Received December 28, 2020, accepted January 10, 2021, date of publication January 13, 2021, date of current version January 28, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3051409

Soft Fault Diagnosis Using URV-LDA Transformed Feature Dictionary

CEN CHEN, (Member, IEEE), YUN YANG, XUERONG YE, (Senior Member, IEEE), AND GUOFU ZHAI, (Member, IEEE)

Department of Electrical Engineering, Harbin Institute of Technology, Harbin 150001, China

Corresponding author: Xuerong Ye (xuelai1981@163.com)

This work was supported in part by the National Research and Development Program of China under Grant 2017YFB1300800 and in part by the National Natural Science Foundation of China under Grant 61671172.

ABSTRACT Dictionary-based fault diagnosis methods, focusing on storing feature patterns of known faults, have been widely used for electromechanical systems. The state of component degradation caused soft faults, however, are continuously changeable. Thus, conventional dictionaries cannot be applied for diagnosis of soft faults with multi-degradation levels. To address this issue, this article develops a new type of dictionary by combining the unit residual signal vector (URV) and the linear discriminant analysis (LDA) for feature transformation, which is referred to as URV-LDA dictionary. The unit residual signal vector keeps the fault feature growth trends but eliminates the degradation severity influence. The linear discriminant analysis is then implemented to find the best projection directions for classification. Specifically, two dictionaries named as the URV-MLDA binary-value dictionary and the URV-SLDA unique-value dictionary are proposed. To validate the efficiency of two developed dictionaries, an electromagnetic relay is carried out and two conventional methods are compared. The comparison results show the developed dictionaries can better solve the soft faults issues with significant increases on diagnostic accuracy.

INDEX TERMS Fault diagnosis, fault dictionary, linear discriminant analysis, and electromagnetic relay.

I. INTRODUCTION

In modern industries, electromechanical systems have become more integrated and complex [1]. The continuously severe operational conditions may cause the degradation of the components in the system, and unexpected instantaneous environment shocks may cause the catastrophic failure of components, which may further cause system faults. The study of fault diagnosis for these electromechanical systems attracts more attention in recent decades [2]–[4]. Compared with component catastrophic failures (hard faults) which are always determined states (such as the open circuit failure or the short circuit failure), degradation caused faults (such as the bearing crack generation), also known as soft faults, are gradually generated with continuously changeable states. Thus, fault diagnosis for soft faults is more difficult [2], [5]–[7].

Fault diagnostic methods can be categorized as data-driven methods and model-based methods. Data-driven methods mainly utilize extracted features of measured signals for fault diagnosis. Some representative data-driven methods include the principal component analysis (PCA)-based methods [8], [9], the support vector machine (SVM) [10], and the artificial neural network (ANN)-based methods. The diagnostic performance of these methods relies on the correctness and completeness of extracted healthy and faulty features [11]. However, the data mining and the feature extraction process can be regarded as a random searching process, since the usability of a specific feature cannot be predicted before it is extracted. Thus, these methods are better suitable for systems with large amounts of historical and online measured signals/features. When data acquisition or diagnostic time is limited, it requires more effort to minimize the number of monitored features and simplify the diagnostic model. Model-based methods, however, utilize the knowledge of the system model to directly figure out the relationship between fault states and monitored physical features for fault diagnosis. The knowledge can be obtained from physical principles, fault mechanisms, and relevant expertise, *et al*. Compared with the data-driven methods, model-based methods are suitable for systems whose fault effect and theoretical model are well studied so as to avoid the

The associate editor coordinating the review of this manuscript and approving it for publication was Qingchao Jiang.

random searching of useful features. It can be further classified into two types. One is parameter identification-based methods, which focus on establishing the quantitative system model between component parameters (or state parameters) and the system-level observers (testable features of the system). The predicted features derived from the model and the monitored features are compared to isolate the fault location [12], [13]. However, such quantitative models are difficult to be constructed for complex systems. The other type, the dictionary-based methods, prefer to establish qualitative relationships between given system states and observers. The dictionary-based methods only map the weak dependency relationship between predetermined fault states and a series of system features, which do not require the precise quantitative system model [14]. Traditionally, such dictionaries can be derived from various theoretical or numeric models such as the expert system models [15], [16], the graph theory models [3], [17]–[19], and fault simulation models [20]–[22], *et al*. As the signal processing technique develops, more data-driven methods can also be used to construct dictionaries. Due to its simplicity and directness, the dictionary-based methods are widely used for soft fault diagnosis in analog circuits [17]–[19], [23]–[25], digital circuits [26], [27], and modular systems [28], [29].

Although the dictionary-based methods are widely used, they still have drawbacks. The conventional dictionaries only focus on describing the feature distributions of deterministic fault states, thus, it is unable to diagnose the faults with undetermined states. Since the states of soft faults are continuously changeable, the conventional dictionary-based diagnostic methods cannot be applied for soft faults [21]. On the other hand, the simplicity and directness of the dictionary demand the sacrifice of the feature distribution precision. In the conventional dictionaries, the feature distribution of each fault state is represented using a series of independent feature intervals. Without considering the dependency among features, the conventional dictionaries will ‘‘expand’’ the real feature distribution of each fault state. Thus, the impractical overlap of the feature distributions under different fault states occurs, which may waste the usability of features and may further lead to an improper diagnostic strategy. To solve the soft faults issue, a proper idea is to find out new features which are insensitive to the severity level of faults but are sensitive to the fault modes. Taking the linear analog circuit as an example, [21] developed a dictionary using the slope of the node-voltage as features. Even the severity level of faults changes, the slope is invariant for a given fault mode. In [30], the current residual is selected as the feature, and the sign of which is used to classify different faults of a pulse width modulation (PWM) rectifier. Although the magnitude of the residual varies as the fault severity increases, the sign of the residual still keeps constant. In [31], the features derived from the residual signal are approximately linearly increased as the fault severity increases. It also indicates the residual signal can show the fault growth directions. For the overlapping issue, it can be solved by projecting original features into a new feature space. In the field of the data-driven fault diagnosis studies, various linear transformation methods (such as the principal component analysis (PCA), and the linear discriminant analysis [32], [33]) and nonlinear methods (such as ANN, and SVM [34]) are available to solve this issue. Compared with the nonlinear transformation methods, the linear transformation methods have simple structures and are more efficient to avoid the overfitting problem. As mentioned before, applications of model-based methods are usually with relatively small feature sample size. In that case, the overlapping issue is more likely to occur. Thus, a simple linear transformation method is more suitable. The linear discriminant analysis, which is such a robust and simple linear transformation method, has been successfully applied for Tennessee Eastman process [32], [34], and bearings [35]–[37] to improve the diagnostic accuracy.

Inspired by the above studies, this article proposes a new type of dictionary— the unit residual vector linear discriminant analysis (URV-LDA) dictionary, to drawback the soft fault issue of conventional dictionary-based diagnosis methods. By utilizing the unit residual signal features instead of the original features, the new dictionary stores the fault feature growth trends, which can better consider the ‘‘expandability’’ of the feature distribution caused by degradation. Thus, a better diagnostic ability for soft faults can be obtained. Using the transformed unit residual signal features, the linear discriminant analysis is then implemented to further learn best projection directions without losing original growth trends.

This article is organized as follows. In Section II, the dictionary-based fault diagnosis methods are summarized. For the first time, three types of dictionaries are categorized based on the value form of elements in the dictionary. Section III describes the proposed two developed new dictionaries represented with the combination of the residual signal vector and the linear discriminant analysis. To validate the efficiency of the developed dictionaries, the fault simulating experiment of an integrated electromechanical device — the electromagnetic relay is carried out in Section IV. And the diagnostic accuracy results, as well as the comparison with other methods, are also discussed. The main contributions are then concluded in Section V.

II. REVISIT OF DICTIONARY-BASED FAULT DIAGNOSIS

The dictionary-based fault diagnosis methods are widely used in many systems. Although such methods all use a dictionary for fault diagnosis, the dictionaries may have slightly different expression types in different application fields. In this part, the authors category dictionaries into three basic types for the first time. The category is based on the value form of elements in the dictionary. Which are named as the binary-value dictionary, the multi-value dictionary, and the unique-value dictionary.

A. BINARY-VALUE DICTIONARY

The binary-value dictionary is named because the elements are binary, which are filled with ‘‘1’’ and ‘‘0’’. Table 1 is

TABLE 1. Example of a Binary-Value Dictionary

	t_1	t_2	t_3
Fault-free	0	0	0
Fault 1	0	0	1
Fault 2	0	1	0
Fault 3	1	0	0

TABLE 2. Example of a Multi-Value Dictionary

	t_1	t_2	t_3
Fault-free	0	0	0
Fault 1	1	1	1
Fault 2	1	2	2
Fault 3	2	1	0

an example of the typical binary-value dictionary with three features $t_1 \sim t_3$ for three faults.

This kind of dictionary is also named as the dependency matrix (or D-matrix), which is widely used for complex systems/ modular systems testability analysis and fault diagnosis. For each feature, a single discriminant rule is predetermined. If the feature value of a given fault state (including the fault-free state) satisfies the discriminant rule, the corresponding element value is set as "0". Otherwise, the element value is set as "1". Generally, the discriminant rule is preferred to be determined according to the fault-free state. Thus, element values of the fault-free state can be regularized as "0"s.

The binary-value dictionary is well suited for the cases whose features are from functional tests since the outcome of functional tests is either "pass" or "fail". To use the binary-value dictionary for fault diagnosis, the binary series of each row should be unique.

B. MULTI-VALUE DICTIONARY

In common with the binary-value dictionary, the multi-value dictionary also only uses a single discriminant rule for each feature. The difference is that the multi-value dictionary allows more than two values for each element. An example of the multi-value dictionary is shown in Table 2. Taking t_1 as an example, its value can be "0", "1", or "2".

The multi-value dictionary is suitable for cases in which each feature under different fault states only exists several countable values or intervals. It is widely used for analog circuit fault diagnosis, which is also known as the "integer-coded dictionary". In cases of the analog fault diagnosis, faults whose feature value intervals are similar are coded with the same integer such as "0", "1", or "2". To make the form consistent, the integer codes referring to the fault-free state are set as "0"s. The integer code series of each row should also be unique for fault diagnosis.

C. UNIQUE-VALUE DICTIONARY

For the first two types of the dictionary, a single discriminant rule is used to determine the element value for each feature (column) under each fault state. Instead of using

TABLE 3. Example of a Unique-Value Dictionary

	t_1	t_2	t_3
Fault-free	$B_{01}^- \sim B_{01}^+$	$B_{02}^- \sim B_{02}^+$	$B_{03}^- \sim B_{03}^+$
Fault 1	$B_{11}^- \sim B_{11}^+$	$B_{12}^- \sim B_{12}^+$	$B_{13}^- \sim B_{13}^+$
Fault 2	$B_{21}^- \sim B_{21}^+$	$B_{22}^- \sim B_{22}^+$	$B_{23}^- \sim B_{23}^+$
Fault 3	$B_{31}^- \sim B_{31}^+$	$B_{32}^- \sim B_{32}^+$	$B_{33}^- \sim B_{33}^+$

the single discriminant rule for each element, the unique-value dictionary expresses the element with a unique interval for each feature under each fault state. The unique-value dictionary is formed as Table 3.

For a given feature t_j under the i-th fault, its corresponding element is expressed as an interval between B_{ij}^- and B_{ij}^+. And the boundaries (B_{ij}^- and B_{ij}^+) of each interval can be any unique values.

Generally speaking, the third type dictionary can be regarded as a generalized form of the first two types. The binary-value dictionary tries to cluster faults into two groups with each feature, the multi-value dictionary tries to cluster faults into multi-groups with each feature, while the unique-value dictionary does not strictly attempt to cluster faults with a single feature. For fault diagnosis, the intervals of each feature are allowed to be overlapped with each other, as long as the feature vector overlapping area of any two fault states is small enough. Since the unique-value dictionary eases the requirement of each feature compared with the first two dictionaries, more features can be considered to construct the dictionary.

III. DIAGNOSIS USING DICTIONARY WITH URV-LDA REPRESENTATION

As mentioned in the Introduction, the soft faults and the feature intervals overlapping issues are two critical problems for dictionary-based fault diagnosis. In this part, the developed new type of dictionary is described, which applies the linear discriminant analysis and the unit residual signal vector for dictionary construction to solve the overlapping issue and the soft fault issue, respectively.

A. LINEAR DISCRIMINANT ANALYSIS

The linear discriminant analysis (LDA) derived from Fisher's linear discriminant, is used to separate different classes by optimizing a series of linear combinations of original features. The LDA is a supervised classification method since it requires labeled feature samples.

The purpose of the LDA is to maximize the between-class sample distance while minimizing the within-class sample distance. Taking the two-class issue as an example, these distances can be expressed in terms of the scatter matrices as follows:

$$\begin{cases} S_b(x) = (\mu_0 - \mu_1)(\mu_0 - \mu_1)^T \\ S_w(x) = \sum_{i=0,x\in X_i}^{1} (x-\mu_i)(x-\mu_i)^T \end{cases} \tag{1}$$

where $S_b(x)$ represents the between-class scatter matrix, $S_w(x)$ represents the within-class scatter matrix, x denotes the $M \times 1$ feature vector of each sample, X_i denotes the sample set

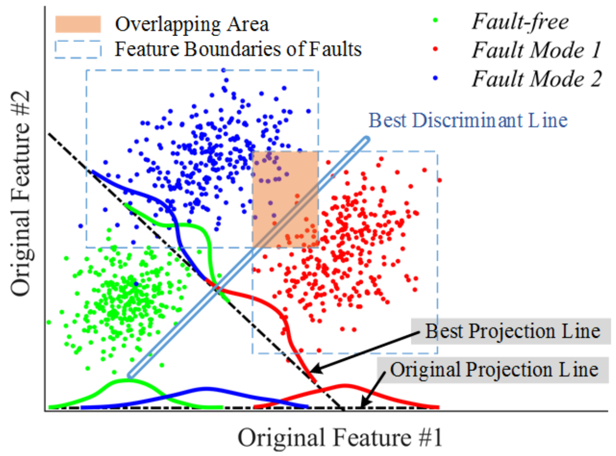

FIGURE 1. LDA used for fault diagnosis with single degradation level.

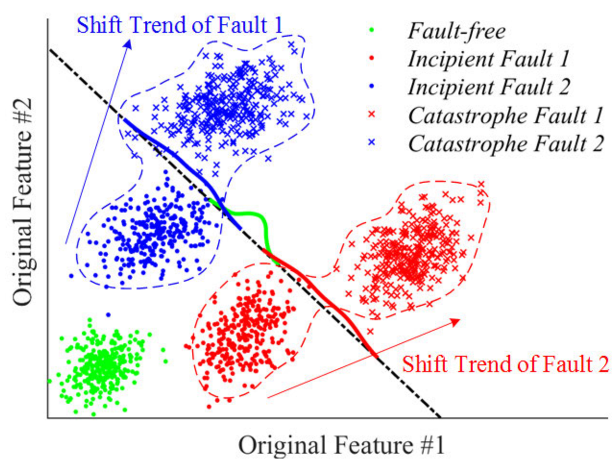

FIGURE 2. LDA used for fault diagnosis with multi-degradation levels.

of the i-th class, μ_0 and μ_1 represents the mean feature vector of two classes, respectively.

The scatter matrix ratio (S_b/S_w), which indicates the discriminant level, is a deterministic matrix for a given feature sample set X. Using an $M \times 1$ non-zero vector w, a new feature can be obtained as $y = w^T x$. Then, the scatter matrix ratio with y can be maximized by optimizing w as (2).

$$\arg\max_{w} J(w) = \frac{S_b(y)}{S_w(y)} = \frac{w^T S_b(x)\, w}{w^T S_w(x)\, w} \tag{2}$$

According to the generalized Rayleigh quotient theory, $J(w)$ reaches its maximum value $\lambda_{\max}$ which is also the biggest eigenvalue of the matrix $S_w(x)^{-1} S_b(x)$, only when w is the corresponding eigenvector.

The advantages of the LDA can be shown in Fig. 1. For the original features of the fault mode 1 and the fault mode 2, the feature distributions defined from the dictionary-based methods overlap heavily. The samples located in the orange overlapping area cannot be identified. Thus, it is hard to find proper discriminant rules to separate the fault mode 1 (or 2) from the other classes. However, by projecting original feature samples on a proper line (the best projection line in Fig. 1), the distribution overlap significantly decreases.

To implement the LDA for fault diagnosis of soft faults with multi-levels, a simple idea is to label each sample by its fault mode as shown in Fig. 2.

B. COMBINATION OF RESIDUAL SIGNAL VECTOR AND LDA

Although the LDA method can improve the fault diagnosis performance of the dictionary-based methods, it still has drawbacks. The LDA only aims at separating different classes, which still lacks the ability to indicate the similarity of classes. Thus, the LDA cannot be directly used for the soft fault diagnosis, since different soft faults with different degradation levels may have similar feature vector intervals.

To address this issue, the primary features should be preprocessed. The residual signal, which refers to the difference between the measured feature and the predicted feature of a given signal, is also a well-known health indicator for fault

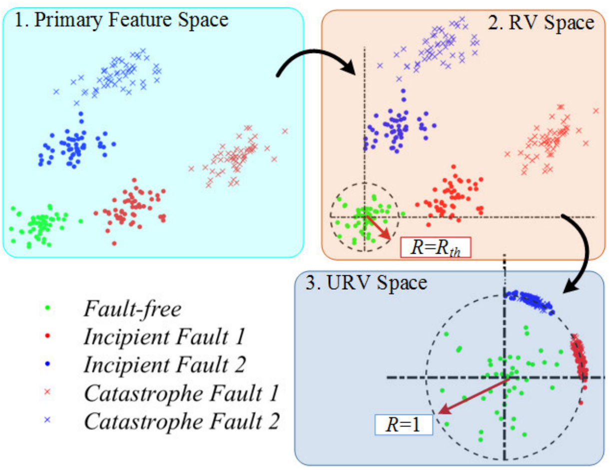

FIGURE 3. Unit residual signal vector transformation process.

diagnosis and prognosis. When the residual signal exceeds a preset threshold, the system is determined as faulty. For a given feature vector x, the normalized residual signal vector (RV) x' can be obtained as follows:

$$x' = \frac{x - \mu_0}{\sigma_0} \tag{3}$$

where μ_0 and σ_0 is the mean feature vector and the standard deviation vector of the fault-free state, respectively. After the transformation, the primary feature vectors are mapped into the RV space as shown in Fig. 3.

To further shrink the RV space into a unit area and eliminate the degradation level influence, a scale parameter $L(x')$ which is shown in (4) can be used.

$$L(x') = \|x'\|_2 = \sqrt{(x_1')^2 + \cdots + (x_n')^2} \tag{4}$$

where $||x'||_2$ denotes the $l2$ norm of x'.

Then, the unit residual signal vector (URV) with a given radius parameter R_{th} can be obtained as:

$$x''(R_{th}) = \begin{cases} L(x')^{-1} \cdot x' & L(x') > R_{th} \\ x' & L(x') \le R_{th} \end{cases} \tag{5}$$

After transforming to the URV space, all the samples fall in a hypersphere centered at the origin as shown at the lower right corner in Fig. 3. By adjusting R_{th} to a proper value, most of the fault-free samples can be mapped to the inside of the hypersphere, while most of the faulty samples are mapped on the surface. The adjustment of R_{th} can be objected as:

$$\begin{cases} \arg\max\limits_{R_{th}} \left(\min(L^{+}, K_R) + \max(L^{-}, K_R)\right) \\ L^{+} = \{\left|L(x') - R_{th}\right|, x' \in X_0'\} \\ L^{-} = \{\left|L(x') - R_{th}\right|, x' \notin X_0'\} \end{cases} \tag{6}$$

where $\min(A, K_R)$ and $\max(A, K_R)$ represents the K_R-smallest scale parameter values and the K_R-largest scale parameter values of the sample set A, respectively.

Since the URV keeps the feature growth trends of each fault and eliminates the effect of the degradation, soft fault diagnosis with multi-levels can be regarded as the typical classification problem. Thus, the LDA can be applied to improve the fault diagnosis accuracy by using the URV as inputs.

C. TWO LDA STRUCTURES FOR DICTIONARY CONSTRUCTION

After transforming the URV samples of each fault mode with LDA into a new feature space, the fault dictionary can be constructed using the new features. In the constructed dictionary, each element represents the expected feature interval which can be obtained from the decision function of the LDA. The LDA structure, which determines the transformation matrix, can significantly affect the classification performance of the new features. In this article, two LDA structures are applied for the fault dictionary construction. One is the combination of a series of typical 2-class LDA which is specified as the "multiple LDA", and the other is a single LDA for multi-classification which is specified as the "single LDA".

1) DICTIONARY CONSTRUCTION WITH MULTIPLE LDA

The multiple LDA (MLDA) structure can be regarded as a combination of a series of 2-class LDA, of which each follows the "one-against-the rest" rule. For each fault mode, a typical 2-class LDA, which is named as a local LDA, is used to discriminate the corresponding fault mode from the others. The transforming vector of each local LDA is assembled to be the global transforming matrix w. The local LDAs follow the same rule as described in Section III (A). Thus, the scatter matrices ($S_w(x)$ and $S_b(x)$) in (1) can be used.

Since the RV transformation described in Section III (B) has already separated feature samples of the fault-free state and that of the faulty states as two parts (with an R_{th}-radial hypersphere), the fault-free state does not need to be considered again during each local LDA solving process. Then, the scatter matrices for the j-th local LDA can be adjusted as:

$$\begin{cases} S_b^{\#}(x; j) = (\mu_j - \tilde{\mu}_j)(\mu_j - \tilde{\mu}_j)^T, \tilde{\mu}_j = \sum\limits_{i=1, i\neq j}^{M-1} \dfrac{\mu_i}{M-2} \\ S_w^{\#}(x) = \sum\limits_{i=1}^{M-1} \left[\sum_{x\in X_i} (x-\mu_i)(x-\mu_i)^T\right] \end{cases} \tag{7}$$

where $S_b^{\#}(x; j)$ is the between-class scatter matrix, $S_w^{\#}(x)$ is the within-class scatter matrix. M is the total class number, X_i denotes the sample set of the i-th fault mode, and μ_i denotes the mean feature vector of X_i.

Using (7), the solution of the multiple LDA for a multi-classification problem can be derived as:

$$\begin{cases} w_{\#} = [w_1, w_2, \cdots, w_{M-1}] \\ w_j : \{A_j w_j = \lambda_{\max} w_j, \lambda_{\max} = eig(A, 1)\} \\ A_j = S_w^{\#}(z)^{-1} S_b^{\#}(z), z = x''(R_{th}) \end{cases} \tag{8}$$

where $eig(A, 1)$ denotes the maximum eigenvalue of a given matrix A.$S_b^{\#}(x; j)$ is the between-class scatter matrix, and $S_w^{\#}(x)$ is the within-class scatter matrix.

The new feature vector can be obtained as $y = w_{\#}^T z$. Since each vector w_j is directly used to discriminate the j-th fault mode, each new feature only requires a single boundary for discrimination. The boundary B_j of the j-th feature y_j can be determined by:

$$B_j = \begin{cases} \dfrac{\max(Y_j^-, K_B) + \min(Y_j^+, K_B)}{2K_B}, & E(Y_j^+) > E(Y_j^-) \\ \dfrac{\max(Y_j^+, K_B) + \min(Y_j^-, K_B)}{2K_B}, & E(Y_j^+) < E(Y_j^-) \end{cases} \tag{9}$$

where $\max(Y, K_B)$ and $\min(Y, K_B)$ denotes the K_B-largest values and the K_B-smallest values of a given sample set Y, respectively. Y_j^+ is the feature set of j-th fault, and Y_j^- is the feature set of the other faults.

Since each new feature is responsible for discriminating a single fault, the URV-MLDA dictionary is sparse and can be regarded as the binary-value dictionary.

2) DICTIONARY CONSTRUCTION WITH SINGLE LDA

Besides using an MLDA structure to solve the multi-classification problem, a single LDA (SLDA) can also achieve the goal. Compared with the MLDA, the SLDA follows the "one-against-the rest" rule. However, the scatter matrix ratio cannot be calculated directly, since the scatter matrices may not be scalars for multi-classification. To obtain the optimized w, the objective of the scatter matrix ratio should be modified as:

$$\arg\max_{w} J(w) = \frac{\prod diag\left(w^T S_b(x) w\right)}{\prod diag\left(w^T S_w(x) w\right)} = \prod_{j=1}^{D} \frac{w_j^T S_b(x) w_j}{w_j^T S_w(x) w_j} \tag{10}$$

where $diag(A)$ refers to the main diagonal of a given matrix A, D refers to the dimension of the expected output vector. The scatter matrix ratio is approximately expressed to be a ratio of the products of two scatter matrices. It can be further expressed as the product of a series of generalized Rayleigh quotients as shown in the second line of (10). Thus, the optimized w can be determined as the corresponding eigenvectors of the D-largest eigenvalues of $S_w(x)^{-1}S_b(x)$.

With the "one-against-the rest" rule, the scatter matrices $S_w(x)$ and $S_b(x)$ should also be modified as follows:

$$\begin{cases} S_b^*(x) = \sum_{i=1}^{M-1}\left[(\mu_i - \mu_X)(\mu_i - \mu_X)^T\right] \\ S_w^*(x) = \sum_{i=1}^{M-1}\left[\sum_{x\in X_i}(x-\mu_i)(x-\mu_i)^T\right] \end{cases} \tag{11}$$

where μ_X denotes the mean feature vector of the entire sample set X. With the above modification, the distance between each class center and the global center is calculated to evaluate the between-class sample distance.

Thus, the solution of the single LDA can be described as:

$$\begin{cases} w_* = [w_1, \cdots, w_j, \cdots, w_D], 1 \le D \le M-2 \\ w_j = \left\{v : Av = \lambda_j v, \lambda_j = eig(A, j)\right\} \\ A = S_w^*(z)^{-1} S_b^*(z), z = x''(R_{th}) \end{cases} \tag{12}$$

where $eig(A, j)$ denotes the j-th largest eigenvalue of a given matrix A.

For the new $D \times 1$ feature vector $y = w_*^T z$, the distributions of a given feature under different faults may be overlapped with each other. According to the review in Section II, the URV-SLDA dictionary should be categorized as the unique-value dictionary. Thus, the boundaries of each feature under each fault state should be determined respectively as:

$$\begin{cases} B_{ij}^- = K_B^{-1} \times \min\left(Y_{ij}, K_B\right) \\ B_{ij}^+ = K_B^{-1} \times \max\left(Y_{ij}, k\right) \end{cases} \tag{13}$$

where $\max(Y_{ij}, K_B)$ and $\min(Y_{ij}, K_B)$ denotes the K_B-largest values and the K_B-smallest values of a given sample set Y_{ij}, Y_{ij} denotes the j-th transformed feature samples under i-th fault. And B_{ij}^- and B_{ij}^+ denote the lower boundary and the upper boundary of the j-th feature under i-th fault, respectively.

D. FAULT DIAGNOSIS PROCEDURE OF URV-LDA DICTIONARIES

The training and diagnosis procedure using the URV-LDA transformed dictionary can be shown in Fig. 4.

The purpose of the training process is to determine the transforming process parameters using given training samples and training hyper-parameters. The transforming process parameters include the URV transformation parameters μ_0, δ_0, and R_{th}, and the LDA transforming matrix w ($w_\#$ or w_*). After obtaining the new features, the dictionary boundaries can also be determined.

Once the transformation process is determined, the fault diagnosis can be achieved. The diagnosis can be separated into the distance-based fault detection step and the pattern-based fault identification step. The fault detection

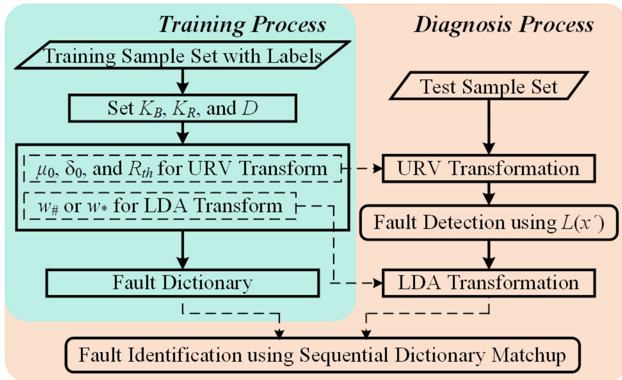

FIGURE 4. The training and diagnosis process with the developed URV-LDA dictionary(s).

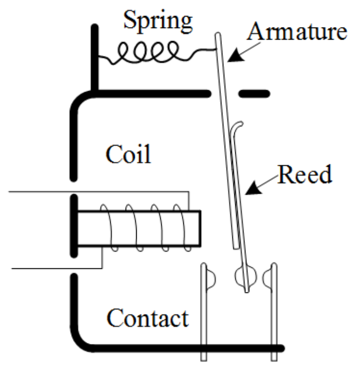

FIGURE 5. Structure of the electromagnetic relay.

step is realized by comparing $L(x')$ and R_{th}, the sample whose $L(x')$ is smaller than R_{th} is regarded as the fault-free sample. Samples with $L(x')$ greater than R_{th} are further be LDA transformed for fault identification. The fault identification is realized by matching up the feature vector with each row of the fault dictionary.

IV. VALIDATION EXPERIMENT WITH THE RAILWAY SIGNAL ELECTROMAGNETIC RELAY

To validate the effectiveness of the developed dictionary, the fault seeding experiment of an electromagnetic relay is carried out. Four fault modes of the relay are simulated, and three different degradation levels are considered. Two developed methods are compared with two conventional diagnostic methods.

A. FAULT SEEDING AND DATA ACQUISITION DESCRIPTION

The electromagnetic relay used in the fault seeding experiment is a railway signal relay, the simplified structure of which is shown in Fig 5.

As the relay used in the railway system, the degradation of the coil, the spring, the reed, and the armature can affect the performance of the relay, which may further affect the signal transmission for the entire system. There are four faults related to the parts degradation, which are the spring loosing, the armature stuck, the coil resistance increasing, and the reed fatigue. These four faults can be simulated by decreasing the spring preload, inserting thin spacers to the armature gap,

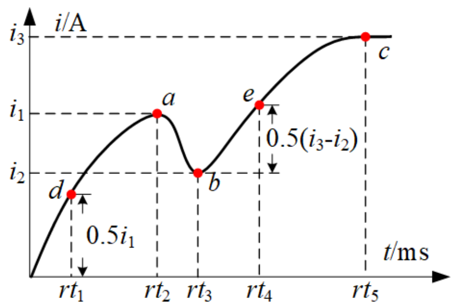

FIGURE 6. Driving Curve of the Electromagnetic Relay.

increasing drive resistance, and modifying elastic modulus, respectively. Also, the degradation level of these faults can be controlled.

According to the theory analysis, the relay faults can affect the shape of the driving signal. Fig. 6 shows the driving current curve, which is a non-monotonic curve. To extract the shape information from the curve, the time, and the current value of five specific points shown in Fig. 6 are used.

Thus, eight physical features (rt_1 $\sim rt_5$ and i_1 $\sim$ i_3 as shown in Fig. 6) are determined as the primary features. Four fault modes with three degradation levels, as well as the fault-free state, are simulated, and each state includes 100 samples. Thus, $13 \times 100 = 1300$ samples are obtained from the relay fault seeding experiment.

B. EFFECT ANALYSIS OF TUNABLE PARAMETERS

For the developed two dictionaries, tunable parameters include K_R, K_B, and D. To analyze the influence of these tunable parameters, only feature samples of the fault-free and the incipient faults (faults whose degradation level is relatively low) are used for training. And all samples including multi-level faults are used to calculate the diagnostic accuracy. The diagnostic accuracy is calculated as:

$$Acc = N_C \times N^{-1} \times 100\% \tag{14}$$

where N_C is the number of correctly classified samples, and N is the total number of samples.

Taking a fixed combination of $K_R = 3$, $K_B = 3$, and $D = 3$ for training, the obtained new feature space as well as the corresponding element values of the two developed dictionaries are shown in Fig. 7, Fig. 8, and Table 4, Table 5, respectively.

According to the definition of K_R, it mainly affects the discriminant result between fault states and fault-free state, which is fault detection ability. Given fixed values of $K_B = 5$ and $D = 3$, the effect of K_R is analyzed by changing its value from 1 to 9. For each combination of tunable parameters, 20 randomly generated sample sets with 250 samples (50 samples of each state × 5 fault-free/fault states) for each are used for training, and the diagnostic results are shown in Fig. 9 in the form of boxplots.

As seen in the above results, the diagnostic accuracy of both dictionary increase as the value of K_R increase from 1 to 5. However, when K_R keeps increasing, the diagnostic

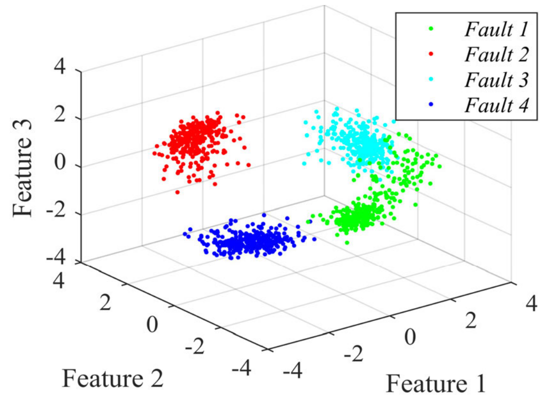

FIGURE 7. Scatter plots of URV-SLDA features.

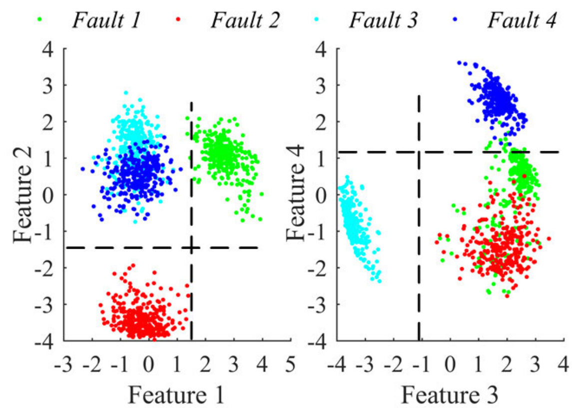

FIGURE 8. Scatter plots of URV-MLDA features.

TABLE 4. URV-SLDA Dictionary Trained With 100 Samples

	t_1	t_2	t_3
Fault 1	−1.42 ~ −0.11	1.67 ~ 3.29	−2.51 ~ −0.43
Fault 2	−2.07 ~ −0.34	−3.06 ~ −2.58	−1.75 ~ −0.83
Fault 3	2.97 ~3.70	−1.04 ~ 0.15	−0.28 ~ −0.99
Fault 4	−2.68 ~−1.92	−0.42 ~0.87	1.27 ~ 1.84

TABLE 5. URV-MLDA Dictionary Trained With 100 Samples

	t_1	t_2	t_3	t_4
Fault 1	> 1.57	> −1.47	> −1.05	< 1.13
Fault 2	< 1.57	< −1.47	> −1.05	< 1.13
Fault 3	< 1.57	> −1.47	< −1.05	< 1.13
Fault 4	< 1.57	> −1.47	> −1.05	> 1.13

accuracy stops increasing, which indicates the influence of K_R mainly exists when its value is relatively small. Moreover, the effect on URV-SLDA is heavier than that of URV-MLDA.

For K_B, it is related to the boundary determination. Given fixed values of $K_R = 5$ and $D = 3$, the effect of K_B is analyzed by changing its value from 1 to 9, and the diagnostic results are shown in Fig. 10 with the form of boxplots. Compared with K_R, the influence of K_B is more complicated.

As K_B increases, the diagnostic accuracy of URV-SLDA decreases, while that of URV-MLDA increases first and

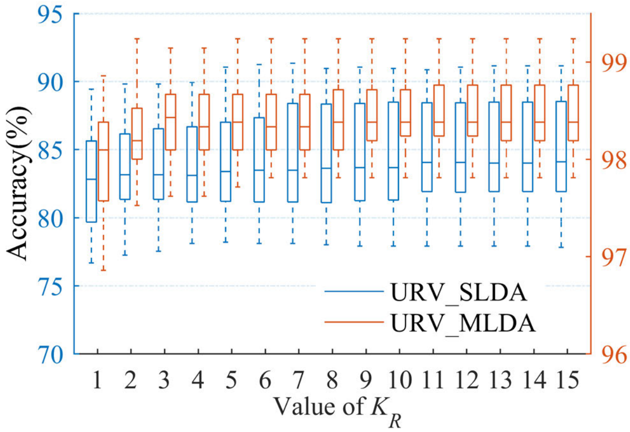

FIGURE 9. Diagnostic accuracy results with different K_R (trained with only fault-free and incipient faults samples).

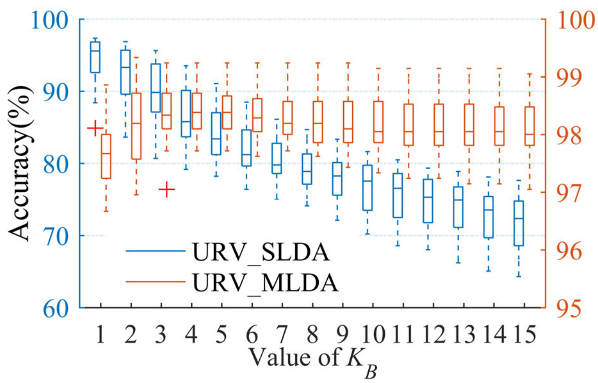

FIGURE 10. Diagnostic accuracy results with different K_B (trained with only fault-free and incipient faults samples).

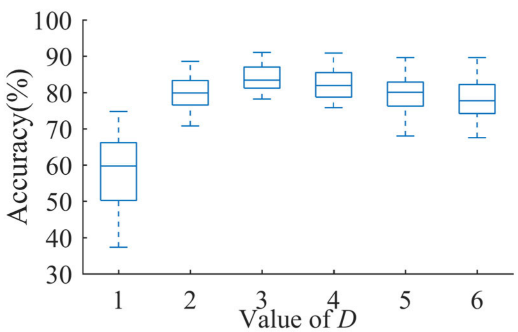

FIGURE 11. Diagnostic accuracy results with different D (trained with only fault-free and incipient faults samples).

gradually decreases. The optimal value of K_B for URV-MLDA ranges from 3 to 5.

The tunable parameter D only exists in the SLDA dictionary, by changing it from 1 to 4 (with $K_R = 5$ and $K_B = 5$), the influence of it can be seen from Fig. 11. As seen in Fig. 11, the optimal D ranges from 2 to 3.

Compared with K_R, the influence of K_B and D is heavier. Although a fixed combination of tunable parameters can be arbitrarily given, an optimization process for tunable parameters is suggested to get a better diagnostic ability.

TABLE 6. Diagnostic Ability Comparison Result (20 Sample Sets)

Methods		Tunable Parameters
M1	Sparse LDA	Lasso penalty factor λ
M2	SVM	Scaling factor of RBF kernel γ; penalty factor c
M3	URV-SLDA	K_R; K_B; D
M4	URV-MLDA	K_R; K_B

TABLE 7. Diagnostic Ability Comparison Result (20 Sample Sets)

		Worst	Best	Average
M1	Sparse LDA	90.5%	94.3%	93.0%
M2	SVM	96.6%	98.6%	97.9%
M3	URV-SLDA	88.6%	94.9%	91.6%
M4	URV-MLDA	99.0%	99.7%	99.3%

C. DIAGNOSTIC ABILITY COMPARISON

To validate the ability to solve the soft fault diagnosis issue, two developed dictionaries are compared with two kinds of conventional diagnostic methods, which are sparse LDA, and SVM.

The sparse LDA for comparison is constructed with a lasso penalty to get better classification ability. The SVM in this part uses the one-against-the-rest strategy and the sequential minimal optimization (SMO) algorithm with a radial basis function (RBF) kernel. To get better classification ability, three-fold cross validation and the grid search method are applied to optimize tunable parameters of these methods. The tunable parameters required to be optimized of these methods are listed as follows.

For each method, 20 sample sets are randomly generated as training samples. For each sample set, fault-free state and four fault modes with three degradation levels (incipient, mid-range & catastrophic levels) are considered. 50% samples for each state ($50\% \times 1300 = 650$ samples in total) are used as training & validation samples, and the rest are used for the test. The results are listed as follows.

As can be seen in Table 7, the proposed URV-MLDA dictionary has the highest averaged diagnostic accuracy, which is much better than that of sparse LDA. Although the result of SVM is 0.1% lower than URV-MLDA, its variation is slightly larger. On the other hand, the URV-SLDA is lower than others. By changing the sample rate from 50% down to 10%, the averaged diagnostic accuracy of each training sample rate is shown in Fig. 12. When the sample rate decreases to 10%, the averaged accuracy of URV-MLDA still keeps at 98.4%, which is 5.3% higher than that of SVM and is 12.8% higher than that of sparse LDA.

When training samples are limited to include only incipient & mid-range levels or only incipient level, averaged diagnostic accuracy results can be seen in Fig. 13 and Fig. 14, respectively. The result of URV-MLDA still keeps higher than 95%, however, those of SVM and sparse LDA all decrease to less than 80%. It indicates when degradation tendency information is limited, the diagnostic ability of SVM or sparse LDA cannot be guaranteed.

It should be noted, URV-SLDA also has better diagnostic accuracy than SVM and sparse LDA, when the degradation

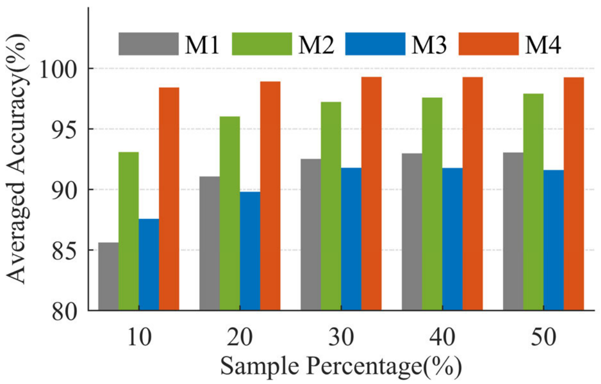

FIGURE 12. Averaged diagnostic accuracy results trained with incipient, mid-range & catastrophic faults samples (sample fault level = 3).

TABLE 8. Averaged Diagnostic Accuracy Comparison With Different Training Sample Sizes (Sample Fault Level = 3)

	10%	20%	30%	40%	50%
Sparse LDA	85.6%	91.1%	92.5%	93.0%	93.0%
SVM	93.1%	96.0%	97.2%	97.6%	97.9%
URV-SLDA	87.6%	89.8%	91.8%	91.8%	91.6%
URV-MLDA	98.4%	98.9%	99.3%	99.3%	99.3%

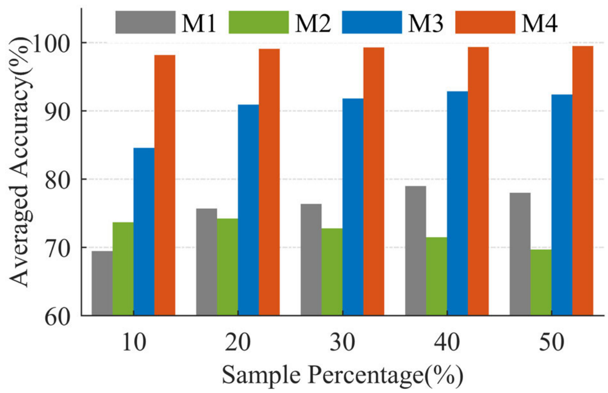

FIGURE 13. Averaged diagnostic accuracy results trained with incipient & mid-range faults samples (sample fault level = 2).

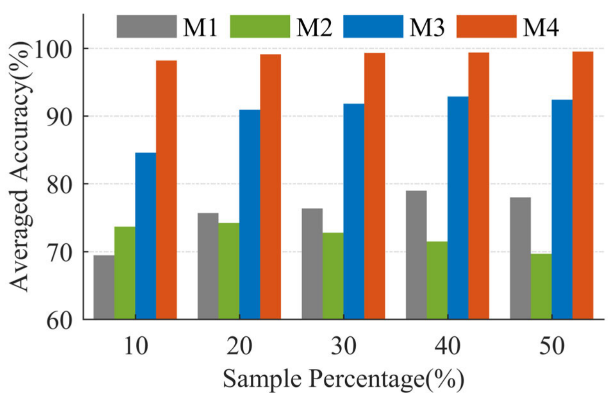

FIGURE 14. Averaged diagnostic accuracy results trained with only incipient faults samples (sample fault level = 1).

level of samples is limited. Thus, it can also be regarded as an option for fault diagnosis when the degradation situation of training fault samples is insufficient.

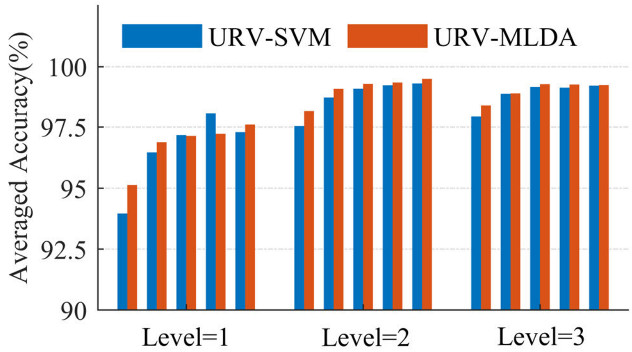

FIGURE 15. Diagnostic accuracy comparison between URV-SVM and URV-MLDA.

As mentioned above, the main reason why conventional nonlinear (SVM with RBF kernel) and linear (sparse LDA) classification methods perform worse is that they rely on the existing sample distribution, which does not consider the "expandability" of the distribution caused by degradation. The URV transformation process, however, can somehow compress the distribution differential caused by degradation. Thus, the combination of URV and LDA can perform better for the diagnosis of soft faults.

Actually, LDA can be regarded as a simple and linear form of SVM. Thus, the combination of URV and SVM should also be able to solve the soft fault diagnosis issue. As seen in Fig.15, the combination of URV and SVM also shows excellent diagnostic ability as URV-MLDA. The subtle difference can be explained by the improper overfitting and rough grid search optimization of SVM, which can be ignored case by case.

V. CONCLUSION

Dictionary-based methods are widely used for hard fault diagnosis of electromechanical systems. Unlike hard faults with deterministic states, states of degradation caused soft faults can be continuously changed. Without deterministic states, soft faults cannot be diagnosed using conventional dictionaries. To address this issue, this article develops a new type of dictionary — the URV-LDA dictionary, which is constructed with the unit residual signal vector (URV) and the linear discriminant analysis (LDA).

Compared with conventional linear or nonlinear transformation used for fault dictionaries, the developed URV-LDA transformation method first introduces the URV to perform the fault feature growth trends instead of the original feature distribution. Thus, the influence of continuous degradation on feature distribution can be eliminated. Based on URV transformation, two dictionaries named as the URV-MLDA binary-value dictionary and the URV-SLDA unique-value dictionary are proposed, and diagnostic accuracy for an electromagnetic relay product is compared with conventional SVM and sparse LDA methods. The results show that the developed URV-MLDA performs better diagnostic ability whether degradation trends are implied in samples or not. While URV-SLDA performs delicately better than two conventional methods, only when fault samples with multi-degradation levels are insufficient. It shows the developed dictionaries can solve the soft fault issue.

As a linear classifier, LDA used in the URV-LDA dictionary prefers to solve cases whose degradation trends are relatively linearly separable such as the electromagnetic relay, which requires simple structure and are robust to avoid over-fitting. To issues whose degradation trends are hard to be linearly separated, the combination of URV and nonlinear SVM can be further applied to construct a dictionary.

ACKNOWLEDGMENT

The authors would like to thank Prof. Michael Pecht, Dr. Michael Azarian, and Dr. Myeongsu Kang from the Center for Advanced Life Cycle Engineering (CALCE) at the University of Maryland for discussion.

REFERENCES

[1] Y. Lei, F. Jia, J. Lin, S. Xing, and S. X. Ding, "An intelligent fault diagnosis method using unsupervised feature learning towards mechanical big data," *IEEE Trans. Ind. Electron.*, vol. 63, no. 5, pp. 3137–3147, May 2016.
[2] 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.
[3] M. Yu, C. Xiao, W. Jiang, S. Yang, and H. Wang, "Fault diagnosis for electromechanical system via extended analytical redundancy relations," *IEEE Trans. Ind. Informat.*, vol. 14, no. 12, pp. 5233–5244, Dec. 2018.
[4] J. A. Carino, M. Delgado-Prieto, J. A. Iglesias, A. Sanchis, D. Zurita, M. Millan, J. A. O. Redondo, and R. Romero-Troncoso, "Fault detection and identification methodology under an incremental learning framework applied to industrial machinery," *IEEE Access*, vol. 6, pp. 49755–49766, 2018.
[5] M. Tadeusiewicz and S. Halgas, "A new approach to multiple soft fault diagnosis of analog BJT and CMOS circuits," *IEEE Trans. Instrum. Meas.*, vol. 64, no. 10, pp. 2688–2695, Oct. 2015.
[6] L. Abboud, A. Cozza, and L. Pichon, "A matched-pulse approach for soft-fault detection in complex wire networks," *IEEE Trans. Instrum. Meas.*, vol. 61, no. 6, pp. 1719–1732, Jun. 2012.
[7] S. Tian, C. Yang, F. Chen, and Z. Liu, "Circle equation-based fault modeling method for linear analog circuits," *IEEE Trans. Instrum. Meas.*, vol. 63, no. 9, pp. 2145–2159, Sep. 2014.
[8] X. Deng, X. Tian, S. Chen, and C. J. Harris, "Nonlinear process fault diagnosis based on serial principal component analysis," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 3, pp. 560–572, Mar. 2018.
[9] X. Deng, X. Tian, S. Chen, and C. J. Harris, "Deep principal component analysis based on layerwise feature extraction and its application to nonlinear process monitoring," *IEEE Trans. Control Syst. Technol.*, vol. 27, no. 6, pp. 2526–2540, Nov. 2019.
[10] L. Ren, W. Lv, S. Jiang, and Y. Xiao, "Fault diagnosis using a joint model based on sparse representation and SVM," *IEEE Trans. Instrum. Meas.*, vol. 65, no. 10, pp. 2313–2320, Oct. 2016.
[11] Y. Xu, Y. Sun, J. Wan, X. Liu, and Z. Song, "Industrial big data for fault diagnosis: Taxonomy, review, and applications," *IEEE Access*, vol. 5, pp. 17368–17380, 2017.
[12] H.-J. Ma and G.-H. Yang, "Observer-based fault diagnosis for a class of non-linear multiple input multiple output uncertain stochastic systems using B-spline expansions," *IET Control Theory Appl.*, vol. 5, no. 1, pp. 173–187, Feb. 2011.
[13] S. Zhai, W. Wang, and H. Ye, "Fault diagnosis based on parameter estimation in closed-loop systems," *IET Control Theory Appl.*, vol. 9, no. 7, pp. 1146–1153, Apr. 2015.
[14] W. G. Fenton, T. M. McGinnity, and L. P. Maguire, "Fault diagnosis of electronic systems using intelligent techniques: A review," *IEEE Trans. Syst., Man Cybern. C, Appl. Rev.*, vol. 31, no. 3, pp. 269–281, Aug. 2001.
[15] H.-J. Lee, D.-Y. Park, B.-S. Ahn, Y.-M. Park, J.-K. Park, and S. S. Venkata, "A fuzzy expert system for the integrated fault diagnosis," *IEEE Trans. Power Del.*, vol. 15, no. 2, pp. 833–838, Apr. 2000.
[16] Y. Zhi-Ling, W. Bin, D. Xing-Hui, and L. Hao, "Expert system of fault diagnosis for gear box in wind turbine," *Syst. Eng. Procedia*, vol. 4, pp. 189–195, Jan. 2012.
[17] S. Deb, K. R. Pattipati, V. Raghavan, M. Shakeri, and R. Shrestha, "Multi-signal flow graphs: A novel approach for system testability analysis and fault diagnosis," *IEEE Aerosp. Electron. Syst. Mag.*, vol. 10, no. 5, pp. 14–25, May 1995.
[18] I. Ghosh, A. Raghunathan, and N. K. Jha, "A design-for-testability technique for register-transfer level circuits using control/data flow extraction," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 17, no. 8, pp. 706–723, Aug. 1998.
[19] G. Niu and X. Huang, "Failure prognostics of locomotive electro-pneumatic brake based on bond graph modeling," *IEEE Access*, vol. 5, pp. 15030–15039, 2017.
[20] T. Golonek and J. Rutkowski, "Genetic-algorithm-based method for optimal analog test points selection," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 54, no. 2, pp. 117–121, Feb. 2007.
[21] C. Yang, S. Tian, B. Long, and F. Chen, "Methods of handling the tolerance and test-point selection problem for analog-circuit fault diagnosis," *IEEE Trans. Instrum. Meas.*, vol. 60, no. 1, pp. 176–185, Jan. 2011.
[22] V. C. Prasad and N. S. C. Babu, "Selection of test nodes for analog fault diagnosis in dictionary approach," *IEEE Trans. Instrum. Meas.*, vol. 49, no. 6, pp. 1289–1297, Dec. 2000.
[23] B. Bosio, P. Girard, S. Pravossoudovitch, and A. Virazel, "A comprehensive framework for logic diagnosis of arbitrary defects," *IEEE Trans. Comput.*, vol. 59, no. 3, pp. 289–300, Mar. 2010.
[24] C. Yang, J. Yang, Z. Liu, and S. Tian, "Complex field fault modeling-based optimal frequency selection in linear analog circuit fault diagnosis," *IEEE Trans. Instrum. Meas.*, vol. 63, no. 4, pp. 813–825, Apr. 2014.
[25] J. A. Starzyk, D. Liu, Z.-H. Liu, D. E. Nelson, and J. Rutkowski, "Entropy-based optimum test points selection for analog fault dictionary techniques," *IEEE Trans. Instrum. Meas.*, vol. 53, no. 3, pp. 754–761, Jun. 2004.
[26] I. Pomeranz and S. M. Reddy, "On dictionary-based fault location in digital logic circuits," *IEEE Trans. Comput.*, vol. 46, no. 1, pp. 48–59, Jan. 1997.
[27] C. Liu and K. Chakrabarty, "Compact dictionaries for fault diagnosis in scan-BIST," *IEEE Trans. Comput.*, vol. 53, no. 6, pp. 775–780, Jun. 2004.
[28] Y. Yan, P. B. Luh, and K. R. Pattipati, "Fault diagnosis framework for air handling units based on the integration of dependency matrices and PCA," in *Proc. IEEE Int. Conf. Autom. Sci. Eng. (CASE)*, Taipei, Taiwan, Aug. 2014, pp. 1103–1108.
[29] S. Zhang, K. R. Pattipati, Z. Hu, and X. Wen, "Optimal selection of imperfect tests for fault detection and isolation," *IEEE Trans. Syst., Man, Cybern. Syst.*, vol. 43, no. 6, pp. 1370–1384, Nov. 2013.
[30] B. Gou, X. Ge, S. Wang, X. Feng, J. B. Kuo, and T. G. Habetler, "An open-switch fault diagnosis method for single-phase PWM rectifier using a model-based approach in high-speed railway electrical traction drive system," *IEEE Trans. Power Electron.*, vol. 31, no. 5, pp. 3816–3826, May 2016.
[31] J. Zhou, Y. Yang, S. X. Ding, Y. Zi, and M. Wei, "A fault detection and health monitoring scheme for ship propulsion systems using SVM technique," *IEEE Access*, vol. 6, pp. 16207–16215, 2018.
[32] W. Yu and C. Zhao, "Sparse exponential discriminant analysis and its application to fault diagnosis," *IEEE Trans. Ind. Electron.*, vol. 65, no. 7, pp. 5931–5940, Jul. 2018.
[33] C.-N. Li, Y.-H. Shao, W. Yin, and M.-Z. Liu, "Robust and sparse linear discriminant analysis via an alternating direction method of multipliers," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 31, no. 3, pp. 915–926, Mar. 2020.
[34] L. H. Chiang, M. E. Kotanchek, and A. K. Kordon, "Fault diagnosis based on Fisher discriminant analysis and support vector machines," *Comput. Chem. Eng.*, vol. 28, no. 8, pp. 1389–1401, Jul. 2004.
[35] J. Harmouche, C. Delpha, and D. Diallo, "Improved fault diagnosis of ball bearings based on the global spectrum of vibration signals," *IEEE Trans. Energy Convers.*, vol. 30, no. 1, pp. 376–383, Mar. 2015.
[36] X. Jin, M. Zhao, T. W. S. Chow, and M. Pecht, "Motor bearing fault diagnosis using trace ratio linear discriminant analysis," *IEEE Trans. Ind. Electron.*, vol. 61, no. 5, pp. 2441–2451, May 2014.
[37] C. P. Mboo and K. Hameyer, "Fault diagnosis of bearing damage by means of the linear discriminant analysis of stator current features from the frequency selection," *IEEE Trans. Ind. Appl.*, vol. 52, no. 5, pp. 3861–3868, Sep. 2016.

CEN CHEN (Member, IEEE) was born in 1992. He received the Ph.D. degree in electrical engineering from the Harbin Institute of Technology, Harbin, China, in 2019.

From October 2017 to October 2018, he was a Visiting Research Scholar with the Center for Advanced Life Cycle Engineering (CALCE), University of Maryland at College Park, College Park, MD, USA. He is currently a Lecturer with the Harbin Institute of Technology. His research interests include condition monitoring, fault diagnosis and prognosis, testability design techniques for power electronics systems, and electrical apparatus.

YUN YANG received the B.S. degree in electrical engineering from the Harbin Institute of Technology, Harbin, China, in 2020, where he is currently pursuing the master's degree with the Department of Electrical Engineering. His research interests include failure analysis and reliability design for electronic devices and systems.

XUERONG YE (Senior Member, IEEE) received the B.S., M.S., and Ph.D. degrees in electrical engineering from the Harbin Institute of Technology, Harbin, China, in 2003, 2005, and 2009, respectively. He is currently a Professor with the Department of Electrical Engineering, Harbin Institute of Technology. His research interests include failure analysis and reliability design for electronic devices and systems.

GUOFU ZHAI (Member, IEEE) received the Ph.D. degree from the Harbin Institute of Technology, Harbin, China, in 1998. He is currently a Professor with the Department of Electrical Engineering, Harbin Institute of Technology. He is also the Director of the Center for Electrical Apparatus and Electronic System Reliability Engineering, Harbin Institute of Technology. He has published more than 40 peer-reviewed journal articles. His research interest includes reliability and testing techniques of electronic devices and systems.