Journal of Dynamics, Monitoring and Diagnostics, 2025, 4, 111-120
https://doi.org/10.37965/jdmd.2025.784

ISTP
RESEARCH ARTICLE

Dynamic Relative Advantage-Driven Multi-Fault Synergistic Diagnosis Method for Motors under Imbalanced Missing Data Rates

Zhenpeng Teng,[1,2] Xiaojian Yi,[1,2] and Biao Wang[3]
[1]School of Mechatronical Engineering, Beijing Institute of Technology, Beijing, China
[2]Yangtze Delta Region Academy, Beijing Institute of Technology, Jiaxing, China
[3]State Key Laboratory of Advanced Rail Autonomous Operation, Beijing Jiaotong University, Beijing, China

(Received 16 April 2025; Revised 08 May 2025; Accepted 27 May 2025; Published online 28 May 2025)

Abstract: Missing data handling is vital for multi-sensor information fusion fault diagnosis of motors to prevent the accuracy decay or even model failure, and some promising results have been gained in several current studies. These studies, however, have the following limitations: 1) effective supervision is neglected for missing data across different fault types and 2) imbalance in missing rates among fault types results in inadequate learning during model training. To overcome the above limitations, this paper proposes a dynamic relative advantage-driven multi-fault synergistic diagnosis method to accomplish accurate fault diagnosis of motors under imbalanced missing data rates. Firstly, a cross-fault-type generalized synergistic diagnostic strategy is established based on variational information bottleneck theory, which is able to ensure sufficient supervision in handling missing data. Then, a dynamic relative advantage assessment technique is designed to reduce diagnostic accuracy decay caused by imbalanced missing data rates. The proposed method is validated using multi-sensor data from motor fault simulation experiments, and experimental results demonstrate its effectiveness and superiority in improving diagnostic accuracy and generalization under imbalanced missing data rates.

Keywords: data missing; motor fault; relative advantage; synergistic diagnosis

I. INTRODUCTION

The motor is a crucial power component in sophisticated equipment like armored vehicles, aircraft, and ships, significantly influencing overall performance and reliability [1]. During operation, motors are susceptible to faults from high-frequency operation, inadequate cooling, deferred maintenance, or power variations [2]. If no action is taken in time, these faults will directly affect the equipment's performance, even resulting in substantial economic losses and casualties [3]. Hence, accurate and effective fault diagnosis technology is vital for ensuring equipment safety and improving reliability.

The advancement of industrial information technology has led to the widespread utilization of various sensors in equipment health monitoring, including those for vibration, temperature, and pressure. These sensors are employed to monitor the real-time and comprehensive service status of critical components and systems [4]. Concurrently, deep learning models have gained prominence in the domain of multi-sensor information fusion fault diagnosis due to their exceptional capacity for nonlinear characterization learning [5]. By leveraging deep learning, diagnostic methods can establish a direct mapping from multi-sensor data to fault classification, obviating the necessity for elucidating fault mechanisms or engaging in intricate feature engineering [6]. In recent years, numerous deep learning-based approaches coupled with multi-sensor data have demonstrated efficacy in motor fault diagnosis. For instance, Wei *et al.* [7] proposed a fault diagnosis model based on multi-head self-attention and multi-sensor deep feature fusion network. Li *et al.* [8] proposed a multi-sensor information fusion fault diagnosis model that integrated a two-stage recurrent neural network with a convolutional block attention module. Li *et al.* [9] designed a multi-view graph neural network that combines time domain and frequency domain features for multi-sensor information fusion fault diagnosis.

High-quality multi-sensor datasets are essential for achieving high diagnostic accuracy for deep learning model-based methods in fault diagnosis of motors [10], but it is inevitable in real-world industrial applications that the multi-sensor data collected by monitoring sensors may contain imperfect data, such as drift, loss, mutation, and empty acquisitions, due to the complexity of equipment operating conditions, variability of combat environments, sensor degradation, and other factors [11]. Utilizing such inferior multi-sensor data directly in fault diagnosis algorithms can significantly impact diagnostic performance, leading to reduced accuracy and potential algorithm failure [12,13]. In response to this dilemma, researchers have focused on handling missing data in multi-sensor information fusion fault diagnosis of motors. For instance, Yu *et al.* [14] introduced a novel fault data generation approach that combined physical models and periodic generation to address network variations in bearing fault diagnosis. Shao *et al.* [15] integrated complete data and uncertain information using dynamic Bayesian networks to mitigate data integrity issues and minimize prediction errors. Song *et al.* [16] proposed a fuzzy clustering framework that incorporated weighted non-negative latent factor analysis to cluster incomplete data with high precision. Yang *et al.* [17] presented a data enhancement method based on a diffusion probability model for fault diagnosis in scenarios

Corresponding author: Xiaojian Yi (e-mail: xjyi@bit.edu.cn).

with unbalanced data. In these studies, the reconstruction network based on data generation mechanisms, such as generative adversarial networks, can effectively alleviate the impact of poor-quality data on diagnostic performance, using physical constraints. Thereby, these networks can achieve excellent diagnostic performance under inadequate fault samples. More importantly, they do not require additional structures or auxiliary data.

Although existing methods have achieved promising results in missing data processing, there are still some limitations as follows:

1) They choose to either directly mask the missing data or only consider reconstructing the missing data based on the features of gettable data. Direct masking avoids the interference of missing data on forward feature extraction, but it overlooks potentially useful information. The method that only considers the obtained data features improves data utilization efficiency, but it neglects the effective supervision of missing data across different fault types, which may lead to the performance decline of diagnostic models, especially in the scenarios of high missing data rates. Therefore, a new missing data handling methodology that considers missing data supervision is needed to guide the intelligent diagnosis network to more effectively learn and classify fault types.
2) They consider the issue of missing data but overlook the impact of the imbalance in missing rates among fault types. In industrial scenarios, the operating conditions of motors inevitably change, and the missing data rates of different fault types often exhibit significant differences. The missing rate for certain fault types may be as low as 20%, while for others it may be as high as 80%. This discrepancy triggers the channel imbalance problem, where the strong channels with less missing data dominate model training, while the weak channels with more missing data may be neglected, resulting in inadequate training for some fault types and affecting overall diagnostic performance. Therefore, it is key to effectively balance the training data of different fault types under data missing conditions, ensuring that each fault type is sufficiently learned.

To address the aforementioned issues, this paper proposes a dynamic relative advantage-driven multi-fault synergistic diagnosis method to accomplish accurate fault diagnosis of motors under imbalanced missing data rates. In the proposed method, there are two distinctive compositions: a cross-fault-type generalized synergistic diagnosis strategy and a dynamic relative advantage assessment technique. The cross-fault-type generalization synergistic diagnostic strategy focuses on using the variational information bottleneck theory to obtain an optimal representation of fault types, which is able to ensure sufficient supervision in handling missing data. The dynamic relative advantage assessment technique is composed of a reconstruction loss and a two-layer collaborative optimization algorithm, which reduces diagnostic accuracy degradation caused by missing data imbalance. The effectiveness of the proposed method is validated using multi-sensor data from motor fault simulation experiments. The contributions of this work can be summarized as follows:

1) A cross-fault-type generalized synergistic diagnosis strategy based on variational information bottleneck theory is proposed to enable reconfigured networks to be adequately supervised when missing data are made up. Through minimizing the information bottleneck loss for redundancy reduction and leveraging label-based supervision on missing data, cross-type generalizable representations are progressively generated from the optimized features of other fault types. Therefore, the method can guide the reconfiguration network to maximize the retention of task-relevant features, thereby achieving higher diagnostic accuracy under missing data.
2) A dynamic relative advantage assessment technique is designed to quantify the expressive capability of different fault types, which can facilitate collaborative learning of different fault types. On the one hand, the relative advantage of each fault type is defined dynamically based on the reconfiguration loss. On the other hand, a two-layer collaborative optimization algorithm is designed to dynamically adjust the supervisory weights by means of a synergistic training and adaptive supervisory adjustment, which makes the inductive loss gap between fault types smaller. Therefore, this technique can avoid the degradation of fault diagnosis accuracy caused by imbalance in data missing rate.

The article is structured as follows: Section II presents the pertinent preliminary research; Section III delineates the proposed methodology; Section IV validates the method's efficacy through simulation experiments and comparative analysis with established approaches; Section V encapsulates the paper findings and outlines potential future research directions.

II. PRELIMINARY

A. DEFINITION OF MISSING DATA

Missing data are defined as two types: incomplete data and complete missing data. Incomplete data refer to situations where there are parts of the sensor monitoring data that are missing, usually due to sensor failures or other limitations that result in the loss of certain data segments, but still result in partially valid data. In contrast, complete missing data refer to a situation where the sensor fails to monitor any data, leaving the relevant portion completely empty. This is usually caused by equipment failures, communication breakdowns, or other serious problems that prevent access to any valid information.

B. SYMBOL DEFINITION

Assume that the training data contain K types of faults, and each fault type includes L sensor data. The concatenation of all fault data is utilized to define an auxiliary fault, leading to a total of $L+1$ data entries for simplicity. The training samples are denoted by $\{Z_l^k\}_{k\in[K]}$ and $\{W_l^k\}_{K\in[K+1]}$, where $k \in [K]$ denotes the sample serial number, $Z_l^k \in R^{d_l}$ denotes the feature vector of the sensor, and W_l^k denotes the corresponding fault label. For consistency of expression, all features of fault type L are collected $\boldsymbol{Z}_L^{K+1} = [Z_L^1; Z_L^2; \cdots; Z_L^K]$. senc denotes the adaptive encoder, $\mathfrak{R}_{senc}^k$ denotes its corresponding parameter, and xenc denotes the interactive residual autoencoder, $\mathfrak{R}_{xenc}^k$ denotes its corresponding parameter. $\mathfrak{R}_{dec}^{K+1}$ denotes the parameters of the decoder.

III. DYNAMIC RELATIVE ADVANTAGE-DRIVEN MULTI-FAULT SYNERGISTIC DIAGNOSIS METHOD

The proposed method is composed of two parts: a cross-fault-type generalized synergistic diagnosis strategy and a dynamic relative advantage assessment technique. As shown in Fig. 1, the model architecture is depicted. Stage I consists of the training process of samples from three fault types with complete data, while stages II and III are used to illustrate the model training and inference methods under missing data conditions. In short, the features Z_k^l of the sensors corresponding to each fault type are mapped to their representations $\bar{y}_l^k$ through a parameterized variational encoder $\mathfrak{R}_{senc}^k$. Subsequently, each fault type final representation $\hat{y}_l^k$ is constructed by a parameterized $\mathfrak{R}_{xenc}^k$ cross-fault-type encoder using representations from other fault types $\{\bar{y}_l^k\}_{k\neq k}$, rather than relying solely on its own self-representation $\bar{y}_l^k$.

A. GENERALIZED SYNERGISTIC DIAGNOSIS STRATEGY ACROSS FAULT TYPES

To ensure that the reconstruction module receives adequate guidance under missing data training framework and that each fault type obtains its optimal representation during learning, this section proposes a cross-fault-type

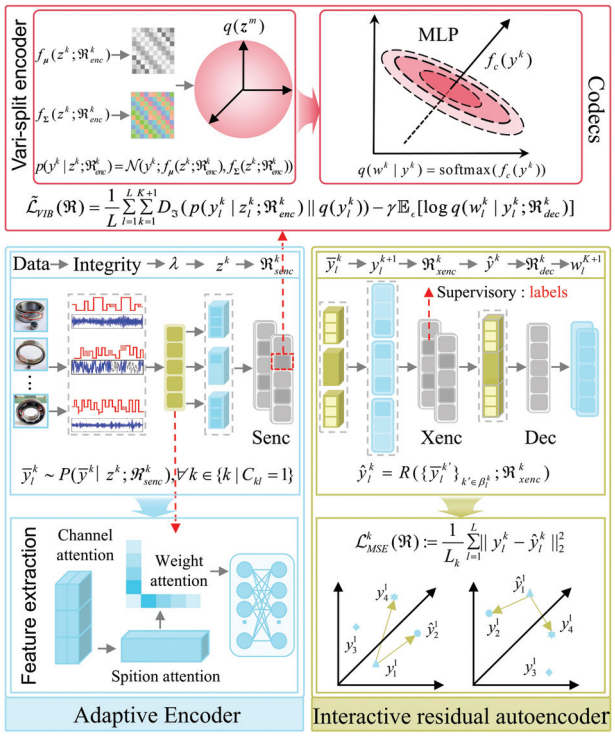

Fig. 2. A cross-fault-type generalized synergistic diagnosis strategy.

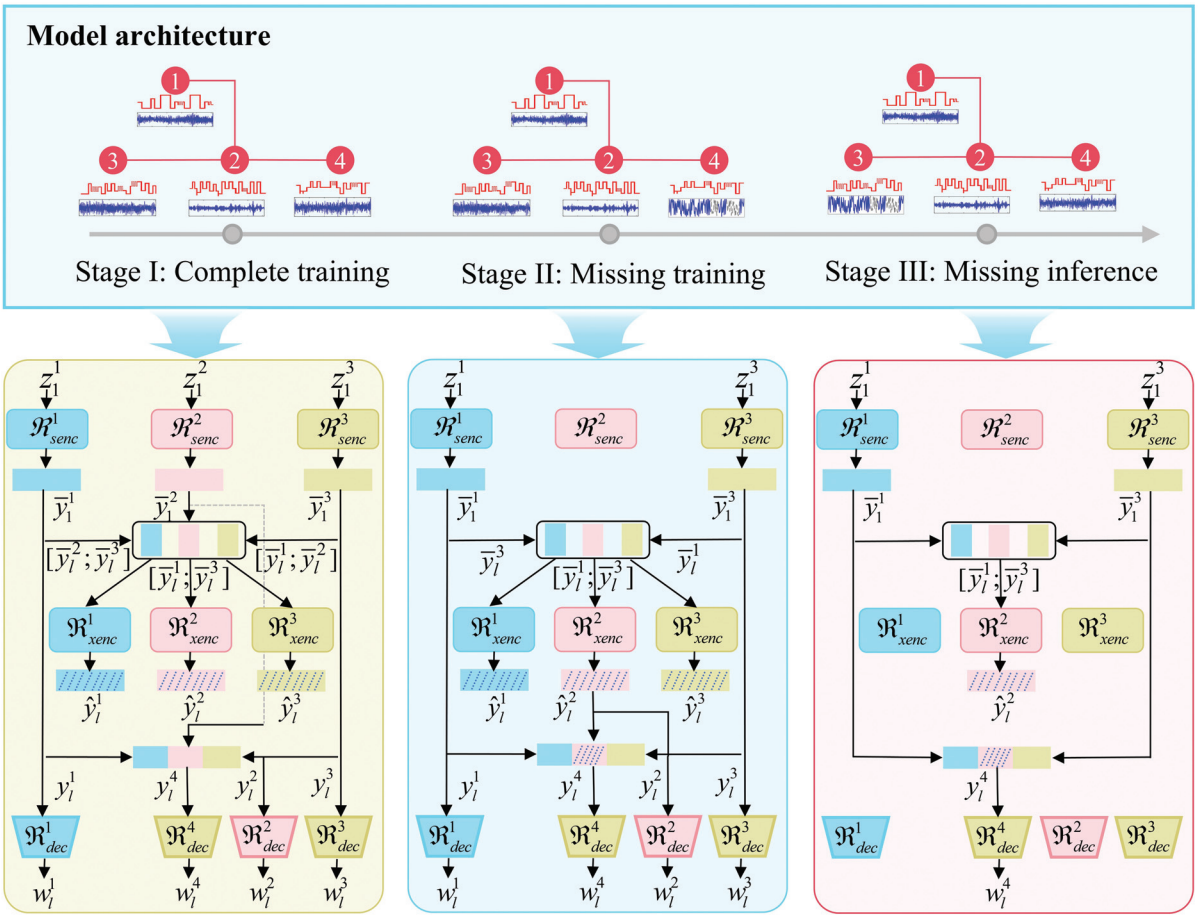

Fig. 1. Model architecture of the dynamic relative advantage-driven synergistic fault diagnosis method.

Algorithm: Synergistic Fault Diagnosis Strategy

Input: Sensor feature matrix $\boldsymbol{Z}_L^{K+1}$,
Completeness matrix $C = \{C_{kl}\}$
Output: Final representations $\hat{y}_l^k$, auxiliary
fault y_l^{K+1}, shared labels W_l^k
Initialize: model parameters θ
for $l \in [1, L]$ do
    for $k \in [1, K]$ do // Self-Representation
via Variational Encoder
        if $C_{kl} = 1$ then
            $\bar{y}_l^k \leftarrow P(\bar{y}^k | \bar{z}^k; \mathfrak{R}_{senc}^k)$
        else $\bar{y}_l^k \sim null$
        $\beta_l^k = \{k' \in [K] | \, C_{lk'} = 1 \,, \; k' \neq k\}$
            // Available fault types
    if $|\beta_l^k| > 0$ then
        $\hat{y}_l^k \leftarrow R(\{\bar{y}_l^{k'}\}_{k' \in \beta_l^k}; \mathfrak{R}_{xenc}^k)$
        else $\hat{y}_l^k = null$
        if $C_{kl} = 1$ then
            $y_k^l \leftarrow \bar{y}_l^k$
        else if $\hat{y}_l^k \neq$ null then
            $y_k^l \leftarrow \hat{y}_l^k$
end for
    $y_l^{K+1} = [y_l^1; y_l^2; \cdots; y_l^K]$
for $k \in [1, K+1]$ do
    $w_l^k \sim R(w^k | y^k; \mathfrak{R}_{dec}^k)$ if $y_k^l \notin null$
  else null
end for
valid $\leftarrow \{W_l^k | W_l^k \neq$ null$\}$
if |valid| >0 then
    majority$\leftarrow$ argmax(w$\in$valid) count(w)
    $W_l^k \leftarrow$majority $\forall \; k \in [1, K+1]$
  end if
end for
return y_1^k, y_l^{K+1}, W_l^k

generalized synergistic diagnostic strategy. As illustrated in Fig. 2, this strategy generates a generalized diagnosis across fault types using the optimal representations of other fault types, regardless of whether the data is missing or available.

1. SYNERGISTIC DIAGNOSTIC STRATEGY DERIVATION. For different sensor features Z_l^k, the generalized synergistic characterization y_l^k of fault types is derived as follows. First, a sample is taken from the complete sensor feature set, denoted as $\bar{y}_l^k \sim P(\bar{y}^k | z^k; \; \mathfrak{R}_{senc}^k), \forall k \in \{k | C_{kl} = 1\}$. Then, the cross-fault type representation for all sensors is calculated as $\hat{y}_l^k = R(\{\bar{y}_l^{k'}\}_{k' \in \beta^k}; \mathfrak{R}_{xenc}^k)$, where β_l^k is defined as the set of all sensor features excluding the features of sensor k itself, denoted as $\beta_l^k = \{k' \in [K], C_{lk'} = 1 \,, \; k' \neq k\}$. Finally, using the above symbols, the final representation of feature $z_k^l, \forall l \in [L], \forall k \in [K]$ can be reduced to $y_k^l = \{\bar{y}_l^k$, if $C_{kl} = 1$; $\hat{y}_l^k$, otherwise.$\}$. Meanwhile, the representation of the auxiliary fault *K+1* is the concatenation of the sensor features of all fault types, denoted as $y_l^{K+1} \coloneqq [y_l^1; y_l^2; \cdots; y_l^K]$, and the encoder parameters for all fault types k are collected using $\mathfrak{R}_{enc}^k = \{\mathfrak{R}_{senc}^k \,, \; \mathfrak{R}_{xenc}^k\}$. When all encoder parameters are gathered using $\mathfrak{R}_{enc}^{K+1} \coloneqq \{\mathfrak{R}_{enc}^K\}_{k \in [K]}$, the above process can be summarized as $y_l^k \sim R(y^k | z^k; \mathfrak{R}_{enc}^k)$. For typical multi-fault-multi-sensor synergistic fault diagnosis, the generalized synergistic representation of the auxiliary fault y_l^{K+1} is used for the final sensor fault label prediction, which is equivalent to passing the representation to a decoder with parameter $\mathfrak{R}_{dec}^{K+1}: w_l^{k+1} \sim R(w^{k+1} | y^{k+1}; \; \mathfrak{R}_{dec}^{k+1})$. Additionally, in this strategy, each individual sensor also performs fault type label prediction, denoted as $\mathfrak{R}_{dec}^k: w_l^k \sim R(w^k | y^k; \mathfrak{R}_{dec}^k) \,, \; k \in [K]$, and all sensors share the same label W_l^{k+1}. The above algorithm illustrates the derivation of the strategy.

2. PARAMETER OPTIMIZATION. The learning objective based on the variational information bottleneck is to maximize the target information while effectively compressing the original features. In the generalized synergistic diagnosis method across fault types, this objective can be transformed into an optimization problem, which is achieved by minimizing the Information Bottleneck (IB) loss.

$$\min_{\mathfrak{R}} \mathrm{L}_{IB}(\mathfrak{R}) \coloneqq \sum_{k \in [K+1]} I(z^k \,, y^k; \mathfrak{R}) - \lambda I(y^k \,, w^k; \mathfrak{R}) \tag{1}$$

where $\mathfrak{R} = \{\mathfrak{R}_{enc}^k \,, \mathfrak{R}_{dec}^k\}_{k \in [K+1]}$ collects all model parameters, $I(\cdot \;, \cdot)$ represents the mutual information between two random variables, and the "bottleneck" width, denoted by λ, realizes the filtering of redundant information and the coefficients $\lambda > 0$ are two of the balanced predefined constant loss functions. However, since calculating the cross-fault type information $I(z^k \,, y^k; \mathfrak{R})$ and $I(y^k \,, w^k; \mathfrak{R})$ is difficult, this paper employs the variational upper bound of $\mathscr{L}_{IB}(\mathfrak{R})$:

$$\begin{aligned} \mathrm{L}_{IB}(\mathfrak{R}) \leq \mathrm{L}_{VIB}(\mathfrak{R}) = \sum_{k=1}^{K+1} \mathrm{E} \\ [D_{\mathcal{J}}(p(y^k | z^k; \mathfrak{R}_{enc}^k) \| q(y^k)) - \lambda \, \log \, q(w^k | y^k; \mathfrak{R}_{dec}^k)] \end{aligned} \tag{2}$$

where $E[\cdot]$ is the expectation for the random variables Z^k, Y^k, and W^k; $D_{\mathfrak{J}}(\cdot \parallel \cdot)$ represents the KL divergence between two distributions; and $q(y^k)$ and $q(w^k | y^k; \mathfrak{R}_{dec}^k)$ represent the variational distributions approximating $p(y^k | z^k; \mathfrak{R}_{enc}^k)$ and $q(w^k | y^k; \mathfrak{R}_{dec}^k)$.

The variational encoder adopts a Gaussian form $p(y^k | z^k; \mathfrak{R}_{enc}^k) = N(y^k; f_{\boldsymbol{\mu}}(z^k; \mathfrak{R}_{enc}^k) \,, \; f_{\boldsymbol{\Sigma}}(z^k; \mathfrak{R}_{enc}^k))$, where $f_{\boldsymbol{\mu}}(\cdot; \mathfrak{R}_{enc}^k) \in \mathrm{R}^d$ and $f_{\boldsymbol{\Sigma}}(\cdot; \mathfrak{R}_{enc}^k) \in \mathrm{R}^{d \times d}$ are the mean vector and covariance matrix determined by the network model, and $q(y^k)$ is fixed as a dimensional spherical Gaussian, i.e., $q(y^k) = N(y^k; \mathbf{0}_d \,, \; \boldsymbol{I}_d)$, where 0d and Id are the d-dimensional zero vector and the d × d identity matrix, respectively. The decoder is a simple classifier model in the form of $q(w^k | y^k) = \mathrm{softmax}(f_c(y^k))$, where $f_c(y^k) \,, \; k \in [K+1]$ maps the representation to the logic used for classification through an MLP, resulting in the following:

$$y^k = f_{\boldsymbol{\mu}}(z^k; \mathfrak{R}_{enc}^k) + \grave{o} \odot f_{\boldsymbol{\Sigma}}(z^k; \mathfrak{R}_{enc}^k)) \tag{3}$$

where $\odot$ denotes the element-wise product. Therefore, $\mathrm{L}_{VIB}(\mathfrak{R})$ can be approximated as the following $\tilde{\mathrm{L}}_{VIB}(\mathfrak{R})$:

$$\begin{aligned} \tilde{\mathrm{L}}_{VIB}(\mathfrak{R}) = \frac{1}{L} \sum_{l=1}^{L} \sum_{k=1}^{K+1} D_{\mathcal{J}}(p(y_l^k | z_l^k; \mathfrak{R}_{enc}^k) \| q(y^k)) \\ - \gamma E_{\grave{o}}[\, \log \, q(w_l^k | y_l^k; \mathfrak{R}_{dec}^k)] \end{aligned} \tag{4}$$

where ò is the random variable in the reparameterization trick.

This formula bypasses the expectation in the equation, allowing us to compute unbiased random gradients through backpropagation with a single sample dynamic relative

advantage assessment technique. Minimizing $\tilde{\mathrm{L}}_{VIB}(\boldsymbol{\theta})$ helps achieve the optimal representation for each sensor, thereby meeting the goal of the information bottleneck theory, which is to maximize prediction accuracy while compressing features. In the generalized synergistic diagnosis strategy, the system learns the features of all sensors, regardless of whether a fault type has complete sensor data. For both missing and available data, the cross-fault type encoder receives supervision for the labels by minimizing the VIB loss $\tilde{\mathrm{L}}_{VIB}(\mathfrak{R})$. For complete data, the cross-fault type encoder receives additional supervision from this feature by minimizing the following mean squared error (MSE):

$$\mathrm{L}^{k}_{MSE}(\mathfrak{R}) = \frac{1}{L_k}\sum_{l=1}^{L} \|y^{k}_{l} - \hat{y}^{k}_{l}\|^{2}_{2} \tag{5}$$

where $\mathrm{L}^{k}_{MSE}(\mathfrak{R})$ denotes the loss of mean square error, $L_k = \sum C_{lk}$ represents k fault types. The overall loss function for training is:

$$\mathrm{L}(\mathfrak{R}) \coloneqq \tilde{\mathrm{L}}_{VIB}(\mathfrak{R}) + \sum_{k\in[K]} \eta^{k}\mathrm{L}^{k}_{MSE}(\mathfrak{R}) \tag{6}$$

where the coefficient η^{k} balances the supervisory loss for different fault types.

In summary, the supervisory methods for the encoder include two types: the adaptive encoder ($\mathfrak{R}^{k}_{senc}$, $k \in [K]$) and the interactive residual autoencoder ($\mathfrak{R}^{k}_{xenc}$, $k \in [K]$). When training for a fault type with missing sensors, parameters $\mathfrak{R}^{k}_{senc}$ and $\mathfrak{R}^{k}_{xenc}$ rely not only on the supervision of sensor features but also on label supervision. However, in the case of uneven missing rates, the parameters corresponding to sensors with higher missing rates experience less frequent supervision. This situation may lead to overfitting of some predictive labels, while others may be underfitted.

B. A DYNAMIC RELATIVE ADVANTAGE ASSESSMENT TECHNIQUE

In order to better quantify the strength of each fault type relative to other fault types, the concept of relative advantage is proposed in this paper. As shown in Fig. 3, the relative advantages of different fault types are defined and adjusted by analyzing the learning effect and interpolation loss of the fault types during the training process. In this way, balanced learning of different fault types can be achieved by adaptively adjusting the supervision intensity ϑ^{k} during the learning process of each sensor.

1. DEFINITION OF RELATIVE ADVANTAGE. A strong advantage of a fault type is characterized by a better learning effect or a smaller missing rate in the interpolation task, which results in a smaller interpolation loss during the training process. This phenomenon inspires us to define the relative advantage of fault types based on the interpolation loss. For subsequent analysis, the average interpolation loss for all fault types is first defined, denoted as:

$$\bar{\mathrm{L}}_{MSE}(\mathfrak{R}) = \sum_{k\in[K]} \mathrm{L}^{k}_{MSE}(\mathfrak{R})/K \tag{7}$$

For $\bar{\mathrm{L}}^{m}_{MSE}(\mathfrak{R})$, the failure type relative advantage is formally defined as:

$$RA^{k}(\mathfrak{R}) = [\bar{\mathrm{L}}_{MSE}(\mathfrak{R}) - \mathrm{L}^{k}_{MSE}(\mathfrak{R})]/\bar{\mathrm{L}}_{MSE}(\mathfrak{R}) \tag{8}$$

When $RA^{k}(\mathfrak{R}) > 0$, fault k is considered strongly dominant relative to other faults; when $RA^{k}(\mathfrak{R}) = 0$ the interpolation loss of fault type k is comparable to the average and is in equilibrium, and when $RA^{k}(\mathfrak{R}) < 0$, the interpolation loss of fault k is greater than the average loss, implying that the fault type performs weakly in the training and belongs to weakly dominant faults.

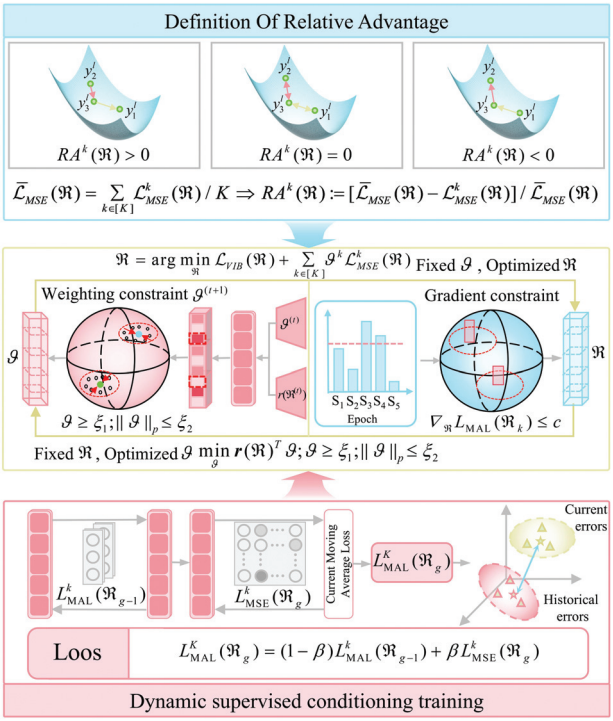

Fig. 3. A dynamic relative advantage assessment technique.

2. A TWO-LAYER COLLABORATIVE OPTIMIZATION ALGORITHM. The relative advantage vector for each fault type is set to $r(\mathfrak{R}) = [RA^{1}(\mathfrak{R}), RA^{2}(\mathfrak{R}), \cdots, RA^{K}(\mathfrak{R})]^{T}$, and the weight vector of all the supervised adjustments is set to $\vartheta = [\vartheta^{1}, \vartheta^{2}, \cdots, \vartheta^{K}]^{T}$. To ensure the effectiveness of learning, the supervised conditioning is inversely related to the relative advantage of each fault type. Therefore, the objective of supervised regulation is to minimize changes in the relative advantage during the training process. The synergistic training and supervised regulation is formulated as the following two-layer collaborative optimization problem:

$$\begin{aligned} &\min_{\vartheta,\mathfrak{R}} \boldsymbol{r}(\mathfrak{R})^{T}\vartheta; \vartheta \geq \xi_1; \|\vartheta\|_p = \xi_2 \\ &\mathfrak{R} = \arg\min_{\mathfrak{R}} \mathrm{L}_{VIB}(\mathfrak{R}) + \sum_{k\in[K]} \vartheta^{k}\mathrm{L}^{k}_{MSE}(\mathfrak{R}) \end{aligned} \tag{9}$$

where ξ_1 is a constant to ensure that the supervision weights are positive; ξ_2 denotes the ℓp-paradigm that constrains the supervision and ensures the stability of the training; and ϑ denotes the self-adaptive tuning of the supervision by minimizing the correlation between the supervision and the relative advantage.

The core of the above two-layer collaborative optimization algorithm is to optimize the weights ϑ in the outer layer and the model parameters $\mathfrak{R}$ in the inner layer to minimize the weighting error, the weights and the model parameters are interdependent and alternately iterated. Fix ϑ, optimize $\mathfrak{R}$, and update the model parameters to minimize the weighting error after weighting. Fix $\mathfrak{R}$, optimize ϑ, calculate the current relative advantage $r(\mathfrak{R})$, and adjust the weights ϑ to reflect the performance of the model. Inner

layer optimization: optimize the model parameters $\mathfrak{R}$, so that the weighted error is minimized:

$$\mathfrak{R} = \arg\min_{\mathfrak{R}} \mathrm{L}_{VIB}(\mathfrak{R}) + \sum_{k\in[K]} \vartheta^k \mathrm{L}^k_{MSE}(\mathfrak{R}) \quad (10)$$

where $\mathrm{L}_{VIB}(\mathfrak{R})$ denotes the underlying training error term and $\sum_{k\in[K]} \vartheta^k L^k_{MSE}(\mathfrak{R})$ denotes the weight-weighted model error term.

To ensure that each parameter update disrupts model training due to unusually large gradients, the process of introducing gradient constraints into the inner optimization is introduced, and each time the gradient is computed, the dynamic smoothing of the gradient is performed first, and then the gradient constraints are applied:

$$\nabla_{\mathfrak{R}} L_{\text{clipped}}(\mathfrak{R}_k) = \begin{cases} \nabla_{\mathfrak{R}} L_{\text{MAL}}(\mathfrak{R}_k) \text{ if } \nabla_{\mathfrak{R}} L_{\text{MAL}}(\mathfrak{R}_k) \le c \\ c \frac{\nabla_{\mathfrak{R}} L_{\text{MAL}}(\mathfrak{R}_k)}{\|\nabla_{\mathfrak{R}} L_{\text{MAL}}(\mathfrak{R}_k)\|} \text{ if } \nabla_{\mathfrak{R}} L_{\text{MAL}}(\mathfrak{R}_k) > c \end{cases} \quad (11)$$

where $\nabla_{\mathfrak{R}} L_{\text{clipped}}(\mathfrak{R}_k)$ is the gradient after applying the gradient constraint. Outer layer optimization: Optimize the weight ϑ, so that the weight distribution is reasonable:

$$\min_{\vartheta} \boldsymbol{r}(\mathfrak{R})^T \vartheta; \vartheta \ge \xi_1; \|\vartheta\|_p \le \xi_2 \quad (12)$$

The weights are updated according to the following rules:

$$\vartheta^{(t+1)} = \mathrm{Proj}_{[\xi_1,\xi_2]}\left(\vartheta^{(t)} - \alpha r(\mathfrak{R}^{(t)})\right) \quad (13)$$

where $\mathrm{Proj}_{[\xi_1,\xi_2]}$ denotes the projection of the updated weights into the constraint range, and α denotes the learning rate, which controls the update step size.

3. DYNAMIC SUPERVISED CONDITIONING TRAINING. Since the training process is affected by stochastic gradients, variations in the error may lead to unstable model performance. Therefore, a dynamic supervision technique is proposed. The key to dynamic supervision is the use of moving average loss to smooth out the error variations:

$$L^K_{\text{MAL}}(\mathfrak{R}_g) = (1-\beta)L^k_{\text{MAL}}(\mathfrak{R}_{g-1}) + \beta L^k_{\text{MSE}}(\mathfrak{R}_g) \quad (14)$$

where $\beta \in [0, 1]$ denotes the smoothing coefficient, which controls the weight of the historical error and the current error. The larger β is, the higher the weight of the current error is. The smaller β is, the more emphasis is placed on the historical error.

IV. EXPERIMENTS AND RESULTS

A. DATA DESCRIPTION

In order to validate the effectiveness of the proposed dynamic relative advantage-driven synergistic multi-fault diagnosis method under the condition of imbalanced data missing rate, multi-channel signals under various motor health states were acquired from the motor experimental setup shown in Fig. 4. The signal data consist of nine channels with a sampling frequency of 25.6 kHz. The channel signals are mainly from the three-axis acceleration signals at the drive end and the fan end, as well as the current signals. As shown in Fig. 5, eight different motor health states were simulated. Each sample in the dataset was constructed using a sliding window of length 1024 with non-overlapping windows. Table I lists the 12 operating conditions in the experiment. At each motor speed, 1200 samples were collected, of which 80% were used for training and 20% for testing.

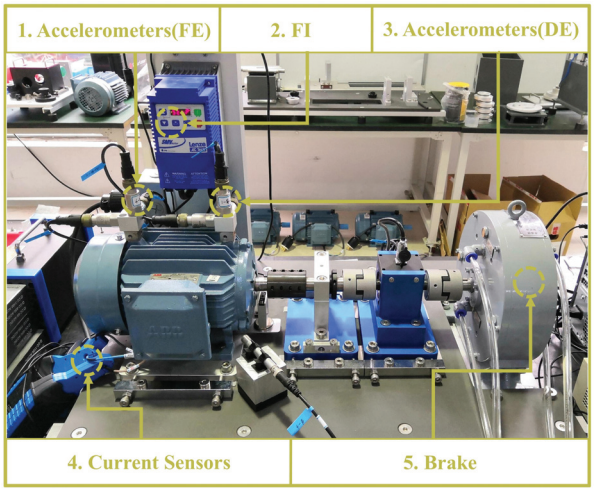

Fig. 4. Motor failure simulation experiment device.

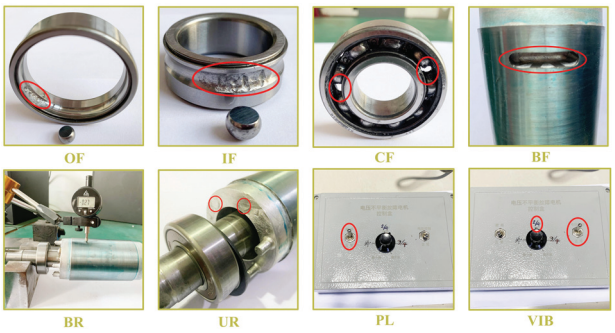

Fig. 5. Failure simulation photos.

Table I. Working conditions of motors

Serial No.	Load/Rotating Speed	Serial No.	Load/Rotating Speed
1	0N·m/15Hz	7	0N·m/35Hz
2	10N·m/15Hz	8	10N·m/35Hz
3	15N·m/15Hz	9	20N·m/35Hz
4	0N·m/25Hz	10	0N·m/45Hz
5	10N·m/25Hz	11	10N·m/45Hz
6	20N·m/25Hz	12	20N·m/45Hz

B. DATA PRE-PROCESSING

In order to simulate the missing data and unbalanced missing rates in real-world environments, the number of missing values for each feature in the tensor was randomly quantified, mainly for the outer gear ring faults, inner gear ring faults, and rolling body faults. And five types of randomized missing rates were set while keeping the unbalanced missing rates for three fault types to reflect the different levels of data availability in real scenarios. Table II shows the missing rates.

Table II. Random missing rate and symbol

Description of symbols	Defect rate
(I_1, O_1, C_1)	(0.2, 0.5, 0.4)
(I_2, O_2, C_2)	(0.4, 0.6, 0.3)
(I_3, O_3, C_3)	(0.6, 0.3, 0.2)
(I_4, O_4, C_4)	(0.7, 0.4, 0.6)
(I_5, O_5, C_5)	(0.5, 0.95, 0.6)
(I_6, O_6, C_6)	(0.5, 0.9, 0.95)
(I_7, O_7, C_7)	(0.9, 0.9, 0.95)

Table III. Related parameter settings

Parameter symbol	Numeric size
λ	0.5
P	3
ξ_1	0.9
ξ_2	0.3
α	1e-3
β	5e-3

C. REALIZATION DETAILS

Using Transformer as an adaptive encoder for all fault types, interactive residual autoencoder as a cross-fault type encoder, and multilayer perceptron as a decoder. The number of self-attentive layers for all adaptive encoders is 4, and all layers have 16 heads and 256-dimensional hidden embeddings. All cross-fault type encoders consisted of 4 interactive residual autoencoder blocks with RA layer sizes of 256-128-64-32-64-128-256, and the outputs were 128-dimensional vectors as representations of the corresponding fault types, and the model was trained using the optimizer Adam with a learning rate of 1×10^{-4} and a batch size of 256. The specific network structure is shown in Table IV. The parameter settings are shown in Table III, and in the experimental validation, a weighted F1 score was used as the performance metric, and the final result was derived from the average of 3 or more replicated experiments.

D. RESULTS AND EFFICIENCY OF EXPERIMENTS

1. EXPERIMENTAL VALIDATION AT HIGH DELETION RATES. To simulate extreme data scarcity scenarios, we conducted systematic tests on synthetic datasets with three different sensors with up to 95% missing rate under the same fault type and designed three differentiated scenarios: scenario 5 with missing rate of (0.5, 0.95, 0.6) (70.3% on average), scenario 6 with missing rate of (0.5, 0.9, 0.95) (80% on average), and scenario 7 with missing rate of (0.9, 0.9, 0.95) (mean 91.7%).

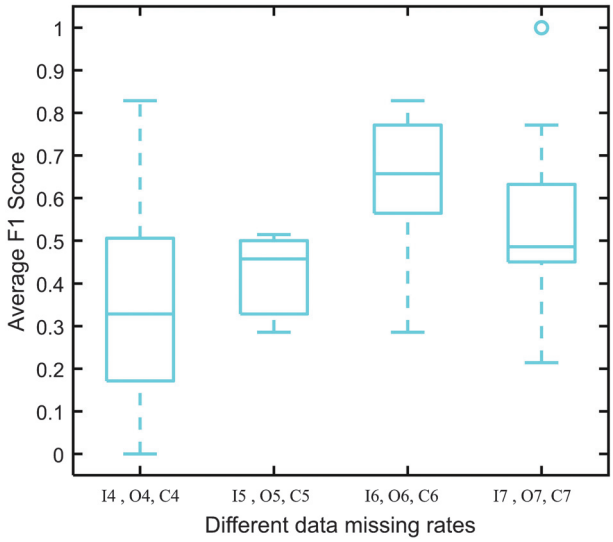

Fig. 6. Mean F1 index under a sequential increase in mean deletion rate.

The experimental results show that the method performs well in the baseline condition (40%–60% missing rate) with an accuracy of 95.2% and an F1 value of 0.948; as the average missing rate increases, the accuracy drops to 88.7% and an F1 value of 0.872 in Scenario 1 (70.3% missing rate), and the accuracy plummets to 62.1% and an F1 value of 0.589 in Scenario 3 (91.7% missing rate). When the average missing rate exceeds 85%, the model accuracy falls below 70%, indicating that the method relying on the generalized joint diagnosis strategy across fault types fails due to data sparsity. The underlying reason is that when the valid sensor data are less than 15%, the correlation features across fault types are too sparse to generate meaningful representations, resulting in the diagnostic performance degrading to a level similar to that of random prediction. Figure 6 visualizes the trend of the average F1 index of the proposed method for four scenarios with increasing average missing rate.

2. TRAINING TIME. For the real-time requirements of industrial applications, the training time comparison between the proposed method and the baseline model is shown in Table V. The experiments were conducted in the same hardware environment (NVIDIA RTX 4060 GPU, 32GB video memory) and the results are shown below:

The training time of the proposed method is longer than that of the traditional method C, but it is 35.6% and 17.0%

Table IV. Network configuration

Layer	Parameters
Input (A/L/V)	Size: 130 (A), 1024 (L), 384 (V)
Transformer Encoder	Layers: 3, Heads: 8, Embd Size: 128, FFN Hidden: 512 (128×4), Activation: GELU
Embedding Fusion	Size: 384 (128×3, A+L+V concatenation)
Residual AE Blocks	Layers: 128→64→32 (3 blocks), BN: - (use_bn=False), Dropout: 0
Residual XE Cross-modal	Input: 256 (A+L/V+A/L+V), Output: 128 (target modality), Layers: 128→64→32 (3 blocks)
Classifier Fc Classifier	Layers: 128→128, Dropout: 0.3, BN: Optional (opt.bn)
Optimizer	Adam (lr=opt.lr, β1=0.9, β2=0.999), LR Scheduler: Cosine Annealing + Warmup

Table V. Comparison of training time of the proposed method with the baseline model

Method	Training time	Instruction
Proposed method	18.5 ± 1.2	Contains VE/CFE dual coding with shared label constraints
Method D	28.7 ± 2.5	Contains adversarial training loop, slower convergence
Method E	22.3 ± 1.8	Requires sample pair similarity computation, higher complexity
Method F	5.6 ± 0.5	Contains only single feature extraction, no deep network

shorter compared to GAN and diffusion models, respectively. This is due to the lightweight cross-fault encoder design that avoids complex adversarial games or global feature comparisons. In batch training scenarios, the training time of the proposed method is in the acceptable range (about 20 seconds per epoch), which is suitable for the daily model iteration requirements of most industrial production lines.

E. ABLATION EXPERIMENT

1. IMPACT OF DIFFERENT MODULES ON THE PROPOSED METHODOLOGY. In the ablation experiment, by comparing the following three scenarios: 1) the absence of cross-fault type generalized synergistic diagnosis strategy (Method A); 2) the absence of the dynamic relative advantage assessment technique (Method B); and 3) the F1 score of the method proposed in this experiment under conditions of missing data with imbalanced missing rates. The results are shown in Fig. 7(a). The method proposed outperforms the other two methods in diagnostic classification accuracy under different imbalanced missing rates for outer race fault, inner race fault, and rolling element fault. Fig. 7(b) and (c) show the fault diagnosis classification results for missing rates of (0.4, 0.6, 0.3) and (0.7, 0.4, 0.6), respectively. Figure 7(e) further illustrates the learning curve accuracy and F1 score for the three methods with a missing rate of (0.4, 0.6, 0.3), providing further evidence of superiority.

Additionally, Fig. 7(f) and (g) present the relative advantages of the three methods during training, with missing rates of (0.4, 0.6, 0.3) for the outer race fault, inner race fault, and rolling element fault. Figure 7(f) indicates that when the relative advantage awareness supervision technique is not implemented, the sensor data corresponding to the rolling element fault predominates over the other two fault types. A comparison between Fig. 7(f) and (g) clearly shows that the supervision technique facilitates more balanced learning, thereby reducing the inductive loss gap between fault types and enhancing the model's generalization ability.

2. IMPACT OF HYPERPARAMETERS ON MODEL PERFORMANCE. The sensitivity analysis of systematic parameters shows that the combination ($\lambda = 0.5$, $\xi_1 = 0.9$, $\xi_2 = 0.3$, $\alpha =$ 1e-3, $\beta =$ 5e-3) performs optimally in terms of accuracy, F1 score, convergence speed and noise robustness, which verifies the reasonableness of the theoretical derivation; among them, the accuracy rate decreases significantly when λ deviates from 0.5, which embodies the key role of information bottleneck trade-off, $\xi_1<0.9$ or $\xi_2>0.3$ will lead to a significant reduction in noise robustness, and α,β exceeding the recommended range will increase the noise robustness. 0.9 or $\xi_2>0.3$ will lead to a significant decrease in noise robustness, highlighting the filtering ability of the threshold on industrial noise, and α/β exceeding the recommended range will increase the number of convergence rounds affecting the real-time efficiency; in summary, the optimal intervals for the hyperparameters are clarified as $\lambda \in$ [0.4, 0.6], $\xi_1 \in$ [0.85, 0.95], $\xi_2 \in$ [0.25, 0.35], $\alpha \in$ [5e-4, 1.5 e-3], $\beta \in$ [3e-3, 7e-3]. The detailed parameter correlations are shown in Fig. 8.

F. COMPARISON EXPERIMENT

To validate the advantages of the proposed method for fault diagnosis of motor under conditions of data missingness and imbalanced missing rates, this paper compares the proposed method with three existing approaches.

1) Method C [15] proposed a new method to compensate the data integrity problem and minimize the prediction error by integrating complete data with uncertain information through a dynamic Bayesian network.
2) Method D [16] proposed a novel fuzzy clustering framework that combines nonnegative latent factor analysis with feature-weighted fuzzy double C-mean clustering, achieving high accuracy for incomplete data clustering.
3) Method E [17] proposed a novel data augmentation method based on a deep diffusion probability model to address fault diagnosis challenges under unbalanced data distributions with varying missing rates.
4) Method F [18] proposed an improved deep convolutional generative adversarial network with discriminator gradient gap regularization for specific application scenarios.
5) Method G [19] proposed a strategy based on a feature fusion deep convolutional generative adversarial network architecture.

The comparison results, shown in Fig. 9, can be summarized as follows: When the missing data rate is low (i.e., 0.0 and 0.1), the proposed method outperforms all comparison models because other models have not been fully exposed to data missing scenarios and struggle to handle such situations. In contrast, our method continues to perform cross-fault-type representation learning effectively, even when no data is missing. When the missing data rate is moderate (e.g., between 0.2 and 0.6), despite using a less data-intensive supervised training technical, the proposed method's performance is at least on par with other models, demonstrating its advantage in missing data supervision. When the missing data rate is high (e.g., between 0.7 and 0.9), our method maintains relatively stable performance with only slight to moderate degradation, whereas comparison models exhibit significant performance declines. These results indicate that our method effectively leverages the supervisory information from available fault types, with label information playing a key role in improving performance. Overall, the method demonstrates

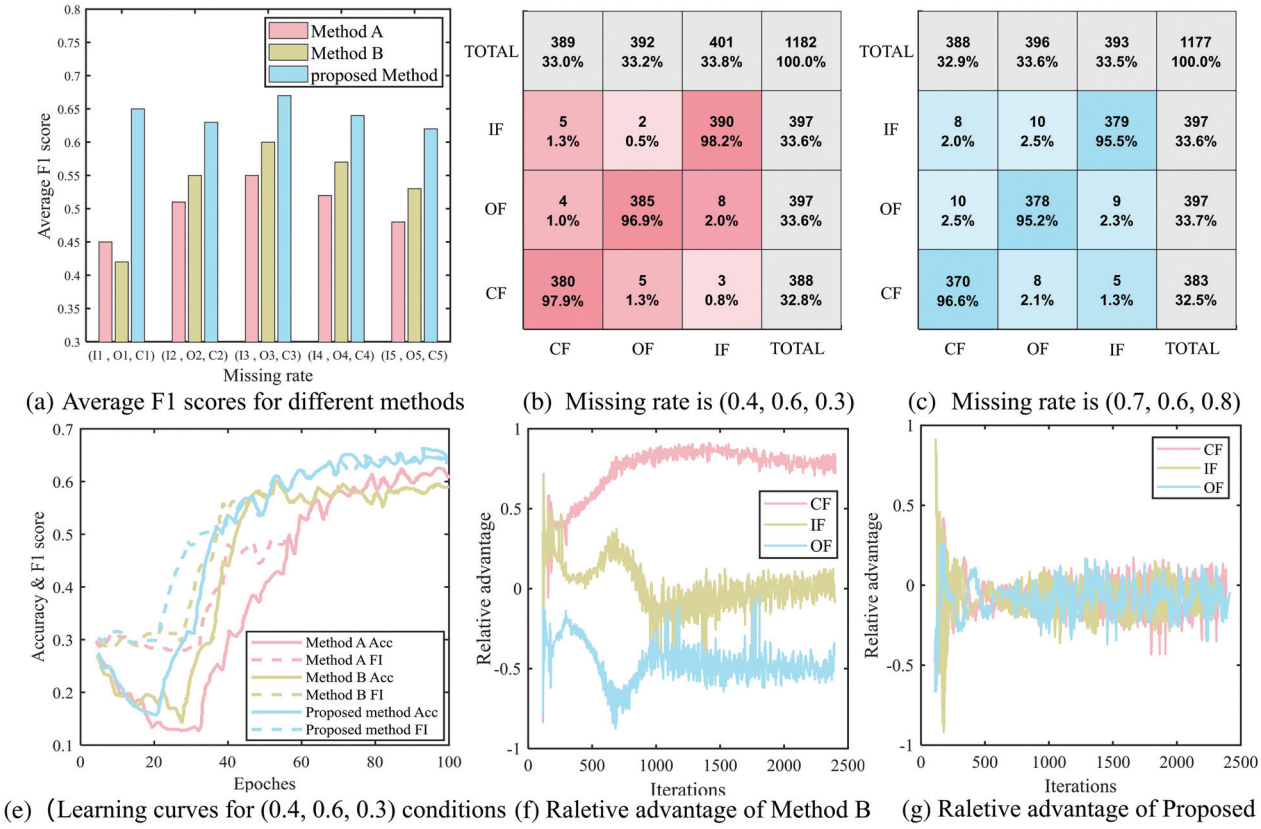

Fig. 7. Comparison of ablation experiment results.

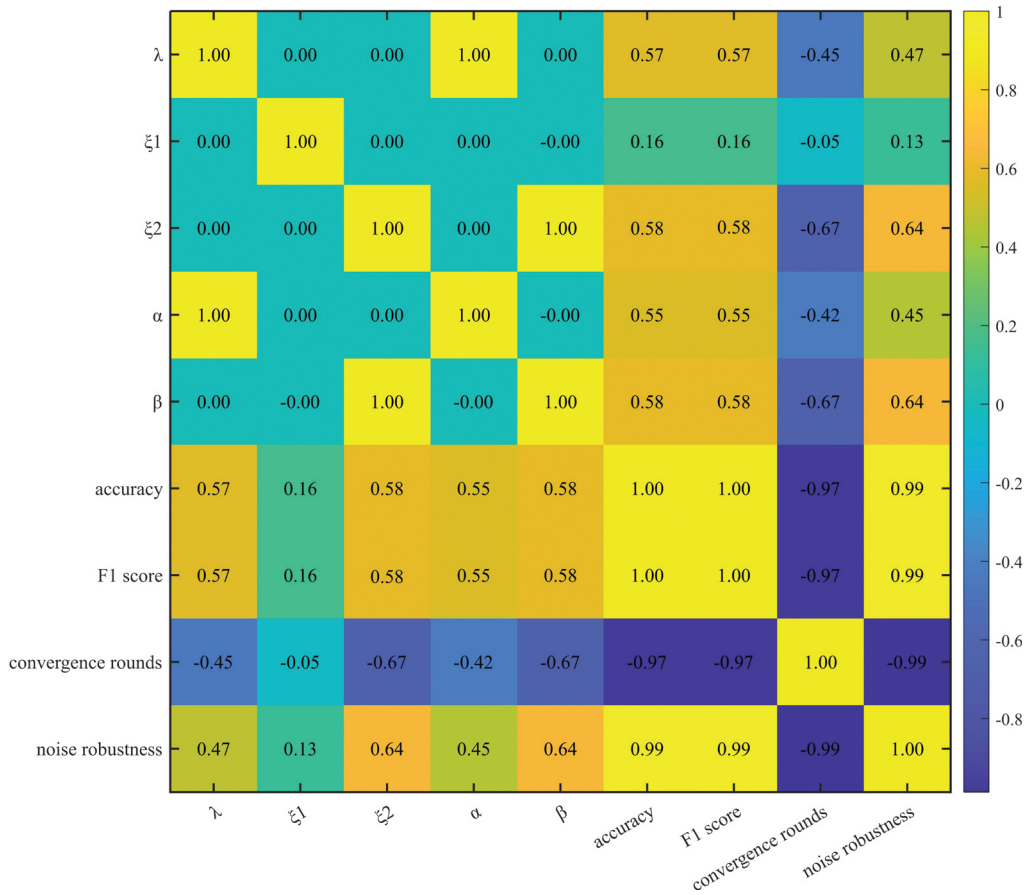

Fig. 8. Parameter correlation analysis.

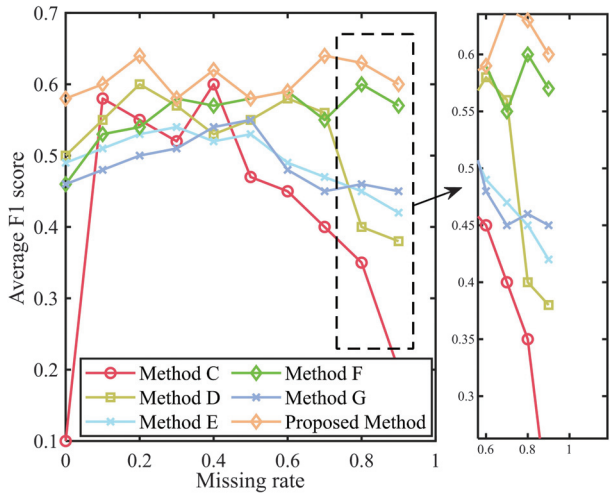

Fig. 9. Comparison results of the proposed method and the other existing methods.

excellent robustness under varying missing data rates, proving its effectiveness and broad applicability.

V. CONCLUSIONS

In this paper, a dynamic relative advantage-driven synergistic diagnosis method is proposed for motors under imbalanced missing data rates. First, a generalized synergistic diagnosis strategy across fault types based on the variational information bottleneck theory is constructed, which is supervised by the optimal representation of each fault type and the available data and labels in order to efficiently utilize incomplete samples of the missing data and perform effective supervision. Additionally, a dynamic relative advantage assessment technique is designed to reduce diagnostic accuracy degradation caused by imbalanced missing data rates. This is achieved by quantifying the relative advantage of each fault type through reconstruction loss and employing a two-layer collaborative optimization algorithm with adaptive supervision weights to adjust the training process. The effectiveness of the method is verified through motor failure simulation experiments, and the robustness of the proposed method is demonstrated in the absence of data. Future research will focus on exploring the convergence of the algorithm and further improving the network performance.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

REFERENCES

[1] L. A. Sikinyi, C. M. Maina, and L. Ngoo, "Condition monitoring and fault diagnosis of induction motor faults using deep learning: a review," *2023 IEEE AFRICON*, pp. 1–6, 2023, doi: 10.1109/AFRICON55910.2023.10293578.

[2] S. K. Gundewar and P. V. Kane, "Condition monitoring and fault diagnosis of induction motor," *J. Vib. Eng. Technol.*, vol. 9, pp. 643–674, 2021.

[3] S. Gawde et al., "Multi-fault diagnosis of industrial rotating machines using data-driven approach: a review of two decades of research," *Eng. Appl. Artif. Intell.*, vol. 123, p. 106139, 2023.

[4] C. Wang et al., "Richly connected spatial–temporal graph neural network for rotating machinery fault diagnosis with multi-sensor information fusion," *Mech. Syst. Signal Process.*, vol. 225, p. 112230, 2025.

[5] X. Xu et al., "Intelligent monitoring and diagnostics using a novel integrated model based on deep learning and multi-sensor feature fusion," *Measurement*, vol. 165, p. 108086, 2020.

[6] Y. Zhang et al., "Multi-modal data cross-domain fusion network for gearbox fault diagnosis under variable operating conditions," *Eng. Appl. Artif. Intell.*, vol. 133, p. 108236, 2024.

[7] Q. Wei et al., "WSAFormer-DFFN: a model for rotating machinery fault diagnosis using 1D window-based multi-head self-attention and deep feature fusion network," *Eng. Appl. Artif. Intell.*, vol. 124, p. 106633, 2023.

[8] J. Li, Y. Liu, and Q. Li, "Intelligent fault diagnosis of rolling bearings under imbalanced data conditions using attention-based deep learning method," *Measurement*, vol. 189, p. 110500, 2022.

[9] C. Li et al., "Noise-robust multi-view graph neural network for fault diagnosis of rotating machinery," *Mech. Syst. Signal Process.*, vol. 224, p. 112025, 2025.

[10] D. Li et al., "A two-stage feature-extraction-based fault detection method for industrial processes with imperfect data," *IEEE Trans. Instrum. Meas.*, vol. 74, pp. 1–11, 2025.

[11] D. Liu, D. Liang, and C. Wang, "A novel three-way decision model based on incomplete information system," *Knowl-Based Syst.*, vol. 91, pp. 32–45, 2016.

[12] Y. Ma et al., "Regionwise generative adversarial image inpainting for large missing areas," *IEEE Trans. Cybern.*, vol. 53, no. 8, pp. 5226–5239, 2022.

[13] C. Ou et al., "Missing-data imputation with position-encoding denoising autoencoders for industrial processes," *IEEE Trans. Instrum. Meas.*, vol. 73, pp. 1–11, 2024.

[14] Y. Yu et al., "A novel bearing fault data generation strategy combining physical modeling and CycleGAN variant for fault diagnosis without real samples," *IEEE Trans. Instrum. Meas.*, vol. 74, pp. 1–17, 2025.

[15] X. Shao et al., "Artificial intelligence enhanced fault prediction with industrial incomplete information," *Mech. Syst. Signal Process.*, vol. 224, p. 112063, 2025.

[16] Y. Song et al., "Nonnegative latent factor analysis-incorporated and feature-weighted fuzzy double c-means clustering for incomplete data," *IEEE Trans. Fuzzy Syst.*, vol. 30, no. 10, pp. 4165–4176, 2022.

[17] X. Yang et al., "A novel data augmentation method based on denoising diffusion probabilistic model for fault diagnosis under imbalanced data," *IEEE Trans. Ind. Inform.*, vol. 20, no. 5, pp. 7820–7831, 2024.

[18] X. Song, Z. Liu, and Z. Wang, "Rolling bearing fault diagnosis in electric motors based on IDIG-GAN under small sample conditions," *Meas. Sci. Technol.*, vol. 35, no. 10, p. 106105, 2024.

[19] L. Ye, K. Zhang, and B. Jiang, "Synergistic feature fusion with deep convolutional GAN for fault diagnosis in imbalanced rotating machinery," *IEEE Trans. Ind. Inform.*, vol. 21, no. 2, pp. 1901–1910, 2025.