Incipient Fault Diagnosis for High-Speed Train Traction Systems via Stacked Generalization

Zehui Mao, *Member, IEEE*, Mingxuan Xia, Bin Jiang, *Fellow, IEEE*, Dezhi Xu, *Senior Member, IEEE*, and Peng Shi, *Fellow, IEEE*

***Abstract*—Diagnosing the fault as early as possible is significant to guarantee the safety and reliability of the high-speed train. Incipient fault always makes the monitored signals deviate from their normal values, which may lead to serious consequences gradually. Due to the obscure early stage symptoms, incipient faults are difficult to detect. This article develops a stacked generalization (stacking)-based incipient fault diagnosis scheme for the traction system of high-speed trains. To extract the fault feature from the faulty data signals, which are similar to the normal ones, the extreme gradient boosting (XGBoost), random forest (RF), extra trees (ET), and light gradient boosting machine (LightGBM) are chosen as the base estimators in the first layer of the stacking. Then, the logistic regression (LR) is taken as the meta estimator in the second layer to integrate the results from the base estimators for fault classification. Thanks to the generalization ability of stacking, the incipient fault diagnosis performance of the proposed stacking-based method is better than that of the single model (XGBoost, RF, ET, and LightGBM), although they can be used to detect the incipient faults, separately. Moreover, to find out the optimal hyperparameters of the base estimators, a swarm intelligent optimization algorithm, pigeon-inspired optimization (PIO), is employed. The proposed method is tested on a semiphysical platform of the CRH2 traction system in CRRC Zhuzhou Locomotive Company Ltd. The results show that the fault diagnosis rate of the proposed scheme is over 96%.**

***Index Terms*—Fault diagnosis, high-speed train, incipient faults, pigeon-inspired optimization (PIO), stacked generalization.**

I. Introduction

DUE TO high speed, loading capacities, and the ability to stay on schedule, high-speed trains have presently become one of the most important forms of transportation. To guarantee the safety and reliability of these railway transportation systems, the intelligent control strategies with the advanced automatic train control (ATC) system have motivated considerable research during the past few decades (see [1]–[5]). However, owing to the harsh operating environment and with the running time increasing, various types of faults may inevitably occur in the high-speed train, which may reduce the operational efficiency and even threaten the safety of trains and passengers. The fault detection, diagnosis, and fault-tolerant control technologies for high-speed trains play a pivotal role in the modern and intelligent transportation system (see [6]–[9]).

Regarded as the heart of the high-speed train, the traction system provides traction power for the entire system. The faults in the traction system may appear in different locations, such as motor, converters, control units, sensors, and so on [10]. These faults can affect the effectiveness of traction force, and result in reduction of speed or may even stop the train. The existing fault diagnosis methods for high-speed train traction systems can be classified into two major categories: 1) mathematical model-based and 2) data-based methods. Model-based fault diagnosis methods are continuously emerging in recent years, such as adaptive techniques [11]–[13] and sliding-mode observer [14]–[16]. It should be noted that the model-based methods acquire the priori knowledge and fault models [17], [18]. Unlike model-based methods, data-based fault diagnosis approaches are independent of the mathematical modeling, and can be realized by analyzing the large amounts of processed data. Due to the ease of implementation, some results for traction systems using data-based methods can be obtained [19]–[21]. Considering that the traction system is a closed-loop system with multiple coupling and complex fault propagation, it is difficult to identify the faults with higher accuracy, which motivates us to study the data-driven-based fault diagnosis methods for high-speed train traction systems.

On the other hand, the electrical components (i.e., capacitance, resistance, etc.) in the traction system always degenerate with time and could accelerate degradation under the harsh environment, which is considered as the incipient faults in the fault diagnosis field. Incipient faults refer to those whose amplitudes are small and symptoms are obscure especially in early stage [22]. Compared with the relatively obvious features of a serious fault, the weak features of incipient fault are easily hidden in the time-varying process [23]. Furthermore, the intrinsic fault tolerance ability of controllers also increases the difficulty of diagnosing the incipient faults [24]. Until now,

Manuscript received 2 July 2020; revised 10 October 2020; accepted 25 October 2020. Date of publication 10 December 2020; date of current version 19 July 2022. This work was supported in part by the National Natural Science Foundation of China under Grant 61922042, Grant 61773131, and Grant 61973140; in part by the Qing Lan Project; in part by the 111 Project under Grant B20007; and in part by the Fundamental Research Funds for the Central Universities under Grant NE2019101. This article was recommended by Associate Editor G. P. Liu. *(Corresponding author: Bin Jiang.)*

Zehui Mao, Mingxuan Xia, and Bin Jiang are with the College of Automation Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China (e-mail: zehuimao@nuaa.edu.cn; binjiang@nuaa.edu.cn).

Dezhi Xu is with the School of Internet of Things Engineering, Jiangnan University, Wuxi 214122, China (e-mail: xudezhi@jiangnan.edu.cn).

Peng Shi is with the School of Electrical and Electronic Engineering, University of Adelaide, Adelaide, SA 5005, Australia (e-mail: peng.shi@adelaide.edu.au).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TCYB.2020.3034929.

Digital Object Identifier 10.1109/TCYB.2020.3034929

there has been some research on the incipient fault diagnosis for high-speed trains. In [25], a moving window is introduced into the kernel principal component analysis (PCA) model, and new fault detection indices with their control limits are proposed to achieve fault detection, isolation, and estimation. In [26], three tools, including: 1) SVM; 2) convolutional; and 3) recurrent networks, are compared to diagnosis incipient faults of a dc motor. In [27], a deep PCA-based method is proposed to diagnose incipient faults of sensors. However, the PCA-based methods require data to obey the Gaussian distribution, while the data of the traction system cannot satisfy the Gaussian distribution. Moreover, some methods can detect the incipient fault but not diagnose the fault type, and perhaps cannot achieve high precision. The incipient fault diagnosis problem is not fully studied and needs more attention.

Stacked generalization (Stacking), as a kind of ensemble learning method [28], can combine the best hypotheses of different date-driven/machine-learning models, which might be an ideal technique to archive the incipient fault diagnosis and improve the diagnosis performance. It has been proved that the classification accuracy of the stacking model is always better than that of the best individual classifier in the entire stacking model [29], [30]. Although many fields, such as imaging processing, hyperspectral data classification, and steel structures optimization, have benefitted from the stacking method [31], few applications on fault detection and diagnosis have been reported.

Considering the difficulty of diagnosing incipient fault by single machine-learning algorithms, we will propose a stacking-based incipient fault diagnosis scheme for the traction system of high-speed trains. The main contributions of this work include the following.

1) A stacking-based incipient fault diagnosis scheme is proposed, in which four different machine-learning models, that is, extreme gradient boosting (XGBoost), random forest (RF), extra trees (ET), and light gradient boosting machine (LightGBM), are combined as base estimators in the first layer of the stacking model to deeply mine the fault features, and generated fault features are classified by the meta estimator in the second level to reduce variance and improve generalization accuracy.
2) In order to simplify the hyperparameters tuning process, the pigeon-inspired optimization (PIO) algorithm is applied to search the optimal hyperparameters for the base estimators. Furthermore, the proposed incipient fault diagnosis scheme is verified by the data from a semiphysical platform of the CRH2 traction system in CRRC Zhuzhou Locomotive Company Ltd. (see [32] and [33]).

The remainder of this article is organized as follows. Section II introduces the traction system, presents the description of incipient faults, and illustrates the fault types. Section III describes the structure and principle of stacked generalization and introduces the PIO algorithm. Section IV proposes the incipient fault diagnosis scheme via stacking and PIO algorithm, and summarizes the fault diagnosis process. Section V presents the experimental results, and conclusions of this work are remarked in Section VI.

II. Problem Formulation

This section briefly introduces the traction system of high-speed train and shows the incipient faults in this system. Then, the incipient fault diagnosis problem is formulated with the design objective.

A. Incipient Fault in Traction System

The traction system provides the power for high-speed train, which includes the pantograph, transformer, inverter, rectifier, intermediate dc circuit, three-phase asynchronous ac motor, and so on. The traction system works under the strong current and high voltage, such as, single-phase ac with 1500 V and 50 Hz, intermediate dc between 2600 and 3000 V, three-phase ac with adjustable voltage 0–2300 V, and frequency 0–220 Hz. The rigorous operating condition leads to the degradation fault of the electrical components, that is, the incipient fault. The incipient fault has a small amplitude and has characteristics of early changing and slow developing, which can deteriorate with time into something more serious.

For the traction system of high-speed trains, the incipient faults easily occur in the current sensors, speed sensors, resistance, capacitance, insulated-gate bipolar transistor (IGBT), and induction motor. Generally, the incipient faults can be divided into two types: 1) the incipient abrupt fault and 2) incipient time-varying fault [22]. The difference between an incipient abrupt fault and a common abrupt fault lies in the gain percent of deviation. Compared with the actual value under normal conditions, the incipient fault signal is quite small, which ranges from 1% to 10% [23]. For example, when the current sensors have the incipient faults with fixed deviations, the measurements of sensors deviate from the actual values by slight bias. This bias could lead to the change of the system behavior, but could not result in instability [34]. The incipient fault of the current sensor can be formulated as

$$I_f = I_h(1+\alpha), \quad 0 < \alpha \leq 0.1 \tag{1}$$

where I_f is the current value of the faulty sensor, I_h is the health current value, and α is a tiny bias of the current signal.

Besides, some components in the traction system, such as the resistances and capacitances in the dc link, are highly susceptible to performance degradation. This kind of fault has a small amplitude initially but may deteriorate with time into something more serious, which is considered an incipient time-varying fault. For example, the resistance degradation in the dc link is expressed as

$$R_f = R_h e^{-\eta t_f}, \quad 0 < \eta \leq 1 \tag{2}$$

where R_f is the faulty resistance value, R_h is the health resistance value, and η represents the attenuation index of resistance. The incipient time-varying fault usually manifests as measurements of sensors to change slowly. As a consequence, the single machine-learning model-based classifiers can hardly accomplish the incipient fault diagnosis object.

B. Design Objective

For the CRH2 high-speed train, the traction system is equipped with multiple sensors to monitor the operation state

TABLE I
SENSOR SIGNALS

Symbol	Description	Units
I_n	Transformer current output	A
U_{net}	Grid voltage	V
U_{d1}	Bridge voltage	V
U_{d2}	Bridge voltage	V
T_{oq}	Motor torque	N*m
I_a	Three-phase ac output current	A
I_b	Three-phase ac output current	A
I_c	Three-phase ac output current	A
W_r	Motor speed	rpm

of the train [35]. As shown in Table I, the monitored signals include the voltage of the grid U_{net}; output current of transformer I_n; two bridge voltages U_{d1} and U_{d2}; three-phase ac current I_a, I_b, and I_c; motor torque T_{oq}; and motor speed W_r.

In addition, the closed-loop control and coupling result in the complex propagation path of incipient faults, which makes it difficult to establish the faulty model of the entire traction system. Thanks to the available signals shown in Table I, the data-driven method can be used to design an incipient fault diagnosis scheme independent of the mathematical model, where the stacked generalization is chosen as the benchmark method. Thus, the design objective of this article can be summarized as a stacked generalization fault diagnosis that is to be designed for the incipient faults of the traction systems, which can deal with the low amplitude and obscure initial feature problems, and be verified through the data from a semiphysical platform of the CRH2 traction system in CRRC Zhuzhou Locomotive Company Ltd. (see [32] and [33]).

III. PRELIMINARIES

In this section, the structure of stacked generalization is introduced with the methods for choosing the base learners and metalearners. Moreover, to optimize the hyperparameters in the base estimators, the PIO algorithm is employed.

A. Stacked Generalization

Stacked generalization is an ensemble machine-learning method which contains two or more layers of different models to achieve more excellent performance of classification [28]. In the two-layer stacking structure, the models in the first layer, called the base estimators, use the same target function to obtain the best hypothesis. Then, the second layer called the meta estimator, coordinates the balance among the hypothesis generated from the first layer, and makes the final decision. When the stacking model has more than two layers, the estimators in middle layers use the hypothesis generated from the previous layer, and also generate the hypothesis for the next layer.

Different from traditional ensemble methods, such as Adaboost and RF, stacking uses a meta estimator to integrate the hypothesis from the base estimators, which is a nonlinear fusion. In a stacking, the base estimator and meta estimator can be any machine-learning models. In this article, a two-layer stacking is used to construct the fault diagnosis scheme, in which the first layer is to obtain the obscure feature and the second layer is to reduce the generalization error.

B. Estimators Selecting

It can directly be seen that the performance of the stacking determines by the selection of base estimators and meta estimators, which also affects the accuracy of the fault diagnosis. The basic idea of stacking is to minimize the generalization error rate by combining different estimators, which means the base estimators should generate a diverse set of hypothesis, and meta estimator should find out the best way to combine the outputs hypothesis of base estimators [36]. The higher the output diversity of base estimators, the stronger the generalization ability of the stacking model. Therefore, the base estimators should preferably be different from each other. From this point and considering the obscure fault features, various base estimators are needed to obtain the weak features. Here, four different and independent models for base estimators are chosen as XGBoost (XGB) [37], RF [38], ET [39], and LightGBM (LGB) [40], which construct the first layer of the fault diagnosis stacking model.

As each model (XGB, RF, ET, and LGB) can achieve the clarification, which means each model can diagnosis the faults but the accuracy could not be satisfied, the meta estimator must be designed to fuse the results of the base estimators to optimize fault diagnosis. Furthermore, considering the computational cost and overfitting, the logistic regression (LR) model is selected as the meta estimator.

From now on, the incipient fault diagnosis scheme is constructed by the stacked generalization, in which the XGB, RF, ET, and LGB are combined as the base estimators to make an initial decision, and the LR model is chosen as the meta estimator to improve the accuracy.

C. Hyperparameters Searching via Pigeon-Inspired Optimization

In the XGB, RF, ET, and LGB models, the hyperparameters, which are always artificially chosen, indeed affect the fault diagnosis performance. The hyperparameters in the four single models include the number of estimators, max depths, min samples leaf, etc., which are always chosen by experiments and would not achieve the desired performance in a short time. So, it motivates us to find an optimization algorithm to decide these parameters. As the PIO algorithm is a swarm intelligent optimization algorithm, which can quickly find the optimal solution globally and overcome the difficulty of adjusting hyperparameters manually, the PIO algorithm will be used to optimize the hyperparameters of the stacking to achieve the incipient fault diagnosis. The PIO is inspired by the homing process of pigeons, in which the pigeons use two different navigation mechanisms. Then, there are two operator models corresponding to these two navigation mechanisms [41].

1) *Map and Compass Operator:* Pigeons sense the Earth's magnetic field and form a homing map in their mind. They change directions according to the attitude of the

sun. When they are close to the destination, they depend less on the sun and magnetic particles.

2) *Landmark Operator:* Near the destination, pigeons rely on the nearby landmarks for navigation. In this operator model, pigeons familiar with the landmarks can move straightly to their destination, whereas, pigeons who are not familiar with the landmarks will follow the pigeons who are familiar with landmarks.

When using PIO to optimize the hyperparameters, we take the coordinate values of each pigeon in the D-dimensional space as the values of the D different hyperparameters in a model. The homing process can be regarded as the optimal solution searching process, and the last remaining pigeon who is closest to the destination represents the global optima. Considering n pigeons, the position and velocity of the ith pigeon can be described as

$$Z_i = [z_{i1}, z_{i2}, \ldots, z_{iD}] \tag{3}$$

$$V_i = [v_{i1}, v_{i2}, \ldots, v_{iD}] \tag{4}$$

where i is the index of the pigeon, and $i = 1, 2, \ldots, n$, D is the dimension of the search space of the hyperparameters optimization. According to the map and compass operator 1), the position Z_i and velocity V_i are updated during every iteration. The Z_i and V_i in the N_cth iteration are updated by

$$V_i^{N_c} = V_i^{N_c-1} e^{-R\times N_c} + \text{rand}\cdot\left(Z_{\text{gbest}} - Z_i^{N_c-1}\right) \tag{5}$$

$$Z_i^{N_c} = Z_i^{N_c-1} + V_i^{N_c} \tag{6}$$

$$F\left(Z_i^{N_c-1}\right) = \begin{cases} \dfrac{1}{\text{fitness}\left(Z_i^{N_c-1}\right)+\varepsilon}, & \text{fitness}_{\min}\left(Z_i^{N_c-1}\right) \\ \text{fitness}\left(Z_i^{N_c-1}\right), & \text{fitness}_{\max}\left(Z_i^{N_c-1}\right) \end{cases} \tag{7}$$

$$Z_{\text{gbest}} = \begin{cases} \text{argmin}\left[F\left(Z_i^{N_c-1}\right)\right], & \text{fitness}_{\min}\left(Z_i^{N_c-1}\right) \\ \text{argmax}\left[F\left(Z_i^{N_c-1}\right)\right], & \text{fitness}_{\max}\left(Z_i^{N_c-1}\right) \end{cases} \tag{8}$$

where $N_c = 1, 2, \ldots, N_{c_1}^{\max}$ is the current number of iteration, $N_{c_1}^{\max}$ is the maximum iteration in map and compass operator, $R \in [0, 1]$ is the map and compass factor, rand $\in [0, 1]$ is a random number, $F(Z_i^{N_c} - 1)$ is the fitness value of each pigeon in the $N_c - 1$ iteration, fitness$(\cdot)$ is the fitness function, fitness$_{\min}(\cdot)$ means the target is to minimize the objective function, and fitness$_{\max}(\cdot)$ means the target is to maximize the objective function. Z_{gbest} is the global optimal position obtained by comparing the fitness functions of all pigeons after the $N_c - 1$ iteration. When the maximum iteration number $N_{c_1}^{\max}$ is reached, the operation of the map and compass operator stops and the landmark operator works.

When the landmark operator 2) works, half of pigeons will be left in every iteration. Those who are far away from their destination will be abandoned, as they are not familiar with the landmarks. In the N_cth iteration, the position Z_i of the ith pigeon is updated by

$$n^{N_c} = \frac{n^{N_c-1}}{2} \tag{9}$$

$$Z_i = Z_i^{N_c-1} + \text{rand}\cdot\left(Z_{\text{center}}^{N_c-1} - Z_i^{N_c-1}\right) \tag{10}$$

where $Z_i^{N_c-1}$ and n^{N_c-1} are the position Z_i and the number of individuals at the $(N_c - 1)$th iteration, $N_c = N_{c_1}^{\max} + 1, \ldots, N_{c_2}^{\max}$, $N_{c_2}^{\max}$ is the maximum iteration in the landmark operator. The center position of the rest of the individuals in the $(N_c - 1)$th iteration will be considered as the landmark in the next iteration. The center position $Z_{\text{center}}^{N_c-1}$ is calculated by

$$Z_{\text{center}}^{N_c-1} = \frac{\sum_{i=1}^{N_c-1} Z_i^{N_c-1} F\left(Z_i^{N_c-1}\right)}{n^{N_c-1}\sum_{i=1}^{N_c-1} F\left(Z_i^{N_c-1}\right)}. \tag{11}$$

Likewise, when the iteration reaches the maximum iteration $N_{c_2}^{\max}$, the landmark operator stops operating. After that, all hyperparameters of base estimators are searched and the best models are obtained, which can be used to extract the features of incipient faults.

IV. Stacked Generalization-Based Incipient Fault Diagnosis

In this section, the incipient fault diagnosis scheme is proposed using the stacking and PIO algorithm, whose overall structure is depicted in Fig. 1. Using the training data X_1, the base estimator models are constructed by the PIO. Then, the constructed base estimator models are employed to generate the new features from the training data X_2, which are used to train the LR model of the meta estimator. The obtained stacking model can achieve the incipient fault diagnosis as the new data input. The details and process of each step in the proposed incipient fault diagnosis scheme are illustrated in the following sections.

A. Training Base Estimators With PIO

In order to obtain the optimal hypothesis of base estimators, each model (XGB, RF, ET, and LGB) is trained using the same data. The optimal hyperparameters are searched by PIO as the following steps. First, the original data are divided into training and testing datasets. After the preprocess, the training data $X \in R^{m\times n}$ are divided into two subsets

$$X_1 = \left[x_1, x_2, \ldots, x_{n_1}\right] \in R^{m\times n_1} \tag{12}$$

$$\text{and} \quad X_2 = \left[x_{n_1+1}, x_{n_1+2}, \ldots, x_n\right] \in R^{m\times(n-n_1)} \tag{13}$$

where m represents the number of variables, n represents the number of samples, and n_1 represents the number of samples in dataset X_1. The data X_1 are treated as the training data for each base estimator, while X_2 is used to generate the new features for the meta estimator.

Constructed by the classification and regression trees (CARTs), the base estimators (XGB, RF, ET, and LGB) have the similar structures, which lead to the hyperparameters of them being also similar. Generally, the hyperparameters of these models are max depth, n estimators, colsample bytree, and so on, which are summarized in Table II. h_d^ρ represents the hyperparameters, where $\rho \in \{\text{XGB, RF, ET, LGB}\}$ represents the different models of XGB, RF, ET, and LGB, and $d = 1, 2, \ldots, D$ with D being the dimension of the solution space of each model.

Here, we take the XGB model as an example to show how the PIO search the optimal hyperparameters. The objective

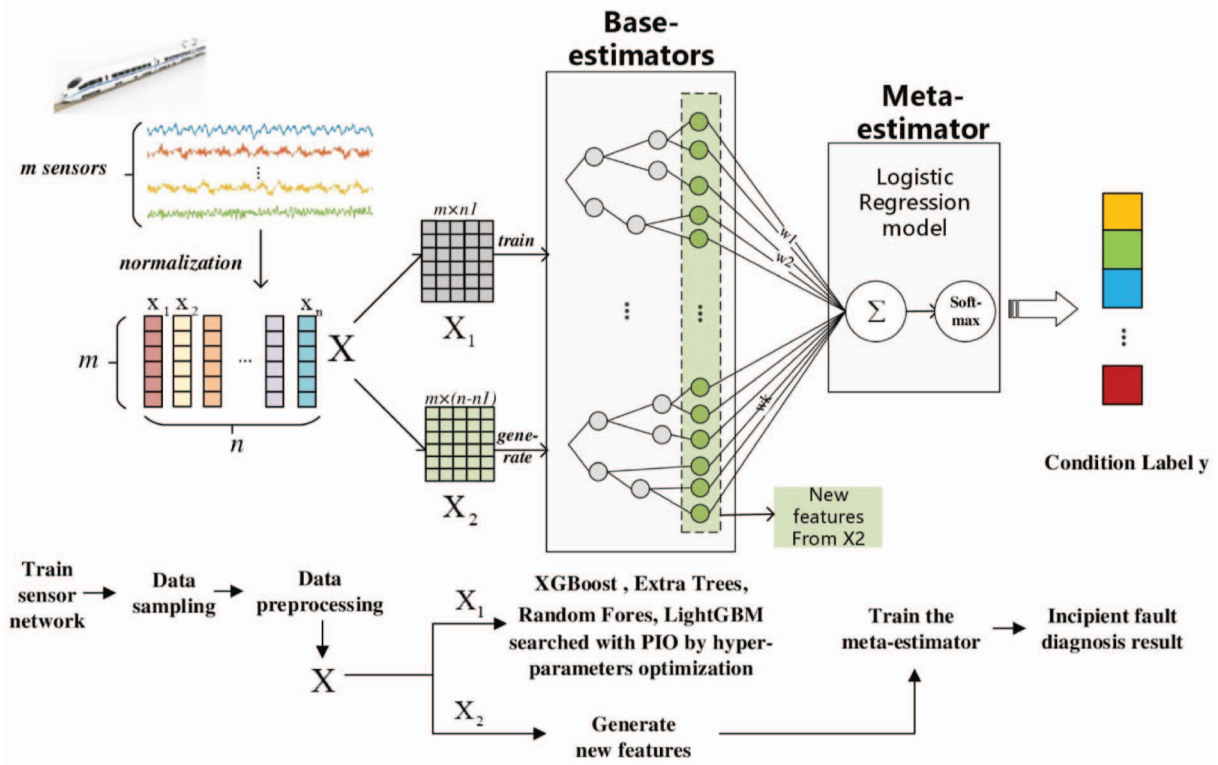

Fig. 1. Stacking-based fault diagnostics.

of the PIO is to obtain a set of hyperparameters $h_d^{\mathrm{XGB}} = [h_1^{\mathrm{XGB}}, h_2^{\mathrm{XGB}}, h_3^{\mathrm{XGB}}, h_4^{\mathrm{XGB}}, h_5^{\mathrm{XGB}}, h_6^{\mathrm{XGB}}]$ to maximize the fitness value F^{XGB}, which is the accuracy of the cross-value prediction of the XGB model. The training process of the XGB model with PIO can be achieved by the algorithm in Table III (Algorithm I).

B. Generation of New Features

As the base estimators are constructed, the training data X_2 are put into the obtained base estimators and are marked as

$$X_2 = \left[x_{n_1+1}, x_{n_1+2}, \ldots, x_n\right] \in R^{m\times(n-n_1)} \tag{14}$$

where n_1 is the number of samples in X_1, and m is the dimension of features.

For the ith sample $x_i \in X_2$, the output of the constructed XGB model is a series of values $[y_i^1, y_i^2, \ldots, y_i^{T_{\mathrm{XGB}}}]^T$

$$\begin{aligned} y_i^1 &= f_1(x_i) \\ y_i^2 &= f_2(x_i) \\ &\vdots \\ y_i^{T_{\mathrm{XGB}}} &= f_{T_{\mathrm{XGB}}}(x_i) \end{aligned} \tag{15}$$

where T_{XGB} represents the number of CARTs in the XGB model, and f_k is the function of the kth CART. It should be noted that the CART cannot output the probability values by itself. Due to the need of the probability values for the meta estimator, we use a sigmoid function $(1/[1+\exp(-y_i^j)])$, $j = 1, 2, \ldots, T_{\mathrm{XGB}}$ to normalize them into probability values, that is, values between [0, 1].

TABLE II
HYPERPARAMETERS OF THE STACKING MODEL

Hyper-parameters	Random Forest	Extra Trees	XGBoost	LightGBM
n_estimators	h_1^{RF}	h_1^{ET}	h_1^{XGB}	h_1^{LGB}
max_depths	h_2^{RF}	h_2^{ET}	h_2^{XGB}	h_2^{LGB}
learning_rate	–	–	h_3^{XGB}	h_3^{LGB}
subsample	–	–	h_4^{XGB}	h_4^{LGB}
colsample_bytree	–	–	h_5^{XGB}	h_5^{LGB}
min_samples_leaf	h_3^{RF}	h_3^{ET}	–	–
min_child_weight	–	–	h_6^{XGB}	–
num_leaves	–	–	–	h_6^{LGB}

Let $x_i^{\mathrm{XGB}} = [(1/[1+\exp(-y_i^1)]), (1/[1+\exp(-y_i^2)]), \ldots, (1/[1+\exp(-y_i^{T_{\mathrm{XGB}}})])]^T \in R^{T_{\mathrm{XGB}}}$. Then, the new features extracted from X_2 can be denoted as $X_{\mathrm{new}}^{\mathrm{XGB}} = [x_{n_1+1}^{\mathrm{XGB}}, x_{n_1+2}^{\mathrm{XGB}}, \ldots, x_n^{\mathrm{XGB}}] \in R^{T_{\mathrm{XGB}}\times(n-n_1)}$. Then, the features extracted from the base estimators can be denoted as

$$X_{\mathrm{new}} = \left[X_{\mathrm{new}}^{\mathrm{XGB}},\ X_{\mathrm{new}}^{\mathrm{RF}},\ X_{\mathrm{new}}^{\mathrm{ET}},\ X_{\mathrm{new}}^{\mathrm{LGB}}\right]^T \tag{16}$$

where $X_{\mathrm{new}}^{\mathrm{RF}} \in R^{T_{\mathrm{RF}}\times(n-n_1)}$, $X_{\mathrm{new}}^{\mathrm{ET}} \in R^{T_{\mathrm{ET}}\times(n-n_1)}$, and $X_{\mathrm{new}}^{\mathrm{LGB}} \in R^{T_{\mathrm{LGB}}\times(n-n_1)}$ are the features generated from RF, ET, and LGB, respectively. The feature X_{new} will be used to train the meta estimator for the fault recognition.

C. Fault Classification

The meta estimator works to classify the features X_{new} from the base estimators. Considering that the fault diagnosis is a multiclassification problem, we choose the extended form of LR as the meta estimator, in which the input vector should

TABLE III
ALGORITHM OF PIO SEARCHING

Algorithm 1 The algorithm of PIO searching

Initialization
Set the initial values of n, R, $N_{c1}^{\max}$ and $N_{c2}^{\max}$, with $N_{c2}^{\max} > N_{c1}^{\max}$
Initialize the velocity V_i and position Z_i of each pigeon randomly
Calculate fitness values $F(Z_i)$ of each pigeon by Eq. (7)
Initialize Z_{gbest} by Eq. (8)

Map and compass operations
Repeat
Update each pigeon's velocity $V_i^{N_c}$ and position $Z_i^{N_c}$ by Eqs. (5)-(6)
Evaluate fitness value $F(Z_i^{N_c})$, and update Z_{gbest} by Eqs. (7)-(8)
$N_c = N_c + 1$
Until
$N_c = N_{c1}^{\max}$

Landmark operations
Repeat
Permute the position $Z_i^{N_c-1}$ of each individual in order of $F(Z_i^{N_c-1})$
Retain half of the pigeons whose fitness values are higher
Update each pigeon's position $Z_i^{N_c}$ by Eq. (10)
Evaluate $F(Z_i^{N_c})$, and update Z_{gbest} by Eqs. (7)-(8)
$N_c = N_c + 1$
Until
$N_c = N_{c2}^{\max}$

Output Z_{gbest}

be a set of probability values. For the ith sample $x_i \in X_{\text{new}}$, the label values of y_i are $\{1, 2, \ldots, K\}$, where K is the total number of categories. Then, we denote the probability that y_i belongs to the jth category as

$$P(y_i j|x_i;\theta) = \frac{e^{\theta_l^T x_i}}{\sum_{l=1}^{K} e^{\theta_l^T x_i}} \tag{17}$$

where $j = 1, \ldots, K$ and $l = 1, \ldots, K$, $\theta = [\theta_1, \theta_2, \ldots, \theta_k]$ is the weight of features in the LR model, and $\theta_k \in R^{n-n_1}$. So the probability value for each category is a K-dimensional vector $h_\theta(x_i)$, which is calculated by a nonlinear softmax function

$$\begin{aligned} h_\theta(x_i) &= \begin{bmatrix} P(y_i = 1|x_i;\theta) \\ P(y_i = 2|x_i;\theta) \\ \vdots \\ P(y_i = K|x_i;\theta) \end{bmatrix} \\ &= \frac{1}{\sum_{l=1}^{K} e^{\theta_l^T x_i}} \begin{bmatrix} e^{\theta_1^T x_i} & e^{\theta_2^T x_i} \ldots & e^{\theta_K^T x_i} \end{bmatrix}^T. \end{aligned} \tag{18}$$

The term $1/\sum_{l=1}^{K} e^{\theta_l^T x_i}$ can normalize the probability distribution. Optimized by the gradient descent algorithm, the maximum index of $h_\theta(x_i)$ can suggest the fault diagnosis result

$$H_{\text{diagnostic}} = \arg\max_{l} \frac{e^{\theta_l^T x_i}}{\sum_{l=1}^{K} e^{\theta_l^T x_i}}. \tag{19}$$

Through the obtained base estimators and meta estimators, the stacking model can be built. By reducing both error and variance, the proposed stacking method can achieve high accuracy of the incipient fault diagnosis for the high-speed train traction systems.

D. Incipient Fault Diagnosis

From now on, the stacked generalization-based incipient fault diagnosis scheme for high-speed train traction systems is constructed. As shown in Fig. 1, we summarize the process as the following steps.

Step 1 (Data Preprocess): The raw data formed as $X \in R^{m\times n}$ are preprocessed. The zero averaging and unit variance methods are used to standardize the data

$$\mu = \frac{1}{m}\sum_{i=1}^{m} x_i, \quad \bar{x}_i = x_i - \mu, \quad x_i \in X \tag{20}$$

$$\delta^2 = \frac{1}{m}\sum_{i=1}^{m} \bar{x}_i^2, \quad \tilde{x}_i = \frac{\bar{x}_i}{\delta^2} \tag{21}$$

where $x_i \in R^m$ is the ith sample of data X, μ is the average value, $\bar{x}_i$ is the zero-averaged value, δ^2 is the variance of data X, and $\tilde{x}_i$ is a data sample normalized by variance. After that, the preprocessed data are refined as $X = [\tilde{x}_1, \tilde{x}_2, \ldots, \tilde{x}_n]$.

Step 2 (Train the Base Estimators by PIO): The training set X is divided into two subsets X_1 and X_2. The subset X_1 is used to establish the XGB, RF ET, and LGB models as the base estimators, during which the PIO algorithm (5)–(11) is employed to optimize their hyperparameters. The training process is shown in Section IV-A.

Step 3 (Generate New Features): Using the trained base estimators in step 2, the subsets X_2 are used to generate the features X_{new} through (15) and (16).

Step 4 (Train the Meta estimator): The new feature X_{new} is employed to train the LR model (18) of the meta estimator. The feature generation and classifier for the incipient faults are obtained, which can be used online or offline.

Step 5 (Incipient Fault Diagnosis): New data, which should be preprocessed as that in step 1, is as the input of the trained stacking model. Then, using the new data, the base estimators generate new features, and the output labels of meta estimator (19) represent the fault situation.

V. EXPERIMENTAL RESULTS AND ILLUSTRATIONS

To verify the effectiveness of the proposed fault diagnosis scheme, experiments on a semiphysical platform of the CRH2 traction system in CRRC Zhuzhou Locomotive Company Ltd. (see [32], [33]), are presented in this section.

A. Experiment Condition

The semiphysical platform shown in Fig. 2 includes a DCU controller, a signal conditioning unit, a dSPACE real-time simulator, and a PC. Six different faults are injected into the traction system, which are presented in Table IV. Among them, speed and current sensor bias faults are the incipient abrupt fault, while resistance degradation and capacitance degradation of the dc link are incipient time-varying faults. Motor rotor broken and IGBT open-circuit faults are abrupt faults. All of them are taken into consideration to show the fault diagnosis performance of the proposed method.

In Table IV, Γ is the step function, $\text{TH}_1(M, f_1, s) = A_1 \cdot \cos(2\pi f_{s1} t + \theta_1) + A_2 \cdot \cos(2\pi f_{s2} t + \theta_2)$ and $\text{TH}_2(cs)$ are the threshold functions, where M is the damaged condition, A_1 and A_2 are the amplitude of the sideband frequency components, f_1 is the fundamental frequency, $f_{s1} \approx f_{s2}$, s is the TM

TABLE IV
CONSIDERED SEVEN CONDITIONS

Category	Typical description	Parameter	Fault type
Health	Healthy data	None	Health
Fault I	$I_f = I_h(1+\alpha)$	α =0.005,0.01, 0.02, 0.05, 0.1	Incipient abrupt fault
Fault II	$S_f = S_h(1+\beta)$	β =0.005,0.01, 0.02, 0.05, 0.1	Incipient abrupt fault
Fault III	$X_f = X_h + \Gamma \cdot \mathrm{TH}_1(M, f_{s1}, s)$	M =0.5,f_{s1} = 128.7,s = 0.038	Abrupt fault
Fault IV	$X_f = X_h + \Gamma \cdot \mathrm{TH}_2(cs)$	$cs = -1$	Abrupt fault
Fault V	$R_f = R_h e^{-\gamma t_f}$	γ =0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1	Incipient time-varying fault
Fault VI	$C_f = C_h e^{-\delta t_f}$	$\delta = 0.01, 0.05, 0.1, 0.15, 0.2$ $0.25, 0.3, 0.35, 0.4, 0.5$	Incipient time-varying fault

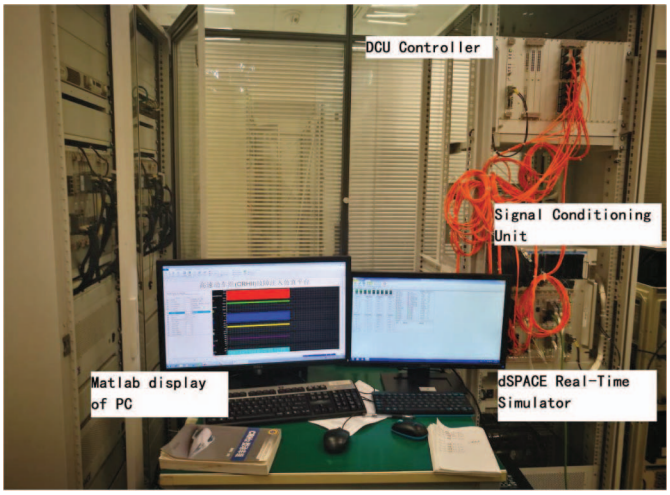

Fig. 2. High-speed train traction system experimental platform.

slip ratio, and θ_1 and θ_2 are the phase angles of the sideband frequency components and are an arbitrary value. In $\mathrm{TH}_2(cs)$, cs is the amplitude of the impulse sequences, and $cs = -1$ denotes an open-circuit fault. The injection of these faults have been taken into consideration in the experiment platform that is developed in the work of [32] and [33]. The sampling frequency of the sensors is 400 μs, and the sensors signals shown in Table I are collected as a dataset, which are collected under the steady-operation speed 110 km/h. Then, 70 000 samples of the acquired data are used to establish the stacking models and the remaining 7700 is treated as the test data.

B. PIO Settings

During the hyperparameter optimization process, the parameters of the PIO algorithm need to be initialized. The initial number of pigeons n depends on the dimensions D of the solution space of different estimators. Here, we choose $n = 5 \times D$. Then, the numbers of pigeons are 15, 15, 35, and 30 in RF, ET, XGBoost, and LightGBM, respectively. The compass factor R is chosen as 0.5, and maximum iterations $N_{c1}^{\max}$ and $N_{c2}^{\max}$ are set as 20 and 25. The search ranges of the hyperparameters in the base estimators are shown in Table V.

In Table V, the hyperparameters that have the same names in different models, such as n_estimators, learning_rate, and min_sample_leaf, have the same functionalities since the structures of XGBoost, RF, ET, and LightGBM models are similar.

TABLE V
SEARCH RANGES OF HYPERPARAMETERS OF BASE ESTIMATORS

Hyper-parameters	Random Forest	Extra Trees	XGBoost	LightGBM
n_estimators	[100, 700]	[100, 700]	[100, 700]	[100, 700]
max_depths	[5, 15]	[5, 15]	[5, 10]	[5, 10]
learning_rate	–	–	[0.005, 0.2]	[0.005, 0.2]
subsample	–	–	[0.5,1]	[0.5,1]
colsample_bytree	–	–	[0.8, 1]	[0.6, 1]
min_samples_leaf	[20, 70]	[20, 70]	–	–
min_child_weight	–	–	[1,6]	–
gamma	–	–	[0,0.5]	–
num_leaves	–	–	–	[30, 150]

TABLE VI
SOLVING RESULTS OF HYPERPARAMETERS

Hyper-parameters	Random Forest	Extra Trees	XGBoost	LightGBM
n_estimators	300	470	510	670
max_depths	12	11	8	7
learning_rate	–	–	0.17	0.16
subsample	–	–	1	0.8
colsample_bytree	–	–	0.8	0.9
min_samples_leaf	40	50	–	–
min_child_weight	–	–	4	4
gamma	–	–	0	–
num_leaves	–	–	–	150

These models are all built by CART. The optimal hyperparameters of XGB, RF, ET, and LGB models searched by the PIO method are given in Table VI, which construct the base estimators.

Then, using the trained base estimators, the new features can be generated. Using the new features and through the steps of the LR model training and fault recognition shown in Section IV-D, the incipient fault can be detected and the results are shown in the next section.

C. Results

To show the performance of the stacked generalization-based fault diagnosis scheme, four used models of base estimators, which are XGB, RF, ET, and LGB, are also tested on the traction system to compare with the stacking model. Each of the them is used for feature extraction and classification to achieve the fault diagnosis, separately. The hyperparameters of these comparison models are the same as those used in the stacking model, which are shown in Table VI. In this article, we use accuracy and $F1$-score to evaluate these

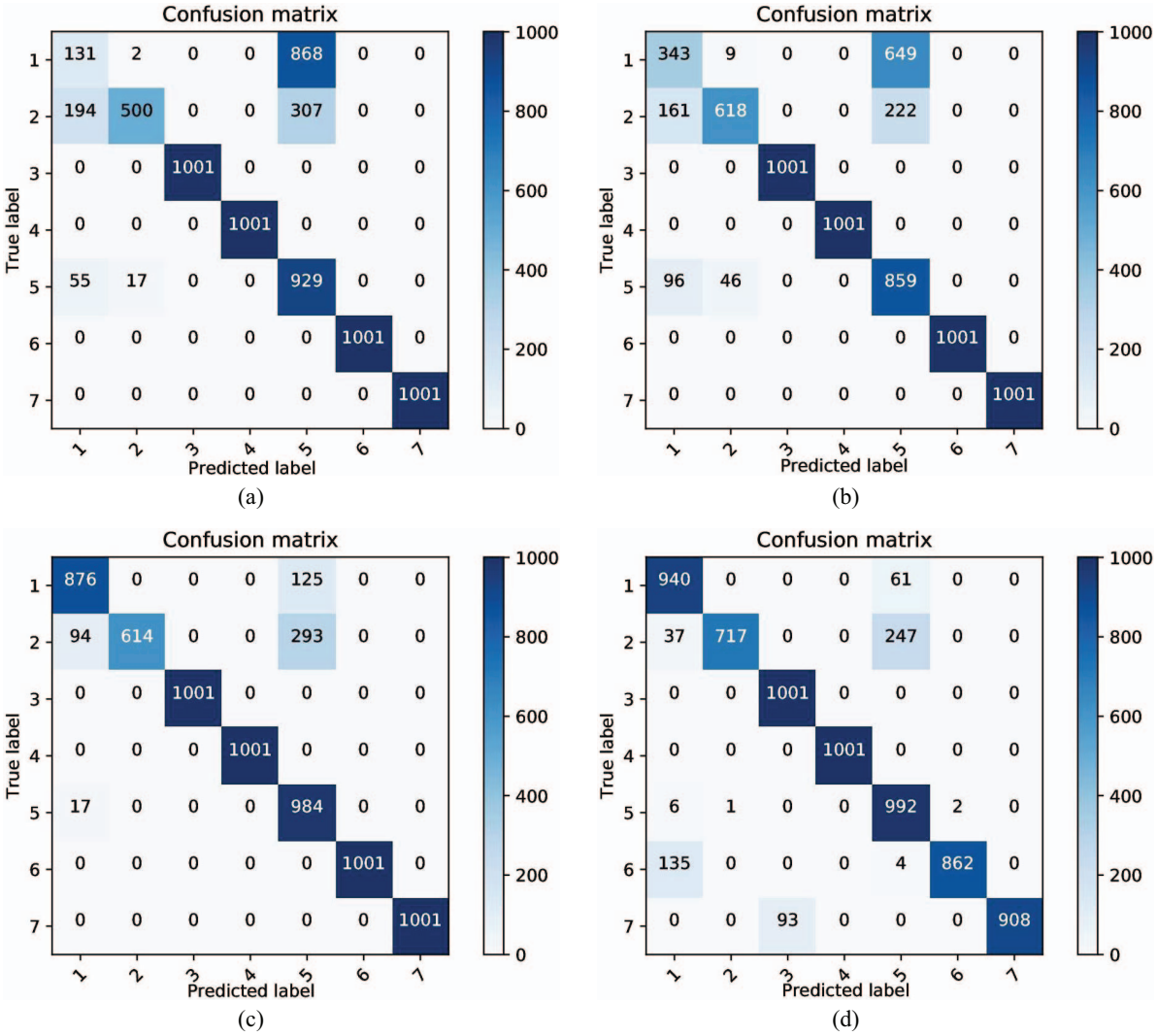

Fig. 3. Confusion matrix of single estimators. (a) *Random Forest*. (b) *Extra Trees*. (c) *XGBoost*. (d) *LightGBM*.

TABLE VII
ESTIMATORS PERFORMANCE COMPARISON

Method	Accuracy	F1-score	Training time
XGBoost	93.36%	90.11%	$112s$
Random Forest	80.16%	77.52%	$4.3s$
Extra Trees	86.82%	82.80%	$2.06s$
LightGBM	95.09%	90.47%	$84s$
Stacking	96.52%	96.05%	$512s$

methods

$$\text{Acc} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{FN} + \text{TN}} \tag{22}$$

$$F1 = \frac{\text{TP}}{\text{TP} + \frac{\text{FN}+\text{FP}}{2}} \tag{23}$$

where the *TP* is the true-positive number, *FP* is the false-positive number, *TN* is the true-negative number, and *FN* is the false-negative number. Besides, the training time is the average time we spend on the training model.

The results are presented in Table VII. It can be seen that the accuracies of RF and ET-based methods are not more than 85%, which cannot satisfy the engineering requirements obviously, while the accuracies of XGB and LGB-based methods can achieve about 90%. On the other hand, it is evident that the proposed stacking method has the best accuracy and $F1$-score compared with other single model methods, which satisfies the requirements of the CRRC Zhuzhou Research Institute. The proposed stacking method is able to diagnose incipient abrupt faults and incipient time-varying faults effectively.

Moreover, the confusion matrices are given in Figs. 3 and 4, in which the horizontal and vertical values represent the corresponding fault status in Table IV. The diagonal elements are the correct numbers of the diagnosed test samples, and the remaining elements are the number of misdiagnosed test samples. For the single model (XGB, RF, ET, and LGB), there are amounts of misdiagnosis on label 1 and label 2, which means that the single-model-based methods are hard to distinguish the motor current sensor incipient fault and healthy condition. However, the confusion matrix in Fig. 4 implies that the misdiagnosis of incipient faults can be substantially reduced by stacking.

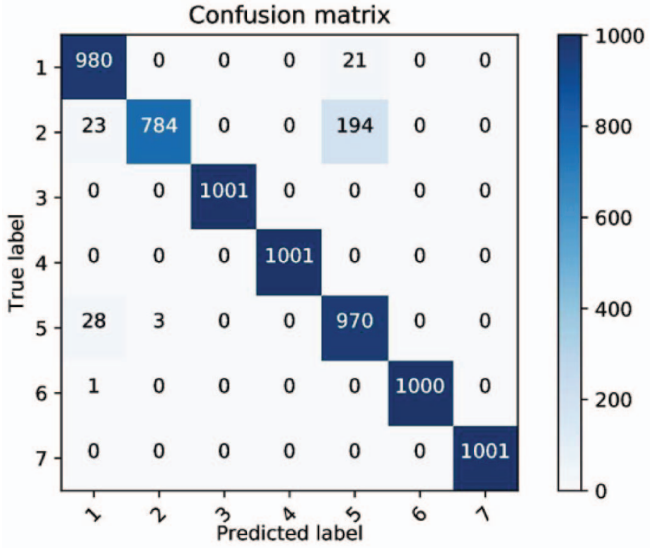

Fig. 4. Confusion matrix of the stack model.

It should be noted that although the training time of the stacking model increases, with the generalization performance of stacking, the model can be trained offline and used online. On the other hand, under the train net, the data can be transferred to the on-the-ground monitor center to be handled by computers with higher computing power.

VI. Conclusion

In this work, the stacked generalization (Stacking)-based incipient fault diagnosis scheme for the traction system of high-speed trains was proposed. The XGBoost, RF, ET, and LightGBM were chosen as the base estimators in the first layer of the stacking to extract the fault feature from the faulty data signals. Then, the LR model was employed as the meta estimator in the second layer to combine the results with the base estimators for fault classification. Moreover, the PIO algorithm was utilized to find the optimal hyperparameters of the base estimators. Finally, the proposed method was verified on a semiphysical platform of the CRH2 traction system. The experimental results showed the proposed approach can diagnose both the incipient faults and the abrupt faults effectively.

In our future work, the compound faults that represent the different faults occurring simultaneously could be studied. We will suppose to employ a multiclass learning model to achieve some possible categories. Moreover, if the remaining lift of the system under an incipient fault can be obtained, the repairing time will be shorted, which motivates the fault prediction. Finally, it would be better to recover the system performance after the fault occurs and cannot be dealt with immediately, especially for the high-speed trains that cannot stop suddenly. So the fault-tolerant control design [42], [43] is also a meaningful research direction.

References

[1] H. Ji, Z. Hou, and R. Zhang, "Adaptive iterative learning control for high-speed trains with unknown speed delays and input saturations," *IEEE Trans. Autom. Sci. Eng.*, vol. 13, no. 1, pp. 260–273, Jan. 2016.

[2] H. Dong, B. Ning, B. Cai, and Z. Hou, "Automatic train control system development and simulation for high-speed railways," *IEEE Circuits Syst. Mag.*, vol. 10, no. 2, pp. 6–18, 2nd Quart., 2010.

[3] R. Cheng, W. Yu, Y. Song, D. Chen, X. Ma, and Y. Cheng, "Intelligent safe driving methods based on hybrid automata and ensemble CART algorithms for multihigh-speed trains," *IEEE Trans. Cybern.*, vol. 49, no. 10, pp. 3816–3826, Oct. 2019.

[4] X. Yao, J. H. Park, H. Dong, L. Guo, and X. Lin, "Robust adaptive nonsingular terminal sliding mode control for automatic train operation," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 49, no. 12, pp. 2406–2415, Dec. 2019.

[5] X. Yao, L. Wu, and L. Guo, "Disturbance-observer-based fault tolerant control of high-speed trains: A Markovian jump system model approach," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 50, no. 4, pp. 1476–1485, Apr. 2020.

[6] Y. Wang, Y. D. Song, H. Gao, and F. L. Lewis, "Distributed fault-tolerant control of virtually and physically interconnected systems with application to high-speed trains under traction/braking failures," *IEEE Trans. Intell. Transp. Syst.*, vol. 17, no. 2, pp. 535–545, Feb. 2016.

[7] D. Zhou, H. Ji, X. He, and J. Shang, "Fault detection and isolation of the brake cylinder system for electric multiple units," *IEEE Trans. Control Syst. Technol.*, vol. 26, no. 5, pp. 1744–1757, Sep. 2018.

[8] S. Wang, W. Hsiung, M. Chiang, and Y. Tsai, "Early stopping fault diagnosis agreement on wireless sensor network of IoT," *Int. J. Innovat. Comput. Inf. Control*, vol. 15, no. 4, pp. 1351–1364, 2019.

[9] Y. Zheng, J. Gao, L. Yi, Y. Qin, and H. Dong, "The fault diagnosis system design of train auxiliary inverter based on LMD and RBFNN," *ICIC Exp. Lett. B, Appl.*, vol. 8, no. 2, pp. 421–427, Feb. 2017.

[10] H. Chen and B. Jiang, "A review of fault detection and diagnosis for the traction system in high-speed trains," *IEEE Trans. Intell. Transp. Syst.*, vol. 21, no. 2, pp. 450–465, Feb. 2020.

[11] Y. Song, Q. Song, and W. Cai, "Fault-tolerant adaptive control of high-speed trains under traction/braking failures: A virtual parameter-based approach," *IEEE Trans. Intell. Transp.*, vol. 15, no. 2, pp. 737–748, Apr. 2014.

[12] T. A. Najafabadi, F. R. Salmasi, and P. Jabehdar-Maralani, "Detection and isolation of speed-, dc-link voltage-, and current-sensor faults based on an adaptive observer in induction-motor drives," *IEEE Trans. Ind. Electron.*, vol. 58, no. 5, pp. 1662–1672, May 2011.

[13] Z. Mao, G. Tao, B. Jiang, and X. Yan, "Adaptive actuator compensation of position tracking for high-speed trains with disturbances," *IEEE Trans. Veh. Technol.*, vol. 67, no. 7, pp. 5706–5717, Jul. 2018.

[14] H. Dong, X. Lin, S. Gao, B. Cai, and B. Ning, "Neural networks-based sliding mode fault-tolerant control for high-speed trains with bounded parameters and actuator faults," *IEEE Trans. Veh. Technol.*, vol. 69, no. 2, pp. 1353–1362, Feb. 2020.

[15] S. Kommuri, M. Defoort, H. Karimi, and K. Veluvolu, "A robust observer-based sensor fault-tolerant control for PMSM in electric vehicles," *IEEE Trans. Ind. Electron.*, vol. 63, no. 12, pp. 7671–7681, Dec. 2016.

[16] K. Zhang, B, Jiang, X. Yan, and Z. Mao, "Sliding mode observer based incipient sensor fault detection with application to high-speed railway traction device," *ISA Trans.*, vol. 63, pp. 49–59, Jul. 2016.

[17] W. Bai, H. Dong, X. Yao, and B. Ning, "Robust fault detection for the dynamics of high-speed train with multi-source finite frequency interference," *ISA Trans.*, vol. 75, pp. 76–87, Apr. 2018.

[18] S. X. Ding, *Model-based Fault Diagnosis Techniques*. Berlin, Germany: Springer, 2008.

[19] Q. Song and Y. Song, "Data-based fault-tolerant control of high-speed trains with traction/braking notch nonlinearities and actuator failures," *IEEE Trans. Neural Netw.*, vol. 22, no. 12, pp. 2250–2261, Dec. 2011.

[20] B. Cai, Y. Zhao, H. Liu, and M. Xie, "A data-driven fault diagnosis methodology in three-phase inverters for PMSM drive systems," *IEEE Trans. Power Electron.*, vol. 32, no. 7, pp. 5590–5600, Jul. 2017.

[21] 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 Trans. Ind. Electron.*, vol. 65, no. 2, pp. 1559–1567, Feb. 2018.

[22] S. Zhang, C. Zhao, and F. Gao, "Incipient fault detection for multiphase batch processes with limited batches," *IEEE Trans. Control Syst. Technol.*, vol. 27, no. 1, pp. 103–117, Jan. 2019.

[23] J. Wang, J. Zhang, B. Qu, H. Wu, and J. Zhou, "Unified architecture of active fault detection and partial active fault-tolerant control for incipient faults," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 47, no. 7, pp. 1688–1700, Jul. 2017.

[24] T. Chen, C. Wang, G. Chen, Z. Dong, and D. J. Hill, "Small fault detection for a class of closed-loop systems via deterministic learning," *IEEE Trans. Cybern.*, vol. 49, no. 3, pp. 897–906, Mar. 2019.
[25] H. Ji, X. He, J. Shang, and D. Zhou, "Incipient sensor fault diagnosis using moving window reconstruction-based contribution," *Ind. Eng. Chem. Res.*, vol. 55, no. 10, pp. 2746–2759, Mar. 2016.
[26] S. Munikoti, L. Das, B. Natarajan, and B. Srinivasan, "Data-driven approaches for diagnosis of incipient faults in dc motors," *IEEE Trans. Ind. Informat.*, vol. 15, no. 9, pp. 5299–5308, Sep. 2019.
[27] 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 Trans. Veh. Technol.*, vol. 67, no. 6, pp. 4819–4830, Jun. 2018.
[28] D. H. Wolpert, "Stacked generalization," *Neural Netw.*, vol. 5, no. 2, pp. 241–259, 1992.
[29] G. Wang, J. Hao, J. Ma, and H. Jiang, "A comparative assessment of ensemble learning for credit scoring," *Expert Syst. Appl.*, vol. 38, no. 1, pp. 223–230, Jan. 2011.
[30] K. Ting and M. Witten, "Issues in stacked generalization," *J. Artif. Intell. Res.*, vol. 10, pp. 271–289, May 1999.
[31] A. Pernia-Espinoza, J. Fernandez-Ceniceros, J. Antonanzas, R. Urraca, and F. J. Martinez-De-Pison, "Stacking ensemble with parsimonious base models to improve generalization capability in the characterization of steel bolted components," *Appl. Soft. Comput.*, vol. 70, pp. 737–750, Sep. 2018.
[32] X. Yang, C. Yang, T. Peng, Z. Chen, B. Liu, and W. Gui, "Hardware-in-the-loop fault injection for traction control system," *IEEE J. Emerg. Sel. Topics Power Electron.*, vol. 6, no. 2, pp. 696–706, Jun. 2018.
[33] C. Yang, C. Yang, T. Peng, X. Yang, and W. Gui, "A fault-injection strategy for traction drive control systems," *IEEE Trans. Ind. Electron.*, vol. 64, no. 7, pp. 5719–5727, Jul. 2017.
[34] Z. Mao, Y. Zhan, G. Tao, B. Jiang, and X. Yan, "Sensor fault detection for rail vehicle suspension systems with disturbances and stochastic noises," *IEEE Trans. Veh. Technol.*, vol. 66, no. 6, pp. 4691–4705, Jun. 2017.
[35] J. Feng, J. Xu, W. Liao, and Y. Liu, "Review on the traction system sensor technology of a rail transit train," *Sensors (Basel, Switzerland)*, vol. 17, no. 6, pp. 1–16, Jun. 2017.
[36] S. Wang, J. Yang, and K. Chou, "Using stacked generalization to predict membrane protein types based on pseudo-amino acid composition," *J. Theor. Biol.*, vol. 242, no. 4, pp. 941–946, 2006.
[37] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.*, Aug. 2016, pp. 785–794.
[38] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, Oct. 2001.
[39] P. Geurts, D. Ernst, and L. Wehenkel, "Extremely randomized trees," *Mach. Learn.*, vol. 63, no. 1, pp. 3–42, Apr. 2006.
[40] G. Ke *et al.*, "LightGBM: A highly efficient gradient boosting decision tree," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 3146–3154.
[41] H. Duan and P. Qiao, "Pigeon-inspired optimization: A new swarm intelligence optimizer for air robot path planning," *Int. J. Intell. Comput. Cybern.*, vol. 7, no. 1, pp. 24–37, Mar. 2014.
[42] L. Liu, Y. Liu, D. Li, S. Tong, and Z. Wang, "Barrier Lyapunov function-based adaptive fuzzy FTC for switched systems and its applications to resistance–inductance–capacitance circuit system," *IEEE Trans. Cybern.*, vol. 50, no. 8, pp. 3491–3502, Aug. 2020.
[43] R. Gao, Y. Wang, J. Lai, and H. Gao, "Neuro-adaptive fault-tolerant control of high speed trains under traction-braking failures using self-structuring neural networks," *Inf. Sci.*, vols. 367–368, no. 1, pp. 449–462, Nov. 2016.

Zehui Mao (Member, IEEE) received the Ph.D. degree in control theory and control engineering from the Nanjing University of Aeronautics and Astronautics, Nanjing, China, in 2009.

She is currently a Professor with the College of Automation Engineering, Nanjing University of Aeronautics and Astronautics. She was a visiting scholar with the University of Virginia, Charlottesville, VA, USA. She worked in the areas of fault diagnosis, with particular interests in nonlinear control systems, sampled-data systems, and networked control systems. Her current research interests include fault diagnosis and fault-tolerant control of systems with disturbance and incipient faults, and high-speed train and spacecraft flight control applications.

Mingxuan Xia received the B.S. degree from Jiangnan University, Wuxi, China, in 2018. She is currently pursuing the M.S. degree with the College of Automation Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing, China.

Her current research interests include incipient fault diagnosis, machine-learning-based fault diagnosis, and their application to the traction systems and the gearbox.

Bin Jiang (Fellow, IEEE) received the Ph.D. degree in automatic control from Northeastern University, Shenyang, China, in 1995.

He is currently a Chair Professor of the Cheung Kong Scholar Program with the Ministry of Education and the Vice President of the Nanjing University of Aeronautics and Astronautics, Nanjing, China. He has authored eight books and over 200 referred international journal papers and conference papers. His current research interests include intelligent fault diagnosis and fault-tolerant control and their applications to helicopters, satellites, and high-speed trains.

Dr. Jiang currently serves as an Associate Editor or an Editorial Board Member for a number of journals, such as the IEEE TRANSACTIONS ON CYBERNETICS and *Neurocomputing*. He is the Chair of Control Systems Chapter in IEEE Nanjing Section, and a Member of the IFAC Technical Committee on Fault Detection, Supervision, and Safety of Technical Processes.

Dezhi Xu (Senior Member, IEEE) received the Ph.D. degree in control theory and control engineering from the Nanjing University of Aeronautics and Astronautics, Nanjing, China, in 2013.

He was a Visiting Fellow with the Department of Biomedical Engineering, City University of Hong Kong, Hong Kong, from 2018 to 2019. He has been with Jiangnan University Wuxi, China, since 2014, as an Associate Professor. His current research interests include data-driven control, fault diagnosis and fault-tolerant control, multiagent systems and CPSs, technologies of renewable energy, motor control, and smart grid.

Dr. Xu was a recipient of the First Class Award of Science and Technology Progression from the China General Chamber of Commerce in 2016 for his research results. He is a Committee Member of the Association of Energy Internet, and Trusted Control in Chinese Association of Automation, and a Committee Member of energy storage in China Renewable Energy Society.

Peng Shi (Fellow, IEEE) received the Ph.D. degree in electrical engineering from the University of Newcastle, Callaghan, NSW, Australia, in 1994, the D.Sc. degree from the University of Glamorgan, Pontypridd, U.K., in 2006, and the D.Eng. degree from the University of Adelaide, Adelaide, SA, Australia, in 2015.

He is currently a Professor with the University of Adelaide. His research interests include intelligent systems, autonomous and robotic systems, network systems, and cyber-physical systems.

Prof. Shi has served on the editorial board of a number of journals, including *Automatica*, IEEE TRANSACTIONS ON AUTOMATIC CONTROL, IEEE TRANSACTIONS ON CYBERNETICS, IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS—I: REGULAR PAPERS, IEEE TRANSACTIONS ON FUZZY SYSTEMS, and IEEE CONTROL SYSTEMS LETTERS. He is currently a Member-at-Large of Board of Governors, IEEE SMC Society, and an IEEE SMC Distinguished Lecturer. He is a Fellow of the Institution of Engineering and Technology and the Institute of Engineers, Australia.