, Apr. 2005.

[2] R. Yan and R. X. Gao, "Energy-based feature extraction for defect diagnosis in rotary machines," *IEEE Trans. Instrum. Meas.*, vol. 58, no. 9, pp. 3130–3139, Sep. 2009.

[3] J. Yu and S. J. Qin, "Variance component analysis based fault diagnosis of multi-layer overlay lithography processes," *IIE Trans.*, vol. 41, no. 9, pp. 764–775, 2009.

[4] S. Simani, C. Fantuzzi, and R. J. Patton, *Model-Based Fault Diagnosis in Dynamic Systems Using Identification Techniques*. New York, NY, USA: Springer, 2013.

[5] J. Lee, F. Wu, W. Zhao, M. Ghaffari, L. Liao, and D. Siegel, "Prognostics and health management design for rotary machinery systems—Reviews, methodology and applications," *Mech. Syst. Signal Process.*, vol. 42, no. 1, pp. 314–334, 2014.

[6] A. K. S. Jardine, D. Lin, and D. Banjevic, "A review on machinery diagnostics and prognostics implementing condition-based maintenance," *Mech. Syst. Signal Process.*, vol. 20, no. 7, pp. 1483–1510, 2006.

[7] I. Hwang, S. Kim, Y. Kim, and C. E. Seah, "A survey of fault detection, isolation, and reconfiguration methods," *IEEE Trans. Control Syst. Technol.*, vol. 18, no. 3, pp. 636–653, May 2010.

[8] C. K. Pang, F. L. Lewis, T. H. Lee, and Z. Y. Dong, *Intelligent Diagnosis and Prognosis of Industrial Networked Systems*. Boca Raton, FL, USA: CRC Press, 2011.

[9] Z. Gao, C. Cecati, and S. X. Ding, "A survey of fault diagnosis and fault-tolerant techniques—Part I: Fault diagnosis with model-based and signal-based approaches," *IEEE Trans. Ind. Electron.*, vol. 62, no. 6, pp. 3757–3767, Jun. 2015.

[10] A. Sapena-Bañó, M. Pineda-Sanchez, R. Puche-Panadero, J. Martinez-Roman, and D. Matić, "Fault diagnosis of rotating electrical machines in transient regime using a single stator current's FFT," *IEEE Trans. Instrum. Meas.*, vol. 64, no. 11, pp. 3137–3146, Nov. 2015.

[11] M. D. Prieto, D. Z. Millan, W. Wang, A. M. Ortiz, J. A. O. Redondo, and L. R. Martinez, "Self-powered wireless sensor applied to gear diagnosis based on acoustic emission," *IEEE Trans. Instrum. Meas.*, vol. 65, no. 1, pp. 15–24, Jan. 2016.

[12] D.-S. Lee, "Effective Gaussian mixture learning for video background subtraction," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 5, pp. 827–832, May 2005.

[13] Q. Ding, J. Han, X. Zhao, and Y. Chen, "Missing-data classification with the extended full-dimensional Gaussian mixture model: Applications to EMG-based motion recognition," *IEEE Trans. Ind. Electron.*, vol. 62, no. 8, pp. 4994–5005, Aug. 2015.

[14] H.-C. Yan, J.-H. Zhou, and C. K. Pang, "New types of faults detection and diagnosis using a mixed soft & hard clustering framework," in *Proc. IEEE ETFA*, Sep. 2016, pp. 1–4.

[15] Q. P. He and J. Wang, "Fault detection using the k-nearest neighbor rule for semiconductor manufacturing processes," *IEEE Trans. Semicond. Manuf.*, vol. 20, no. 4, pp. 345–354, Nov. 2007.

[16] J. W. Sheppard and M. A. Kaufman, "A Bayesian approach to diagnosis and prognosis using built-in test," *IEEE Trans. Instrum. Meas.*, vol. 54, no. 3, pp. 1003–1018, Jun. 2005.

[17] K. Choi *et al.*, "Novel classifier fusion approaches for fault diagnosis in automotive systems," *IEEE Trans. Instrum. Meas.*, vol. 58, no. 3, pp. 602–611, Mar. 2009.

[18] K. E. A. van de Sande, T. Gevers, and C. G. M. Snoek, "Evaluating color descriptors for object and scene recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 9, pp. 1582–1596, Sep. 2010.

[19] J. Vesanto and E. Alhoniemi, "Clustering of the self-organizing map," *IEEE Trans. Neural Netw.*, vol. 11, no. 3, pp. 586–600, May 2000.

[20] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, "Extreme learning machine: Theory and applications," *Neurocomputing*, vol. 70, nos. 1–3, pp. 489–501, 2006.

[21] A. Widodo and B.-S. Yang, "Support vector machine in machine condition monitoring and fault diagnosis," *Mech. Syst. Signal Process.*, vol. 21, no. 6, pp. 2560–2574, 2007.

[22] J.-D. Wu and C.-H. Liu, "Investigation of engine fault diagnosis using discrete wavelet transform and neural network," *Expert Syst. Appl.*, vol. 35, no. 3, pp. 1200–1213, 2008.

[23] F. Zidani, D. Diallo, M. E. H. Benbouzid, and R. Nait-Said, "A fuzzy-based approach for the diagnosis of fault modes in a voltage-fed PWM inverter induction motor drive," *IEEE Trans. Ind. Electron.*, vol. 55, no. 2, pp. 586–593, Feb. 2008.

[24] C. K. Pang, J.-H. Zhou, and H.-C. Yan, "PDF and breakdown time prediction for unobservable wear using enhanced particle filters in precognitive maintenance," *IEEE Trans. Instrum. Meas.*, vol. 64, no. 3, pp. 649–659, Mar. 2015.

[25] J. A. Bilmes, "A gentle tutorial of the EM algorithm and its application to parameter estimation for Gaussian mixture and hidden Markov models," *Int. Comput. Sci. Inst.*, vol. 4, no. 510, p. 126, 1998.

[26] C. M. Bishop, *Pattern Recognition and Machine Learning*. New York, NY, USA: Springer, 2006.

[27] D. Reynolds, "Gaussian mixture models," in *Encyclopedia Biometrics*. New York, NY, USA: Springer, 2009, pp. 659–663.

[28] B. Safarinejadian, M. B. Menhaj, and M. Karrari, "Distributed unsupervised Gaussian mixture learning for density estimation in sensor networks," *IEEE Trans. Instrum. Meas.*, vol. 59, no. 9, pp. 2250–2260, Sep. 2010.

[29] J. P. Vila and P. Schniter, "Expectation-maximization Gaussian-mixture approximate message passing," *IEEE Trans. Signal Process.*, vol. 61, no. 19, pp. 4658–4672, Oct. 2013.

[30] O. Chapelle, B. Schölkopf, and A. Zien, *Semi-Supervised Learning*. Cambridge, MA, USA: MIT Press, 2006.

[31] X. Zhu and A. B. Goldberg, "Introduction to semi-supervised learning," *Synth. Lect. Artif. Intell. Mach. Learn.*, vol. 3, no. 1, pp. 1–130, 2009.

[32] A. Martinez-Uso, F. Pla, and J. M. Sotoca, "A semi-supervised Gaussian mixture model for image segmentation," in *Proc. IEEE ICPR*, Aug. 2010, pp. 2941–2944.

[33] G. Huang, S. Song, J. N. D. Gupta, and C. Wu, "Semi-supervised and unsupervised extreme learning machines," *IEEE Trans. Cybern.*, vol. 44, no. 12, pp. 2405–2417, Dec. 2014.

[34] B. Fernando, E. Fromont, D. Muselet, and M. Sebban, "Supervised learning of Gaussian mixture models for visual vocabulary generation," *Pattern Recognit.*, vol. 45, no. 2, pp. 897–907, 2012.

[35] J. Goldberger and S. T. Roweis, "Hierarchical clustering of a mixture model," in *Proc. Adv. Neural Inf. Process. Syst.*, 2004, pp. 505–512.

[36] M. A. T. Figueiredo and A. K. Jain, "Unsupervised learning of finite mixture models," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 3, pp. 381–396, Mar. 2002.

[37] Z. Zivkovic and F. van der Heijden, "Recursive unsupervised learning of finite mixture models," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 5, pp. 651–656, May 2004.

[38] Y. Zhang and M. S. Scordilis, "Effective online unsupervised adaptation of Gaussian mixture models and its application to speech classification," *Pattern Recognit. Lett.*, vol. 29, no. 6, pp. 735–744, Sep. 2008.

[39] J. Cusido, J. Rosero, E. Aldabas, J. A. Ortega, and L. Romeral, "New fault detection techniques for induction motors," *Electr. Power Quality Utilisation Mag.*, vol. 2, no. 1, pp. 39–46, 2006.

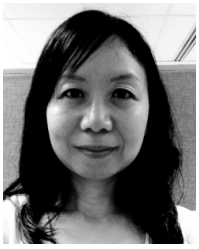
[40] Y.-H. Bae, S.-H. Lee, H.-C. Kim, B.-R. Lee, J. Jang, and J. Lee, "A real-time intelligent multiple fault diagnostic system," *Int. J. Adv. Manuf. Technol.*, vol. 29, no. 5, pp. 590–597, 2006.

[41] Y. Wang, S. X. Ding, H. Ye, and G. Wang, "A new fault detection scheme for networked control systems subject to uncertain time-varying delay," *IEEE Trans. Signal Process.*, vol. 56, no. 10, pp. 5258–5268, Oct. 2008.

[42] L. Yao, J. Qin, H. Wang, and B. Jiang, "Design of new fault diagnosis and fault tolerant control scheme for non-Gaussian singular stochastic distribution systems," *Automatica*, vol. 48, no. 9, pp. 2305–2313, 2012.
[43] J.-H. Zhou, C. K. Pang, Z.-W. Zhong, and F. L. Lewis, "Tool wear monitoring using acoustic emissions by dominant-feature identification," *IEEE Trans. Instrum. Meas.*, vol. 60, no. 2, pp. 547–559, Feb. 2011.
[44] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," *J. Artif. Intell. Res.*, vol. 16, no. 1, pp. 321–357, 2002.

Heng-Chao Yan (S'13) was born in Hanzhong, China, in 1990. He received the B.Eng. degree in automation and the B.Econ. degree in finance from East China University of Science and Technology, Shanghai, China, both in 2012. He is currently pursuing the Ph.D. degree with the Department of Electrical and Computer Engineering, National University of Singapore, Singapore.

He has been a Research Student with the A*STAR Singapore Institute of Manufacturing Technology, Singapore, since 2013. His current research interests include intelligent diagnostics and prognostics, precognitive maintenance, system identification, and data analytics.



Jun-Hong Zhou (M'09) received the Ph.D. degree from Nanyang Technological University, Singapore. Her Ph.D. research focused on advance feature extraction and selection in condition-based maintenance.

She was with industries involving the SCADA system, sensing and measurement for machine tooling condition, intelligent modeling for equipment health prognostics, and process monitoring and product control in manufacturing processes and intelligent systems to maintain serviceability of manufacturing equipment. Since 1996, she has been with the Singapore Institute of Manufacturing Technology (SIMTech), Singapore, where she has been involved in a number of research and industry projects in the areas of process monitoring and control, sensing and advanced signal processing, data analytical, and data mining. She is currently a Principle Researcher with the A*STAR SIMTech, where she is the initiative lead of the area of maximizing overall equipment effectiveness. She has authored over 50 technical papers.

Dr. Zhou was a recipient of the Best Application Paper Award in the 8th Asian Control Conference in 2011 and the Best Paper in the IASTED International Conference on Engineering and Applied Science in 2012.

Chee Khiang Pang (Justin) (S'04–M'07–SM'11) received the B.Eng. (Hons.), M.Eng., and Ph.D. degrees in electrical and computer engineering from the National University of Singapore (NUS), Singapore, in 2001, 2003, and 2007, respectively.

In 2003, he joined the School of Information Technology and Electrical Engineering (ITEE), University of Queensland, St. Lucia, QLD, Australia, as a Visiting Fellow. From 2006 to 2008, he was a Researcher (Tenure) with the Central Research Laboratory, Hitachi Ltd., Kokubunji, Tokyo, Japan. In 2007, he joined the School of ITEE, as a Visiting Academic. From 2008 to 2009, he was a Visiting Research Professor with the Automation and Robotics Research Institute, University of Texas at Arlington, Fort Worth, TX, USA. From 2009 to 2016, he was an Assistant Professor with the Department of Electrical and Computer Engineering, NUS. He is also an Adjunct Assistant Professor in NUS-ECE, a Faculty Associate of the A*STAR Singapore Institute of Manufacturing Technology, Singapore, where he is a Faculty Associate of the A*STAR Data Storage Institute. He is currently an Assistant Professor (Tenure) in engineering cluster with the Singapore Institute of Technology, Singapore. He has authored or edited three research monographs including *Intelligent Diagnosis and Prognosis of Industrial Networked Systems* (CRC Press, 2011), *High-Speed Precision Motion Control* (CRC Press, 2011), and *Advances in High-Performance Motion Control of Mechatronic Systems* (CRC Press, 2013). His current research interests are in data-driven control and optimization, with realistic applications to robotics, mechatronics, and manufacturing systems.

Dr. Pang is a member of ASME. He was a recipient of the Best Application Paper Award in the 8th Asian Control Conference, Kaohsiung, Taiwan, in 2011, and the Best Paper Award in the IASTED International Conference on Engineering and Applied Science, Colombo, Sri Lanka, in 2012. He served as a Guest Editor of the *International Journal of Automation and Logistics*, the *Asian Journal of Control*, the *International Journal of Systems Science*, the *Journal of Control Theory and Applications*, and the *Transactions of the Institute of Measurement and Control*. He serves as an Executive Editor for *Transactions of the Institue of Measurement and Control* an Associate Editor of the *Asian Journal of Control*, the *IEEE Control Systems Letters*, the *Journal of Defense Modeling & Simulation*, and the *Unmanned Systems*, and serves on the Editorial Board of the *International Journal of Automation and Logistics* and the *International Journal of Computational Intelligence Research and Applications*, and on the Conference Editorial Board of the IEEE Control Systems Society.