International Journal of Control, Automation, and Systems 21(9) (2023) 2797-2809
http://dx.doi.org/10.1007/s12555-022-0241-2

ISSN:1598-6446 eISSN:2005-4092
http://www.springer.com/12555

Just-in-time Learning-aided Nonlinear Fault Detection for Traction Systems of High-speed Trains

Chao Cheng, Xiuyuan Sun, Junjie Shao, Hongtian Chen*, and Chao Shang

Abstract: Traction systems in high-speed trains exhibit significant dynamic characteristics, which mainly arise from operation-point changes. Most existing fault detection methods provide static data models for global structures, especially for traditional multivariate statistical analysis methods constrained by constant operating points. The symptoms of incipient faults are slight and easily hidden. Despite the moderate effect of incipient faults, they will compromise the overall performance and remaining life of traction systems in the long run. Therefore, a just-in-time slow feature analysis method is proposed in this study. The salient advantages of the proposed method are: 1) It can be applied to dynamic non-linear systems; 2) It can detect incipient faults subject to environments containing noise and unknown disturbances; 3) It mitigates false alarms caused by parameter mutation during mode-switching. A series of experiments are carried out on a traction system platform to verify the effectiveness and superiority of the proposed method.

Keywords: Incipient faults, just-in-time learning, multi-mode, slow feature analysis, traction systems.

1. INTRODUCTION

With the rapid development of global transportation, high-speed trains provide a solid boost by virtue of their fast speeds and large carrying capacities [1]. As a consequence, the safety and reliability of high-speed trains have received significant attention from researchers in recent years. Traction systems are recognized as central of high-speed trains, which take the traction motor as the control object [2,3]. Thus, if faults in traction systems can not be dealt with in time, the operation of high-speed trains may be disrupted.

The timely detection of incipient faults in traction systems provides an essential guarantee for the operation of high-speed trains. Traction systems, consisting of a large number and various types of components, are vulnerable to incipient faults [4]. For example, under high-temperature conditions, the aging of insulated gate bipolar transistors will accelerate and gradually evolve into faults. Furthermore, short circuits in stator turns and performance degradation of electrolytic capacitors in the DC link are also typical faults in traction systems. Because various components in the system are closely coupled, a fault in one component may migrate and spread over the system along the signal flow direction, which will affect the performance of other components [5,6]. In addition, the operating states of high-speed trains can be divided into start-up acceleration, smooth operation, deceleration when cornering, and braking, which are switched according to the actual operation. The actual operation of high-speed trains makes dynamic nonlinearity an inherent characteristic of traction systems [7,8].

In order to improve the detectability of incipient faults, previous efforts have been focused on the following three aspects [9]: model-based, signal analysis-based, and data-driven methods. The amplitude, frequency, and phase of the gathered signals, which are related to the operation of traction systems, could be utilized to extract specific fault features [10]. For instance, a fault diagnosis approach based on fractional steepest ascent Morlet wavelet transform is developed in [11], which enables identifying faults with small magnitudes in the noise environment. However, prior knowledge of the symptoms in different faults is essential for analyzing the signals collected. The workflow of the model-based methods can be divided into system modeling, residual generation, and

Manuscript received March 18, 2022; revised August 2, 2022 and December 17, 2022; accepted January 13, 2023. Recommended by Associate Editor Jiuxiang Dong under the direction of Senior Editor Bin Jiang. This work was supported by National Natural Science Foundation of China (61903047, U20A20186) and Jilin Science and Technology Department (20200401127GX).

Chao Cheng and Xiuyuan Sun are with the School of Computer Science and Engineering, Changchun University of Technology, Changchun 130012, China (e-mails: {chengx415, sunxy159}@163.com). Junjie Shao is with Data Research Office, CRRC Changchun Railway Vehicles Co., Ltd., Changchun 130062, China (e-mail: shaojunjie@cccar.com.cn). Hongtian Chen is with the Department of Chemical and Materials Engineering, University of Alberta, Edmonton, AB T6G 2V4, Canada (e-mail: chtbaylor@163.com). Chao Shang is with the Department of Automation, Beijing National Research Center for Information Science and Technology, Tsinghua University, Beijing 100084, China (e-mail: c-shang@tsinghua.edu.cn).
* Corresponding author.

residual assessment [6]. For example, Gou *et al*. [12] proposed a dynamic model-based open-switch fault detection approach for rectifiers. Only the switch command signal, the measured current, and DC-link voltage have been employed in the model. The detection result is robust against varying grid voltage and other loads. To account for the non-stationarity of traction rectifiers, a multi-input-multi-output evolution model with dimensionality reduction is developed in [13]. This model enables flexible multivariable control while ensuring that coupling faults do not destabilize the closed loop. Nevertheless, due to the increasing complexity of the system structure, accurate mathematical models are difficult to obtain, which will limit the application of the established fault detection models.

Thanks to the allocation of numerous sensors in traction systems, data are easy to get and store. Data-driven approaches are popular because they do not rely on a priori knowledge [14]. Among them, the methods based on multivariate analysis are of great significance in practice, such as principal component analysis, partial least squares estimation, and canonical correlation analysis [1]. Nevertheless, traditional multivariate analysis methods can only perform ideally in static linear systems [15]. For this reason, improved multivariate analysis methods have been enjoyed popularity. In recent efforts, including the deep-slow feature analysis developed in [16], the original data are separated into six modes, based on which feature extraction is performed using slow feature analysis. In order to satisfy the nonlinear characteristics of traction systems, a linear-nonlinear hybrid model was proposed in [17]. The hybrid model used the principal component analysis and the kernel principal component analysis to extract the traction system's linear and nonlinear features. Furthermore, a new detection method was proposed in [18], considering the limitations of canonical correlation analysis for Gaussian assumptions. The new method addressed this limitation by combining canonical correlation analysis with randomized algorithm-based threshold-setting. Besides, a distributed detection method based on fog computing is designed in [19]. The fault detection and control of each node are realized by an adaptive configuration method. However, these methods only achieve the desired results when the traction system operates stably in a single mode. They do not take into account the effect of operating state-switching on fault detection [20].

The state-switching during the actual operation of a traction system makes multi-mode an inherent characteristic. The switching of operating modes is accompanied by mutations in state parameters. Therefore, fault detection methods based on a single steady state are prone to false fault alarms. The inherent dynamic character of traction systems makes it challenging to achieve the desired results with fault detection methods constrained by static conditions. In addition, the slight symptoms of incipient faults are difficult to trigger alarms. The complex operating environment of traction systems also leads to many external disturbances, such as aerodynamic problems and electromagnetic interference, which can easily cover incipient faults. As a result, existing fault detection methods have many limitations in facing these problems.

Based on these observations, a just-in-time slow feature analysis method is proposed for fault detection in traction systems. The following are the major contributions of this paper

1) The proposed method does not rely on precise mathematical models and a priori knowledge. It can be applied to dynamic non-linear systems.
2) The online update of the local prediction model solves the difficulty of detecting faults caused by sudden changes in parameters during mode switching.
3) The extraction of slowness features eliminates the influence of noise and unknown interference on fault detection. The accuracy and sensitivity of fault detection are improved.
4) This paper addresses the single-mode constraint of traditional multivariate analysis methods by fusing an instantaneous learning model with slow feature analysis. A new solution is provided for the extended application of traditional multivariate analysis.

The remaining part is arranged as follows: Section 2 summarizes the internal structure, operating mechanism, and characteristics of traction systems. Section 3 describes the framework, principles, and specific advantages of the proposed method. Section 4 verifies the method through a platform of traction systems. Finally, Section 5 summarizes this paper and puts forward the further work plan.

2. PRELIMINARIES

In this section, the composition of traction systems is briefly described, in which the measurement variables, incipient faults, noise, and disturbances are expressed in a formula form. To address the challenges of real-time incipient fault detection, the objectives of this study will be presented.

2.1. Traction systems of high-speed trains

As the primary component of high-speed trains, the mechanism of traction systems is to realize the conversion between electric energy and mechanical energy [21]. The schematic diagram of traction systems is shown in Fig. 1. Besides, traction systems mainly consist of high-voltage equipment and power units. High-voltage equipment consists of a pantograph, a vacuum circuit breaker, a current transformer, etc. Power units mainly include a traction transformer, a traction inverter, a traction motor, etc. [22].

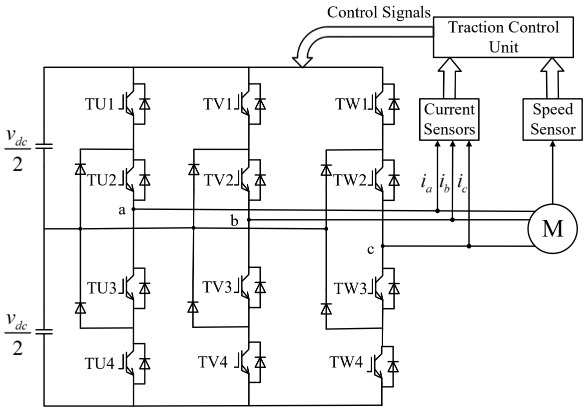


Fig. 1. Schematic diagram of traction systems in high-speed trains.

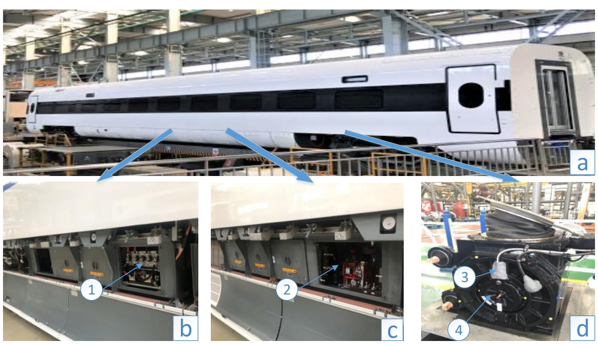


Fig. 2. Main components of traction systems in high-speed trains.

Fig. 2 shows the structural layout of the traction system in the vehicle body, where Fig. 2(a) is the vehicle body, Fig. 2(b-1) is the traction converter, and Fig. 2(c-2) is the low-voltage component that controls the traction converter; Fig. 2(d) is the traction motor, Fig. 2(d-3) is a sensor in the motor, and Fig. 2(d-4) is the cover of sensors.

2.2. Description of incipient faults in traction systems

A complete and accurate model is of great significance for timely detection of incipient faults. Nevertheless, the modeling complexity will increase due to the sophisticated structure of traction systems, which will increase the computational cost and error potential. Therefore, a signal-based description of traction systems and incipient faults is preferred in this paper, which avoids the influence of prior knowledge. Based on the measured signals of the sensors distributed in traction systems, the data matrix E consisting of the main state variables can be written as

$$E = [U_{cd1} \quad U_{cd2} \quad I_a \quad I_b \quad I_c \quad T_e], \tag{1}$$

where $E \in R^{N\times m}$ with N samples collected, m represents the number of variables tested. In this application, $m = 6$. Among them, U_{cd1} and U_{cd2} are the voltages at the upper and bottom support capacitors, respectively; I_a, I_b, and I_c are the three-phase currents in the inverter, and T_e is the electromagnetic torque of the traction motor. Traction systems can be regarded as multiple input multiple output systems due to the internal coupling relationship. Thus, the faults and noise in systems' operation only affect the output variables, rather than the input variables [23,24]. The inputs and outputs are divided as follows:

$$\begin{aligned} &X = [U_{cd1} \quad U_{cd2}], \\ &Y = [I_a \quad I_b \quad I_c \quad T_e]. \end{aligned} \tag{2}$$

There are many types of faults in traction systems, such as offset faults, drift faults, gain faults, etc. To effectively detect all types of faults in traction systems, the system output Y is described as

$$Y = Y^* + F + n, \tag{3}$$

where $Y \in R^{N\times m}$ is the practical measured value of the output, $Y^* \in R^{N\times m}$ is the output value under ideal conditions, $F \in R^{N\times m}$ is the fault in traction systems, and $n \in R^{N\times m}$ is the noise and other unknown disturbances. The noise and unknown disturbances are not only caused by external environmental conditions, but also the mechanical wear, such as friction and impact in the motor [25]. As shown in (4), incipient faults will gradually evolve into common faults.

$$\begin{aligned} &F : F_{inc} \to F_{com}, \\ &\text{s.t.} \quad \|F_{inc}./Y^*\| \leq 0.1. \end{aligned} \tag{4}$$

Consequently, the measured values of the system corresponding to incipient faults and common faults are Y_{inc} and Y_{com}, respectively.

$$\begin{aligned} &Y_{inc} = Y^* + F_{inc} + n, \\ &Y_{com} = Y^* + F_{com} + n, \end{aligned} \tag{5}$$

where F_{inc} and F_{com} are incipient faults and common faults, respectively. In this application, the faults whose fault amplitudes do not exceed 10% of the normal signal are defined as incipient faults. Thus, Y_{inc} and Y_{com} are the system outputs in the different stages of fault evolution.

2.3. Objective and design issues

As mentioned above, the nonlinear dynamic characteristics of traction systems are caused by the actual operation mode, which brings limitations to the application of traditional multivariate analysis methods [26,27]. In addition, as shown in (5), both incipient and common faults may be masked by noises, making them difficult to distinguish [28].

This paper focuses on incipient fault detection within traction systems under actual operations. Unlike the previous fault detection methods, the proposed method considers traction systems' dynamic behaviours, such as mode switching. Moreover, the impact of noise and disturbances on incipient fault detection under actual operations is considered in this paper.

Consequently, this paper intends to develop a real-time detection scheme for incipient faults of traction systems. And the objectives are listed as follows:

1) Build a model with high sensitivity to incipient faults without resort to a complex mathematical model and prior knowledge.
2) Find an online modeling method suitable for dynamic nonlinear systems.
3) Extract incipient fault features from the environment subject to the noise and unknown disturbances for achieving accurate incipient fault detection.

3. METHODOLOGY

A just-in-time slow feature analysis method is proposed to tackle the challenges described above. The proposed method merges the advantages of just-in-time learning with slow feature analysis. It is able to satisfy multi-mode processes in dynamic nonlinear systems by online updating of local dynamic prediction models. Meanwhile, the proposed method identifies incipient faults and disturbances by extracting slow features. Then the framework of the proposed scheme is introduced in more detail.

3.1. Just-in-time learning scheme

The just-in-time learning model belongs to a local learning method based on data, which differs from the traditional method of global modeling [29]. It is difficult for unified models to realize real-time online updates when the systems' operation mode changes [30]. For example, the neural network-based modeling methods will increase the computational complexity when dealing with large data sets because of their complex structure [2]. However, just-in-time learning differs from neural network-based approaches in that there is no standard learning phase. In addition, most of the local models transform complex global problems into simple local issues with the aid of prior knowledge or complex training strategies [31]. In contrast, the local prediction model for just-in-time learning is built by selecting data from the database that is strongly relevant to the query data. In addition to the advantages of online updates, just-in-time learning also has inherent adaptive characteristics by continuously adding fault-free data to the database. The procedure of the just-in-time learning model to predict the output data are given as follows:

1) The database data with high similarity to the query data are selected according to the similarity principle.
2) Based on the query data and the most similar data set, a local prediction model is developed.
3) The output prediction corresponding to the input query data is obtained from the local prediction model. In addition, the model is updated online when new query data arrives.

A suitable similarity criterion is essential to the just-in-time learning model. The similarity criterion needs to be set with a balance between speed and accuracy of selection. The construction of predictive models will be slowed down if the similarity criterion is too complex. Based on this consideration, the similarity is determined by the weighted combination of distance and angle in this paper. For query data x_q, the first l similar data are selected from the database $\{x_i, y_i\}_{i=1,2,\dots,N}$ according to the similarity criterion to form the similar sample set $\{x_i, y_i\}_{i=1,2,\dots,l}$, and the corresponding prediction output $\hat{y}_q$ is obtained. The details of the similarity criterion are given in Subsection 3.3. Furthermore, it is worth noting that

$$k_{\min} \leq l \leq k_{\max}, \tag{6}$$

where l represents the number of similar data selected. The determination in the size of a similar sample set is an essential factor in prediction accuracy. Two parameter values, $k_{\min}$ and $k_{\max}$, need to define beforehand, which are the upper and lower limits of the sample size. The setting of parameters $k_{\min}$ and $k_{\max}$ not only ensures the prediction accuracy but also avoids the expensive calculation.

3.2. Slow feature analysis

Traditional multivariate statistical analysis methods focus on global steady-state variability. Therefore, they cannot determine whether the cause of the alarm is a working state switch or faults. Slow feature analysis has better interpretation capability in terms of temporal correlation. It extracts the main variables based on the temporal variability of the data. Given a multi-dimensional vector input signal, slow feature analysis aims to find slowly changing features from the dynamics point [32,33]. For time-varying signals, slow-changing variables mean that they are the main variables reflecting the changes of system state. Thus, slow feature analysis can realize the distinction between real faults and disturbances. Mathematically, the slow feature analysis can be expressed as follows:

For a set of input signal $u(t) = [u_1(t), u_2(t), \dots, u_m(t)]^T$ with the dimension m, slow feature analysis requires finding a feature function, as shown below.

$$\{l_j(t) = h_j[u(t)]\}_{j=1}^{m}, \tag{7}$$

where $h_j(\cdot)$ is the feature function, and $l_j(t)$ is the required slowness feature.

Slow feature analysis obtains the feature function under the conditions of zero mean, unit variance, and decorrelation

$$\left.\begin{aligned} &\langle l_j\rangle_t = 0,\\ &\langle l_j^2\rangle_t = 1,\\ &\langle l_i l_j\rangle_t = 0,\ (\forall i \neq j)\end{aligned}\right\} \Rightarrow \min_{h_j(\cdot)} \langle \dot{l}_j^2\rangle_t, \tag{8}$$

where $\dot{l}_j$ is the first derivative of the variable along with time, representing the rate of change; $\langle\cdot\rangle_t$ is the mathematical mean of time. These constraints simplify the problem while retaining generality. They remove correlations between variables and ensure that slow features carry different information.

In general, the slowness features can be expressed by vectors and matrices, respectively.

$$\begin{aligned} &l_j(t) = h_j(u(t)) = w_j^T u,\\ &l = Wu, \end{aligned} \tag{9}$$

where w_j^T and W are the mapping coefficient vector and the coefficient matrix, respectively.

Remark 1: It is necessary to arrange the obtained slowness characteristics $\{l_j(t)\}_{j=1}^m$ in ascending order to find the main factors that affect the system state.

3.3. Just-in-time SFA

The data matrix E collected by the traction system in (1) and its further division in (2) are used as the basis for subsequent calculation. The similarity between the test input data x_q and database input data x_i can be performed based on

$$s_i = \alpha\sqrt{e^{-d^2(x_q,x_i)}} + (1-\alpha)\cos(\theta_i), \tag{10}$$

where $\alpha \in [0,1]$ is the weight parameter obtained by cross-validation, θ_i means the angle between query data vector Δx_q and database data vector Δx_i. Therefore, $\cos(\theta_i)$ can be obtained by

$$\cos(\theta_i) = \frac{\Delta x_q^T \Delta x_i}{\|\Delta x_q\|_2 \cdot \|\Delta x_i\|_2}, \tag{11}$$

where $\Delta x_q = x_q - x_{q-1}$ and $\Delta x_i = x_i - x_{i-1}$. The value of s_i ranges from 0 and 1. And the value of s_i reflects the degree of similarity between x_q and x_i.

Remark 2: It should be noted that $\cos(\theta_i)$ should satisfy $\cos(\theta_i) \geq 0$ to ensure that the angle between vectors does not exceed 90°. Otherwise, the database data is not related to the query data and needs to be removed from the database. Moreover, in order to establish the database $\{x_i, y_i\}_{i=1,2,\ldots,N}$, a large amount of data under actual healthy operation is necessary.

To determine the similar dataset $\{x_i, y_i\}_{i=1,2,\ldots,l}$, the similarity vector is arranged in descending order as follows:

$$S = [s_1, s_2, \ldots, s_{N_{train}}]^T \rightarrow S_{sort}. \tag{12}$$

x_i and y_i corresponding to the first l similarity after sorting are selected to form a similar sample set for regression. And the optimal solution of l is calculated by the internal calculation of the prediction model.

To satisfy the demands of real-time detection and ensure prediction accuracy, the ARX model is used as a local prediction model. It is expressed as

$$\hat{y}(k) = z^T(k-1)\,\psi, \tag{13}$$

where $\hat{y}(k)$ denotes the model's prediction, k represents the sampling time, and ψ denotes the model parameter.

The regression vector $z(k-1)$ is defined as

$$\begin{aligned} z(k-1) =& [y(k-1), \ldots, y(k-n_y), u(k-n_d-1),\\ &\ldots, u(k-n_d-n_u)]^T, \end{aligned} \tag{14}$$

where n_y and n_u represent the order of the model, and n_d is the time delay.

The traditional linear prediction model is generally described by

$$\hat{y}_q = x_q^T\left(p^T p\right)^{-1} p^T v, \tag{15}$$

$$\begin{aligned} &\mathrm{p} = W\Phi,\\ &v = WY, \end{aligned} \tag{16}$$

where $W \in \mathrm{R}^{N\times N}$ is a diagonal matrix composed of weight parameters, and $\Phi \in \mathrm{R}^{N\times n}$ is a matrix composed of training samples x_i corresponding to the weight parameters.

By mapping the traditional linear model to the ARX model, the parameters of ARX can be represented as

$$\Psi = \left(p^T p\right)^{-1} p^T v. \tag{17}$$

The similar sample set $\{x_i, y_i\}_{i=1,2,\ldots,l}$ is used to train the ARX model, so the model parameter (16) can be represented as follows:

$$\begin{aligned} &\mathrm{p}_l = W_l \Phi_l,\\ &v_l = W_l Y_l, \end{aligned} \tag{18}$$

$$W_l = \begin{pmatrix} w_1 & \cdots & 0\\ \vdots & \ddots & \vdots\\ 0 & \cdots & w_l \end{pmatrix},$$

$$\begin{aligned} &\Phi_l = [x_{index_1}, x_{index_2}, \ldots, x_{index_l}]^T,\\ &Y_l = [y_{index_1}, y_{index_2}, \ldots, y_{index_l}]^T, \end{aligned} \tag{19}$$

where $index_i$ represents the index value of the element in S_{sort} that is mapped to the original vector S.

To determine the optimal number of correlation samples l as l_{opt}, the error function is introduced as follows:

$$erf_l = \sum_{j=1}^{l}\left(s_j \frac{y_j - \Phi_j^T\left(p_l^T p_l\right)^{-1} p_l^T v_l}{1 - p_j^T\left(p_j^T p_l\right)^{-1} p_j}\right)^2 \Bigg/ \sum\nolimits_{j=1}^{l} s_j^2. \tag{20}$$

Therefore, the optimal solution is obtained when the error function reaches the minimum.

$$l_{opt} = \arg\min_{l} (erf_l). \tag{21}$$

In this way, the optimal parameters and predicted output of the prediction model can be obtained as follows:

$$\Psi_{l_{opt}} = \left(p_{l_{opt}}^T p_{l_{opt}}\right)^{-1} p_{l_{opt}}^T v_{l_{opt}}, \tag{22}$$

$$\hat{y}_q = x_q^T \Psi_{l_{opt}}. \tag{23}$$

For different query data, the local model should be reconstructed to ensure the prediction accuracy.

The residuals between the actual output y_{test}^T and the predicted output $\hat{y}_q$ are the basis of the slow feature analysis. The residuals are calculated as

$$\mathrm{Re} = \hat{y}_q - y_{test}^T, \tag{24}$$

$$U = [u_1,\ u_2,\ \ldots,\ u_m]^T = \mathrm{Re}. \tag{25}$$

The residuals obtained via the just-in-time learning local model are static and linear, which can be suitable for general fault detection methods. Hence, the just-in-time learning can be integrated with slow feature analysis.

Performing a SVD on the covariance matrix of U could be expressed as

$$\left\langle U^T U \right\rangle_t = K \Lambda K^T, \tag{26}$$

where K is the feature matrix, the whitening process is as follows:

$$z = \Lambda^{-\frac{1}{2}} K^T U = QU, \tag{27}$$

in which Q is the whitening matrix.

$$Q = \Lambda^{-\frac{1}{2}} K^T. \tag{28}$$

The ultimate purpose of slow feature analysis is to find out the long-standing and slowly changing features. Thus, the change rate of the decorrelation matrix is calculated as

$$\dot{z}(t) = \frac{z(t) - z(t - \Delta t)}{\Delta t}\ (\Delta t = 1). \tag{29}$$

In this paper, Δt is set to 1. Then, SVD on covariance matrix of $\dot{z}$ could be carried out

$$\left\langle \dot{z}\dot{z}^T \right\rangle_t = P^T \Sigma P. \tag{30}$$

The matrix P is a feature matrix, which not only reflects the decorrelation between variables, but also reflects the slowness of features. The slow feature matrix of the original data is obtained by inversely mapping the feature matrix P to the original data space, i.e.,

$$W = K \Lambda^{-\frac{1}{2}} P^T, \tag{31}$$

$$l = WU. \tag{32}$$

After sorting the slow features according to (12), the approach in [33] can be used to separate the slow features into primary and residual slow features.

$$l \to l_{sort} = l_d + l_e, \tag{33}$$

$$l_d = [l_1,\ l_2,\ \ldots,\ l_M]^T,$$
$$l_e = [l_{M+1},\ l_{M+2},\ \ldots,\ l_{M+M_e}]^T, \tag{34}$$

where M and M_e are the number of primary features and residual features, respectively.

As shown in (35), the T^2 test statistic is selected for incipient fault detection, which mainly reflects the deviation degree between features and the model.

$$T_d^2 = {l_d}^T \Lambda^{-1} l_d = {l_d}^T l_d,$$
$$T_e^2 = {l_e}^T \Lambda^{-1} l_e = {l_e}^T l_e. \tag{35}$$

Remark 3: The eigenvalue matrix obtained from the singular value decomposition of the covariance matrix is composed of variances. As shown in (8), one of the constraints of the slow feature analysis is the unit variance, so the eigenvalue matrix Λ obtained from (26) is a unit diagonal matrix I. Thus, the T^2 test statistic can be simplified to (35).

The test statistic T_d^2 obeys the χ^2 distribution with the degree of freedom M, and the statistic T_e^2 obeys the χ^2 distribution with the degree of freedom M_e.

$$T_d^2 = {l_d}^T l_d \sim \chi_M^2,$$
$$T_e^2 = {l_e}^T l_e \sim \chi_{Me}^2. \tag{36}$$

By setting a α, the control limit can be calculated by

$$J_{T_d^2} = \chi_{M,\alpha}^2,$$
$$J_{T_e^2} = \chi_{Me,\alpha}^2. \tag{37}$$

Therefore, the criteria for incipient fault detection can be summarized as follows:

1) If $T_d^2 \le J_{T_d^2}$ and $T_e^2 \le J_{T_e^2}$, the traction system is in a healthy state and there is no fault.
2) If $T_d^2 > J_{T_d^2}$ or $T_e^2 > J_{T_e^2}$, the traction system is working abnormally and faults have occurred.

Furthermore, this section summarizes the implementation process of the scheme, as shown in Algorithm 1 and Fig. 3.

4. EXPERIMENTAL RESULTS AND ILLUSTRATION

A traction-system platform established by Central South University is selected for verification, which allows the injection of faults with different types and amplitudes [34]. Then, a set of experiments demonstrate the superiority of the proposed scheme.

Algorithm 1: Implementation procedures.

Off-line training

1: Collect fault-free data in all states to construct a database $\{x_i, y_i\}_{i=1,2,\dots,N_{train}}$.
2: Obtain a training set under the same conditions as the data to be tested.
3: Calculate the similarity between the training set and the database according to (10).
4: Calculate the solution of the error function (20) and obtain the parameters (21).
5: Predict $\hat{y}_q$ from (23) and obtain the residual signal from (24).
6: Perform an SVD on the residual signal from (25) and then whiten it according to (26).
7: Calculate the coefficient matrix W according to the change rate $\dot{z}$.
8: Obtain the slow feature l from (33).
9: Calculate the threshold $J_{T_d^2}$ and $J_{T_e^2}$ according to (37).

On-line fault detection

1: Sample the online data.
2: Obtain model parameters refer to steps 3, 4 in Off-line training.
3: Obtain the residual signal U based on $\hat{y}_q$.
4: Compute the slow feature l refer to steps 6, 7, and 8 in Off-line training.
5: Calculate the test statistics ${T_d}^2$ and ${T_e}^2$ via (33) to achieve fault detection.

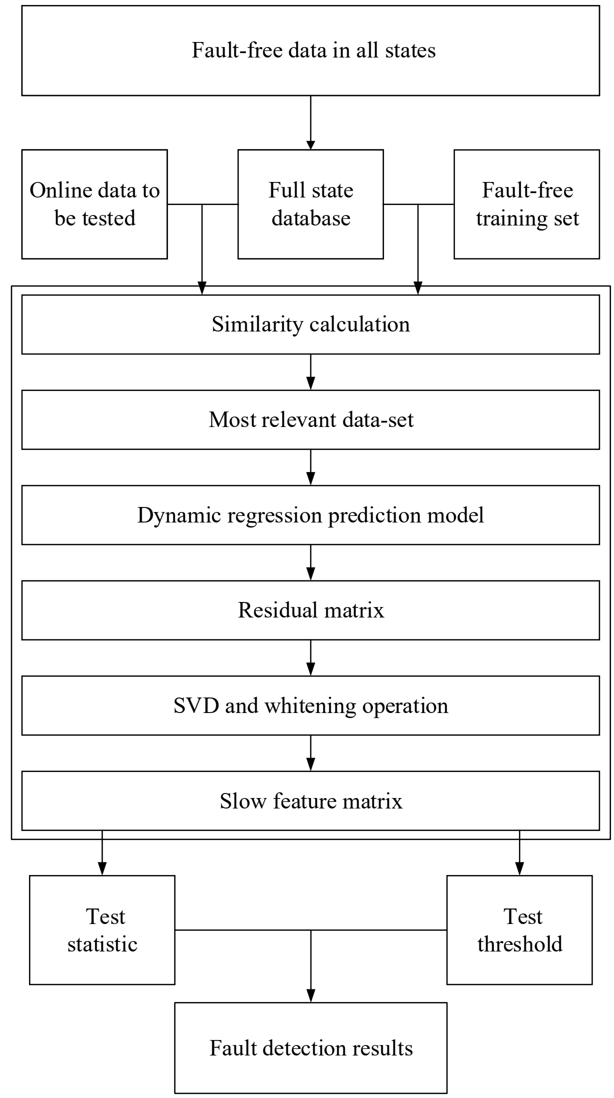


Fig. 3. The proposed fault detection scheme.

4.1. Experimental verification

The basic structure of the traction system in Fig. 4 mainly consists of a transformer, a rectifier, a DC link, a inverter, traction motors, and a traction control unit. The platform adjusts parameters for different components to enable the injection of faults and noise. Injectable fault components of the platform include sensors, traction motors, traction converters, and the traction control unit (TCU).

1) Multi-mode operation: The operating state of high-speed trains is divided into three modes: low-speed start, acceleration, and high-speed smooth running. To realize the operation of the three modes, the speed parameters of the platform are set to 50 km/h, 150 km/h, and and 250 km/h, respectively. The data are measured under multi-mode in Fig. 5, in which each mode contains 200000 samples and a sampling interval is 1 us. The traction speeds of the three modes in Fig. 4 are 50 km/h, 150 km/h, and 250 km/h, respectively.

2) Fault injection: As mentioned above, the voltage outputs (U_{cd1}, U_{cd2}) of the DC link are chosen as inputs, the three-phase current outputs (I_a, I_b, I_c) by the inverter,

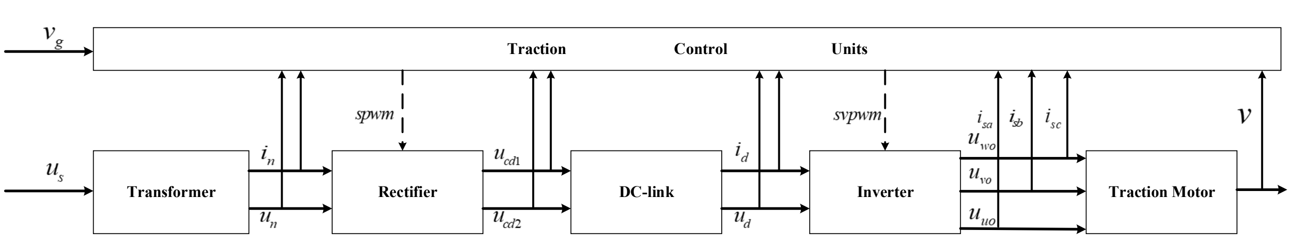


Fig. 4. A diagram of traction systems in high-speed trains.

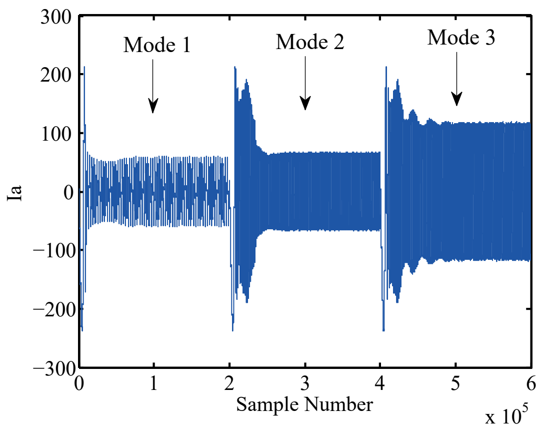


Fig. 5. The current I_a in multimode.

Table 1. Injection fault setting.

Fault no.	Fault type	Fault amplitude
1	Offset sensor fault	Constant value:1
2	Gain sensor fault	Constant value:1
3	Drift sensor fault	Slope,Initial output: [10, 3]
4	Motor bearing fault	Mechanical severity: 0.1
5	TCU fault	Invert Bit: 1, Register-Ts:1e-5

and the electromagnetic torque T_e of the motor are chosen as outputs. Two factors should be considered in determining the number of training and test samples: experimental reliability and computational efficiency. As a result, 200 data are sampled in each mode, and each output corresponds to 600 data for fault detection under multi-mode.

Five types of faults were selected for sensors, motors, and TCU injection. The 100th sampling time in each mode is taken as the injection time of faults and noise. The parameter settings related to types and amplitudes of injected faults as shown in Table 1. It's worth noting that the parameters related to the fault amplitude are adjusted to be less than 20% of the platform default value. Thus, the fault amplitude meets (4).

The setting 20% is the proportion relative to the default fault parameter of the traction system platform. The adjusted fault amplitude is less than 10% of the output signal.

3) Detection of incipient faults: The comparison between the actual outputs (I_a, I_b, I_c, T_e) of fault-free data and their predicted outputs is shown in Figs. 6-9, respectively. The red dotted line represents the actual outputs, and the green asterisk line represents the predicted outputs. The four figures show the prediction accuracy using the local model.

The detection results of the five incipient faults are given in Figs. 10(a)-10(e), where the solid blue line represents the test statistics and the dotted red line represents

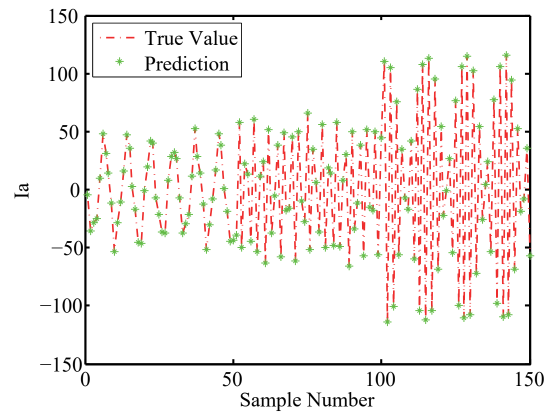


Fig. 6. Prediction results of I_a by JITL.

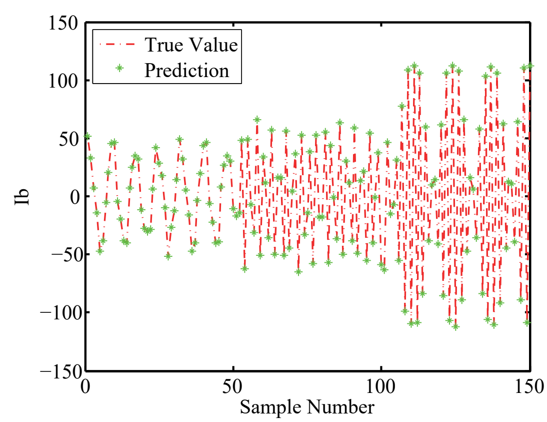


Fig. 7. Prediction results of I_b by JITL.

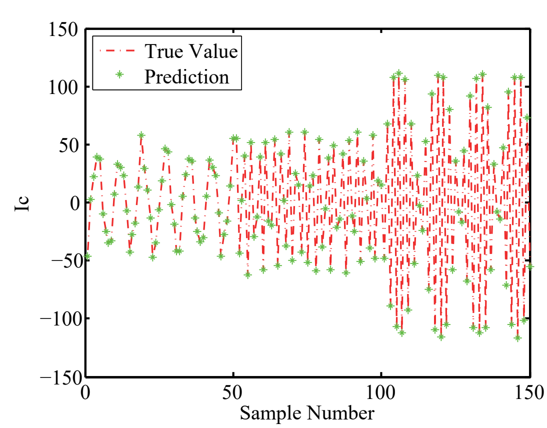


Fig. 8. Prediction results of I_c by JITL.

the threshold.

Fault 1: The signal change of the offset sensor fault is a fixed offset adding the output signal. Fig. 10(a) shows the detection results in the principal space and the residual space. As can be seen, fault 1 has been effectively detected.

Fault 2: The input signal will increase proportionally if

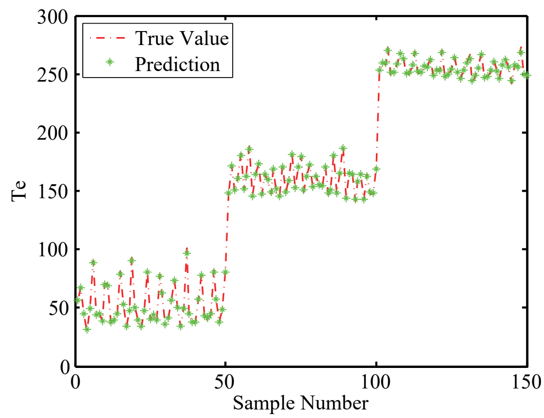
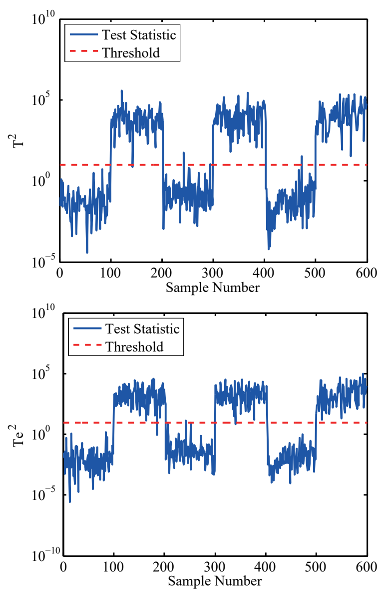


Fig. 9. Prediction results of T_e by JITL.

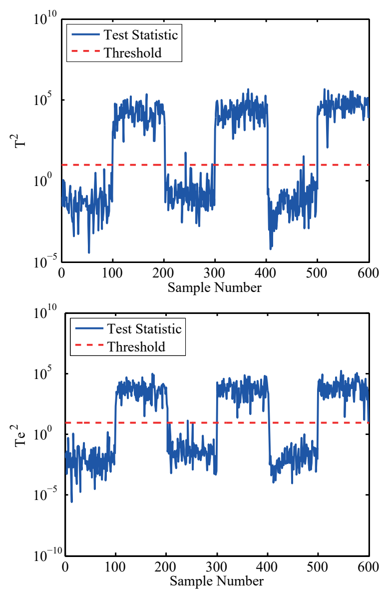
the gain sensor fault happens. As shown in Fig. 10(b), fault 2 can be accurately detected except for the false alarms at two sampling points in the principal space.

Fault 3: The change in electrical characteristics of the drift sensor fault is to add a signal to the original signal, and its value is proportional to the time of the output signal. As seen from Fig. 10(c), the alarm for fault 3 is accurate and obvious.

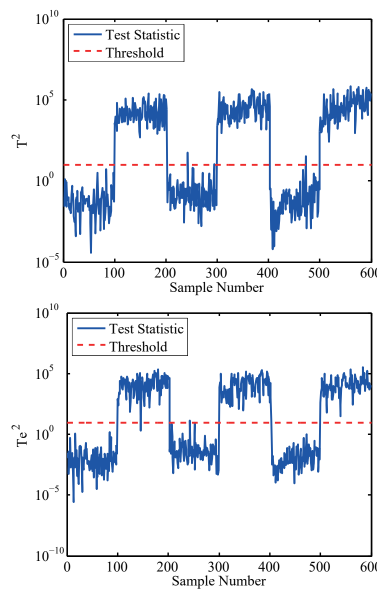
Fault 4: When the motor fault happens, the three-phase current has a specific frequency component in the frequency domain and fluctuates slightly in the time domain. Fig. 10(d) shows the detection results. It can be seen that when fault 4 happens, the result curves immediately rise and become higher than the thresholds.



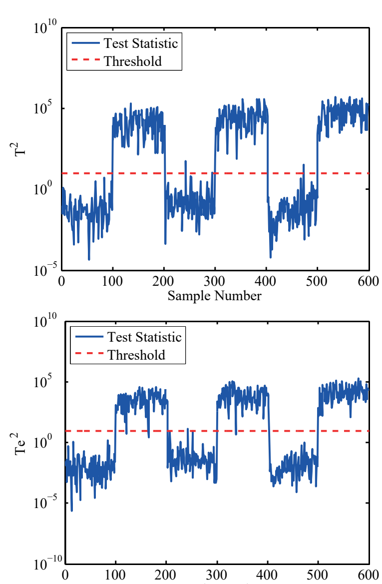
(a) Detection results for fault 1.



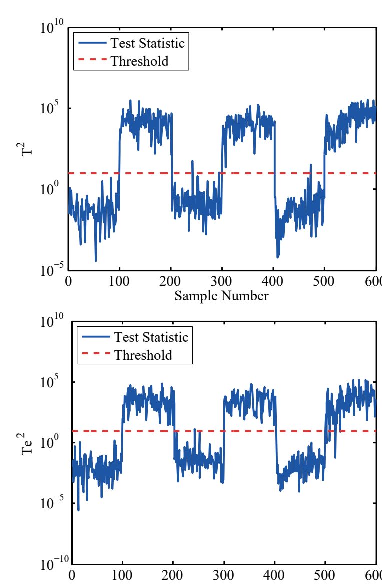
(b) Detection results for fault 2.



(c) Detection results for fault 3.



(d) Detection results for fault 4.



(e) Detection results for fault 5.

Fig. 10. Detection results by the proposed method.

Fault 5: Although TCU packaging makes the specific signal change of the TCU fault uncertain, it will also make the output signal error. As seen from Fig. 10(e), for fault 5, the scheme proposed in this paper can achieve correct and timely fault alarm.

Observed from the figures above, the proposed method can accomplish the detection tasks for incipient faults. These results are satisfactory.

4.2. Discussion

To illustrate the superiority of the scheme, the deep principal component analysis (DPCA), the deep slow feature analysis (DSFA), and the kernel canonical correlation analysis (KCCA) are tested on the traction-system platform. DPCA and DSFA are two recently developed methods for detecting incipient faults, which extract incipient fault information via the depth division of space. Furthermore, their effectiveness was proved in [35] and [16], respectively. Kernel function-based methods have been considered as an effective way to monitor traction systems with multi-mode characteristics. Therefore, KCCA in [36] is selected for performance comparison. Besides the intuitive detection results, this section also adds quantitative indicators for performance comparison.

Figs. 11-13 present the detection results of DPCA, DSFA, and KCCA for fault 2, respectively. As seen from Figs. 11 and 12, false alarms and missing alarms are obvious and hackneyed. It can be concluded that DPCA and DSFA are invalid for incipient fault detection of traction systems in actual operation. As shown in Fig. 13, the KCCA has better fault detectability than DPCA and DSFA, but still difficult to detect incipient faults in the traction system. In addition, JITL-PCA is selected for performance comparison to highlight the contribution of slow feature analysis to fault detectability improvement.

Based on these observations, the missing alarm rate (MAR) and the false alarm rate (FAR) are chosen for performance comparison [37]. Therefore, as shown in Table 2, the method proposed in this paper has salient advan-

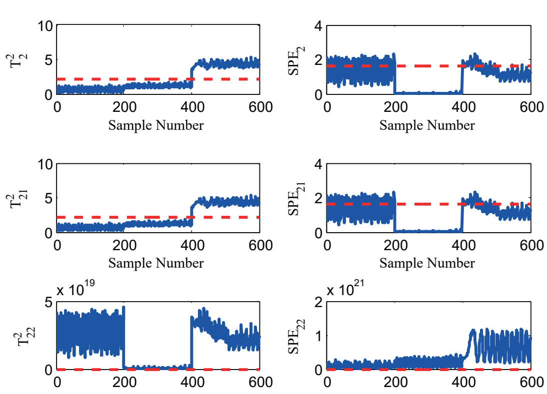


Fig. 11. Detection result for fault 2 by using DPCA.

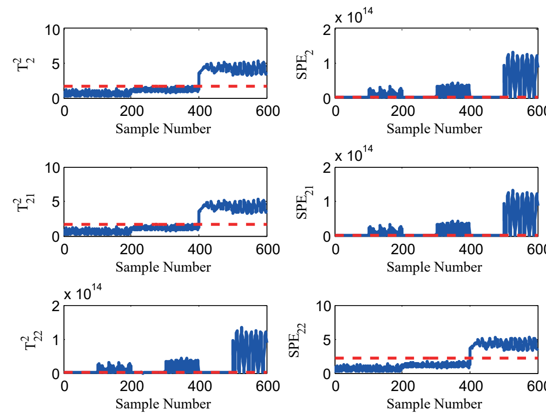


Fig. 12. Detection result for fault 2 by using DSFA.

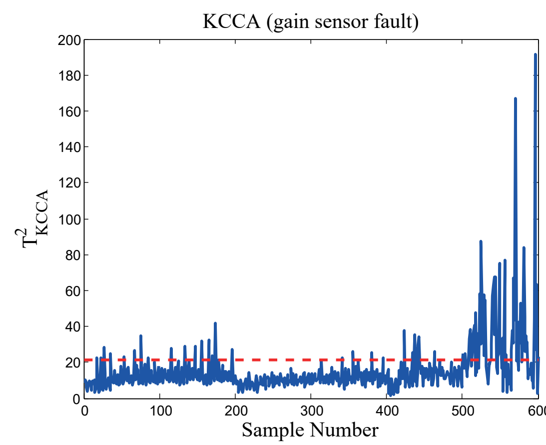


Fig. 13. Detection result for fault 2 by using KCCA.

Table 2. Performance comparisons for five faults.

Method	Fault 1		Fault 2		Fault 3		Fault 4		Fault 5	
	FAR	MAR	FAR	MAR	FAR	MAR	FAR	MAR	FAR	MAR
KCCA	5.50%	77.67%	5.15%	93.67%	5.15%	88.67%	5.15%	75.0%	5.15%	93.67%
DPCA	31.27%	66.67%	37.80%	80.33%	37.11%	77.33%	31.27%	71.33%	37.46%	78.67%
DSFA	31.62%	66.33%	93.13%	65.67%	32.30%	66.00%	31.62%	0.33%	95.88%	0.67%
JITL-PCA	48.33%	28.33%	42.00%	14.67%	49.33%	3.67%	42.00%	12.67%	41.67%	20.00%
JITL-SFA	0.69%	0.33%	0.34%	0%	0.69%	0%	0.34%	0%	0.69%	0%

tages both in FARs and MARs.

From the above detection results and quantitative comparison, it is concluded that the proposed scheme has superior fault detectability for various types of incipient faults in traction systems under actual operations.

5. CONCLUSIONS

This paper presented a just-in-time slow feature analysis method for detecting incipient faults in traction systems. The online modeling method based on just-in-time learning can address the nonlinear dynamics of traction systems. Through slow feature analysis, incipient faults are extracted from the environment containing noise and unknown disturbances. With the assistance of just-in-time learning, the internal constraints of multivariate statistical analysis methods are solved, which provides a new idea for the practical application of multivariate statistical analysis methods. The proposed method is finally tested via a platform of traction systems and shows better fault detectability than conventional methods.

Apart from this study, there are some open room for further studies. For example, the further work will be, on the basis of this effort, the improvement of computational efficiency while ensuring the detection accuracy.

CONFLICTS OF INTERESTS

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

REFERENCES

[1] H. Chen, B. Jiang, S. X. Ding, and B. Huang, "Data-driven fault diagnosis for traction systems in high-speed trains: A survey, challenges, and perspectives," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 3, pp. 1700-1716, March 2022.

[2] H. Chen, Z. Chai, B. Jiang, and B. Huang, "Data-driven fault detection for dynamic systems with performance degradation: A unified transfer learning framework," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, Article Sequence Number 3504712, pp. 1-12, 2021.

[3] D. Ronanki, S. A. Singh, and S. S. Williamson, "Comprehensive topological overview of rolling stock architectures and recent trends in electric railway traction systems," *IEEE Transactions on Transportation Electrification*, vol. 3, no. 3, pp. 724-738, September 2017.

[4] C. Cheng, X. Qiao, B. Zhang, H. Luo, Y. Zhou, and H. Chen, "Multiblock dynamic slow feature analysis-based system monitoring for electrical drives of high-speed trains," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, Article Sequence Number 3514310, pp. 1-10, 2021.

[5] S. Liu, B. Jiang, Z. Mao, and S. X. Ding, "Adaptive backstepping based fault-tolerant control for high-speed trains with actuator faults," *International Journal of Control, Automation, and Systems* , vol. 17, pp. 1408-1420, May 2019.

[6] Q. Shen, B. Jiang, and P. Shi, "Active fault-tolerant control against actuator fault and performance analysis of the effect of time delay due to fault diagnosis," *International Journal of Control, Automation, and Systems*, vol. 15, pp. 537-546, February 2017.

[7] Z. Hou, R. Chi, and H. Gao, "An overview of dynamic-linearization-based data-driven Control and applications," *IEEE Transactions on Industrial Electronics*, vol. 64, no. 5, pp. 4076-4090, May 2017.

[8] Z. Chen and K. E. Haynes, *Chinese Railways in the Ea of High-speed*, Emerald Group Publishing Limited, Bingley, UK, 2015.

[9] H. Chen and B. Jiang, "A review of fault detection and diagnosis for the traction system in high-speed trains," *IEEE Transactions on Intelligent Transportation Systems*, vol. 21, no. 2, pp. 450-465, February 2020.

[10] Z. Liu, H. Wang, J. Liu, Y. Qin, and D. Peng, "Multitask learning based on lightweight 1DCNN for fault diagnosis of wheelset bearings," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, Article Sequence Number 3501711, pp. 1-11, 2021.

[11] C. Yang, Z. Wu, T. Peng, H. Zhu, and C. Yang, "A fractional steepest ascent morlet wavelet transform-based transient fault diagnosis method for traction drive control system," *IEEE Transactions on Transportation Electrification*, vol. 7, no. 1, pp. 147-160, March 2021.

[12] 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.

[13] K. Zhang, B. Jiang, G. Tao, and F. Chen, "MIMO evolution model-based coupled fault estimation and adaptive control with high-speed train applications," *IEEE Transactions on Control Systems Technology*, vol. 26, no. 5, pp. 1552-1566, September 2018.

[14] M. Hamadache and D. Lee, "Principal component analysis based signal-to-noise ratio improvement for inchoate faulty signals: Application to ball bearing fault detection," *International Journal of Control, Automation, and Systems*, vol. 15, pp. 506-517, January 2017.

[15] J. Yang, F. Zhu, X. Wang, and X. Bu, "Robust sliding-mode observer-based sensor fault estimation, actuator fault detection and isolation for uncertain nonlinear systems," *International Journal of Control, Automation, and Systems*, vol. 13, pp. 1037-1046, May 2015.

[16] C. Cheng, X. Qiao, H. Luo, G. Wang, W. Teng, and B. Zhang, "Data-driven incipient fault detection and diagnosis for the running gear in high-speed trains," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 9, pp. 9566-9576, September 2020.

[17] X. Deng, X. Tian, S. Chen, and C. J. Harris, "Nonlinear process fault diagnosis based on serial principal component analysis," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 29, no. 3, pp. 560-572, March 2018.

[18] Z. Chen, S. X. Ding, T. Peng, C. Yang, and W. Gui, "Fault detection for non-gaussian processes using generalized canonical correlation analysis and randomized algorithms," *IEEE Transactions on Industrial Electronics*, vol. 65, no. 2, pp. 1559-1567, February 2018.

[19] H. Luo, H. Zhao, and S. Yin, "Data-driven design of fog-computing-aided process monitoring system for large-scale industrial processes," *IEEE Transactions on Industrial Informatics*, vol. 14, no. 10, pp. 4631-4641, October 2018.

[20] S. Sun, H. Zhang, Y. Wang and J. Zhang, "Dissipativity-based intermittent fault detection and fault-tolerant control for uncertain switched random nonlinear systems with multiple delays," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 52, no. 12, pp. 7457-7468, 2022.

[21] Y. Song, Z. Liu, A. Ronnquist, P. Novik, and Z. Liu, "Contact wire irregularity stochastics and effect on high-speed railway pantograph-catenary interactions," *IEEE Transactions on Instrumentation and Measurement*, vol. 69, no. 10, pp. 8196-8206, October 2020.

[22] S. Zhang, *Fundamental Application Theory and Engineering Technology for Railway High-speed Trains*, Science Press, Beijing, China, 2007.

[23] D. Meng, Y. Jia, J. Du, and F. Yu, "Robust learning controller design for MIMO stochastic discrete-time systems: An H_∞-based approach," *International Journal of Adaptive Control and Signal Processing*, vol. 25, pp. 653-670, 2011.

[24] Y. Lu, X. Peng, D. Yang, M. Yang, and W. Zhong, "Model-agnostic meta-learning with optimal alternative scaling value and its application to industrial soft sensing," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 12, pp. 8003-8013, December 2021.

[25] D. Meng and K. L. Moore, "Robust iterative learning control for nonrepetitive uncertain systems," *IEEE Transactions on Automatic Control*, vol. 62, no. 2, pp. 907-913, February 2017.

[26] H. Ji, X. He, J. Shang, and D. Zhou, "Incipient fault detection with smoothing techniques in statistical process monitoring," *Control Engineering Practice*, vol. 62, pp. 11-21, May 2017.

[27] A. Romanenko, A. Muetze, and J. Ahola, "Incipient bearing damage monitoring of 940-h variable speed drive system operation," *IEEE Transactions on Energy Conversion*, vol. 32, no. 1, pp. 99-110, March 2017.

[28] S. Sun, H. Zhang, C. Liu, and Y. Liu, "Dissipativity-based intermittent fault detection and tolerant control for multiple delayed uncertain switched fuzzy stochastic systems with unmeasurable premise variables," *IEEE Transactions on Cybernetics*, vol. 52, no. 9, pp. 8766-8780, 2022.

[29] S. Yin, H. Gao, J. Qiu, and O. Kaynak, "Fault detection for nonlinear process with deterministic disturbances: A just-in-time learning based data driven method," *IEEE Transactions on Cybernetics*, vol. 47, no. 11, pp. 3649-3657, 2017.

[30] X. Yuan, Z. Ge, B. Huang, Z. Song, and Y. Wang, "Semisupervised JITL framework for nonlinear industrial soft sensing based on locally semisupervised weighted PCR," *IEEE Transactions on Industrial Informatics*, vol. 13, no. 2, pp. 532-541, April 2017.

[31] Q. Jiang, X. Yan, H. Yi, and F. Gao, "Data-driven batch-end quality modeling and monitoring based on optimized sparse partial least squares," *IEEE Transactions on Industrial Electronics*, vol. 67, no. 5, pp. 4098-4107, May 2020.

[32] C. Shang, F. Yang, X. Gao, X. Huang, J. Suykens, and D. Huang, "Concurrent monitoring of operating condition deviations and process dynamics anomalies with slow feature analysis," *AIChE Journal*, vol. 61, pp. 3666-3682, May 2015.

[33] C. Shang, F. Yang, B. Huang, and D. Huang, "Recursive slow feature analysis for adaptive monitoring of industrial processes," *IEEE Transactions on Industrial Electronics*, vol. 65, no. 11, pp. 8895-8905, November 2018.

[34] C. Yang, C. Yang, T. Peng, X. Yang, and W. Gui, "A fault-injection strategy for traction drive control systems," *IEEE Transactions on Industrial Electronics*, vol. 64, no. 7, pp. 5719-5727, July 2017.

[35] H. Chen, B. Jiang, N. Lu, and Z. Mao, "Deep PCA based real-time incipient fault detection and diagnosis methodology for electrical drive in high-speed trains," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 6, pp. 4819-4830, June 2018.

[36] R. T. Samuel and Y. Cao, "Kernel canonical variate analysis for nonlinear dynamic process monitoring," *IFAC-Papers OnLine*, vol. 48, no. 8, pp. 605-610, 2015

[37] H. Chen, L. Li, C. Shang, and B. Huang, "Fault detection for nonlinear dynamic systems with consideration of modeling errors: A data-driven approach," *IEEE Transactions on Cybernetics*, vol. 53, no. 7, pp. 4259-4269, 2023.

Chao Cheng received his Ph.D. degree from Jilin University, Changchun, China, in 2014. He is currently an associated with the Changchun University of Technology, Changchun. He has been a Post-Doctoral Fellow in process control engineering with the Department of Automation, Tsinghua University, Beijing, China, since 2018. He has also been a Post-Doctoral Fellow with the National Engineering Laboratory, CRRC Changchun Railway Vehicles Co., Ltd., China, since 2018. His research interests include dynamic system fault diagnosis and predictive maintenance, wireless sensor networks, artificial intelligence, and data-driven methods.

Xiuyuan Sun received his B.Eng. degree from the Changchun University of Technology, Changchun, China, in 2019. He is currently working toward an M.Eng. degree in computer science and engineer with the Changchun University of Technology. His research interests include complex system fault diagnosis and data-driven fault detection.

Junjie Shao received his B.E. degree in computer science and technology from Changchun Normal University China, in 2011. He is currently the director of the Data Research Office for the CRRC ChangChun Railway Vehicles CO., LTD. He is serious and responsible and he has strong expansion capabilities. He took the lead in building the company's data research platform and continued to carry out basic technology research and engineering promotion for application. The research results in its field have reached an industry-leading level. As a leader, he constructed the company's PHM system, and through continuous optimization of the PHM system, he provided strong support for the company's market bidding, product design, artificial intelligence for IT operations, and vehicle repairs.

Hongtian Chen received his B.S. and M.S. degrees from the School of Electrical and Automation Engineering from Nanjing Normal University, China, in 2012 and 2015, respectively; and he received a Ph.D. degree in College of Automation Engineering from Nanjing University of Aeronautics and Astronautics, China, in 2019. He had ever been a Visiting Scholar at the Institute for Automatic Control and Complex Systems, University of Duisburg-Essen, Germany, in 2018. Now he is a Post-Doctoral Fellow with the Department of Chemical and Materials Engineering, University of Alberta, Canada. His research interests include process monitoring and fault diagnosis, data mining and analytics, machine learning, and quantum computation; and their applications in high-speed trains, new energy systems, and industrial processes. Dr. Chen was a recipient of the Grand Prize of Innovation Award of Ministry of Industry and Information Technology of the People's Republic of China in 2019, the Excellent Ph.D. Thesis Award of Jiangsu Province in 2020, and the Excellent Doctoral Dissertation Award from Chinese Association of Automation (CAA) in 2020.

Chao Shang received his B.Eng. degree in automation and a Ph.D. degree in control science and engineering from Tsinghua University, Beijing, China, in 2011 and 2016, respectively. After working as a Postdoctoral Fellow at Cornell University, he joined the Department of Automation, Tsinghua University in 2018 as an Assistant Professor. His research interests include data-driven modeling, monitoring, diagnosis and optimization with applications to industrial manufacturing processes. Dr. Shang is the recipient of Springer Excellent Doctorate Theses Award, Best Paper Award of 1st International Conference on Industrial Artificial Intelligence, Zijing Scholarship, and Teaching Achievement Award from Tsinghua University.