IF4FD: Multiscale Information Fusion for Zero-Shot Industrial Fault Diagnosis

Chenwei Tang, *Member, IEEE*, Ying Wang, Weijia Wang, Wangyang Ying, Jie Liu, Yong Wang, Nanxu Gong, Wei Ju, *Member, IEEE*, Rong Xiao, and Jiancheng Lv, *Senior Member, IEEE*

***Abstract*—Fault diagnosis aims to identify faults occurring in industrial production processes to prevent personnel injuries and economic losses. However, there are two main challenges in solving fault diagnosis, i.e., *extracting discriminative features from limited sensor data* and *recognizing new classes of faults. To fill these research gaps*, we propose a zero-shot fault diagnosis framework, called IF4FD, based on multiscale information fusion. First, we enhanced raw data from the perspectives of category knowledge, attribute knowledge, and feature knowledge. Then, by drawing on zero-shot learning (ZSL), we can transfer knowledge of trained faults to new classes of faults, enabling the classification of previously unknown faults. The multiscale informative knowledge effectively facilitates knowledge transfer and fault classification, thereby enhancing the accuracy of zero-shot fault diagnosis. Extensive experiments on two industrial fault diagnosis datasets validate the effectiveness of the proposed method, which consistently achieves superior performance compared to representative zero-shot fault diagnosis methods, general ZSL baselines, and several supervised classifiers. A case study on real industrial data from the Cranfield Multiphase Flow Facility also confirms the method's effectiveness in practical applications.**

***Index Terms*—Fault diagnosis, multiscale information, transfer learning, zero-shot learning.**

I. Introduction

FAULTS in industrial production processes are inevitable. They are often caused by factors such as human error, machine aging, and may result in severe consequences, including economic losses and personnel injuries. Thus, intelligent diagnosis of faults plays a crucial role in ensuring the safety and improving efficiency of industrial production processes [1]. There are two main challenges in solving fault diagnosis: 1) identifying new types of faults that have not been encountered before in industrial production processes [2]; 2) extracting informative knowledge from limited sensor data to differentiate between diverse fault types [3]. First, industrial production involves numerous machine components, each with multiple potential fault conditions. However, due to the potentially severe consequences of industrial faults, collecting data on them is challenging. Traditional machine learning heavily relies on the quality and quantity of data. In situations where data are lacking for model training, accurately identifying novel faults becomes a critical challenge in fault diagnosis [4]. Identifying novel faults aims to answer: *How to identify all potential fault conditions with limited data?* Second, data in industrial settings are primarily collected from sensors, yet deploying a large number of sensors to gather diverse machine data proves to be costly [5]. The efficacy of machine learning is profoundly contingent on data. Hence, extracting more informative knowledge from limited data dimensions is important for enhancing machine learning model performance [6]. Extracting informative knowledge aims to answer: *How to distinguish diverse faults with limited sensor data?* Supervised learning methods, widely used in traditional classification tasks, have been applied in this field and partially tackle these two challenges [7]. Due to insufficient data, they struggle to handle previously unseen fault classes. To recognize both seen and unseen faults, the method, called fault description attribute transfer (FDAT), is first proposed to leverage zero-shot learning (ZSL) [8] to solve the fault diagnosis task [9]. As shown in Fig. 1, the zero-shot fault diagnosis (ZSFD) methods extract attributes from the descriptions of fault samples, and then construct an attribute space shared by all classes to achieve ZSFD by transferring attribute knowledge from trained seen classes to

Received 26 November 2025; accepted 12 December 2025. Date of publication 3 February 2026; date of current version 6 April 2026. This work was supported in part by the National Major Scientific Instruments and Equipments Development Project of National Natural Science Foundation of China under Grant 62427820, in part by the Fundamental Research Funds for the Central Universities under Grant 1082204112364, in part by the Science Fund for Creative Research Groups of Sichuan Province Natural Science Foundation under Grant 2024NSFTD0035, in part by the Natural Science Foundation of Sichuan under Grant 2024NSFSC1461 and Grant 2024NSFSC1470, in part by the Tianfu Yongxing Laboratory Organized Research Project Funding under Grant 2023CXXM14, and in part by the Sichuan Province Innovative Talent Funding Project for Postdoctoral Fellows under Grant BX202213 . Paper no. TII-25-8412. *(Corresponding author: Nanxu Gong.)*

Chenwei Tang, Ying Wang, Wei Ju, Rong Xiao, and Jiancheng Lv are with the College of Computer Science, Sichuan University, and Engineering Research Center of Machine Learning and Industry Intelligence, Ministry of Education, Chengdu 610065, China (e-mail: tangchenwei@scu.edu.cn; wangying2yvonne@stu.scu.edu.cn; juwei@scu.edu.cn; rxiao@scu.edu.cn; lvjiancheng@scu.edu.cn).

Weijia Wang is with the Pittsburgh Institute, Sichuan University, Chengdu 610225, China (e-mail: 2022141520173@stu.scu.edu.cn).

Wangyang Ying and Nanxu Gong are with the Engineering Research Center of Machine Learning and Industry Intelligence, Arizona State University, Tempe, AZ 85281 USA (e-mail: wangyang.ying@asu.edu; nanxugong@asu.edu).

Jie Liu is with the Science and Technology on Reactor Fuel and Materials Laboratory, Nuclear Power Institute of China, Chengdu 610213, China (e-mail: liujie5@stu.scu.edu.cn).

Yong Wang is with New Power Systems Research Center, Tianfu Yongxing Laboratory, Chengdu 610213, China (e-mail: wangyong@alu.scu.edu.cn).

Digital Object Identifier 10.1109/TII.2025.3645726

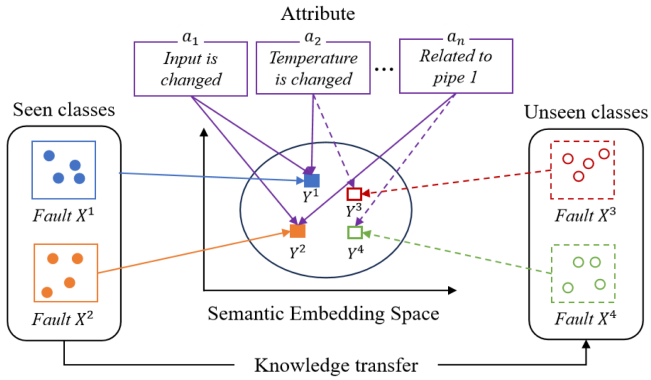

Fig. 1. Zero-shot fault diagnosis uses a shared attribute space to transfer knowledge, enabling the recognition of unseen faults.

unseen test classes. Furthermore, the issue of semantic consistency between the feature space and attribute space is overlooked in FDAT, prompting the proposal of the SCE method [3] to establish a semantically consistent embedding space. By designing a special Barlow matrix, SCE promotes consistency between fault and attribute, and achieves competitive results. However, SCE still fails to address the challenge of excessive fault categories, making precise classification difficult. Therefore, we need a novel perspective to effectively address fault diagnosis.

In this article, we propose **IF4FD**, a novel ZSFD approach based on multiscale information fusion. On the one hand, through ZSL, the **IF4FD** method can leverage existing fault information for transfer learning, enabling the identification of unknown faults. To further enhance the structured exploitation of fault attribute knowledge, this study introduces a novel multimodal semantic encoding framework that seamlessly integrates textual descriptions of faults. On the other hand, we employ data augmentation from three dimensions, i.e., category knowledge, attribute knowledge, and feature knowledge, applied to the original data. By incorporating multiscale information, the **IF4FD** method can better articulate the knowledge within high-dimensional data, thus facilitating transfer learning and improving classification performance. Specifically, the multiscale information consists of the following three components.

1) *Category knowledge information:* We establish a center for each class within the embedding space, dynamically updating these centers within each batch to foster a more compact distribution of data belonging to the same class.
2) *Attribute knowledge information:* We utilize the mutual information between features and attributes in a latent space with the same dimensionality as the attributes to guide the learning of semantic embeddings.
3) *Feature knowledge information:* We also use reconstructed data to enrich the original dataset, providing more detailed features. We simply combine the three types of knowledge (category, attribute, and feature) into one representation. This approach preserves each one's strengths while allowing them to work together, improving both diagnosis performance and model interpretability.

Once information is aggregated at multiple scales, we train a binary classifier for each attribute. The ZSFD task is construed as a constellation of binary classification tasks. Provided we can anticipate the values of each attribute for unseen fault classes, identification of such unseen classes can be facilitated through the nearest neighbor search within the attribute space. The main contributions can be summarized as follows.

1) Our proposed **IF4FD** leverages knowledge-enriched multiscale information extraction to capture highly discriminative representations of sensor data across diverse perspectives, enabling effective diagnosis of unseen fault classes in the ZSFD task.
2) To enhance class separability, we introduce and dynamically update feature centers for each known class, minimizing intraclass variance. Moreover, we employ noise contrastive estimation (NCE) to enforce precise alignment between the learned feature representations and the semantic space, guiding the model toward more accurate and generalizable semantic embeddings.
3) Extensive experiments on two public industrial fault diagnosis datasets demonstrate that **IF4FD** achieves state-of-the-art performance, significantly outperforming existing zero-shot baselines. Furthermore, an in-depth case study on a complex real-world industrial dataset validates **IF4FD**'s robustness and effectiveness in handling dynamic, real-world scenarios.

II. Related Work

ZSL aims to recognize instances belonging to unseen categories that lack labeled data for training [10]. Existing ZSL methods can be divided into two main streams, i.e., embedding-based method and generative method [11]. Embedding-based methods aim to uncover the correlations between features and auxiliary semantic information [12], [13], [14]. Common embedding-based approaches involve mapping features into a semantic space or vice versa. Alternatively, both features and semantics can be embedded into a shared space [15]. Learning a projection function within embedding-based methods facilitates the acquisition of semantic embeddings for unseen classes, subsequently employing the nearest neighbor search to identify matching classes. However, embedding-based methods are susceptible to biases and domain shift issues, where the projection function may tend to predict attribute distributions for seen classes, potentially impeding accurate mapping for unseen classes. With the introduction of generative adversarial nets (GANs) and variational autoencoders (VAEs), the generative methods have presented novel solutions for ZSL tasks [16]. These generative methods can learn mappings from attributes to samples of seen classes, thereby enabling the generation of samples for unseen classes based on their attributes. Consequently, ZSL tasks can be construed as conventional classification tasks [17]. While generative methods offer certain advantages over embedding-based approaches, their training procedures and time requirements are more intricate. Building upon embedding-based methodologies, our approach aims to address the scarcity of raw sensor data in fault diagnosis tasks. To address this, we propose a ZSFD method that leverages multiscale information. This approach aims to improve the

effectiveness of fault diagnosis even when confronted with limited data availability.

Due to its success in addressing image classification tasks with limited labeled samples, ZSL can be extended to various industrial applications to tackle practical challenges [3], [9], [18]. By leveraging the capabilities of ZSL, the limitations of traditional supervised learning approaches can be overcome, where obtaining labeled data for every possible category may be impractical or infeasible [18]. Fault diagnosis, the focus of this article, stands out as a critical industrial application where ZSL can be applied. In order to tackle the challenges in ZSFD, the FDAT method was proposed in 2021, which is the first to leverage ZSL to transfer knowledge from seen fault categories to accurately classify unseen fault categories [9]. FDAT constructs an attribute matrix for knowledge transfer based on fault descriptions. Using extracted features, FDAT trains attribute classifiers, inferring the fault class by obtaining the attribute vector. Building upon FDAT, Hu et al. observed the alignment between features and attributes [3]. They propose the SCE method and introduce a specialized Barlow matrix to enhance semantic consistency, employing two sets of aligned encoders and decoders to guide the network in extracting more reconstruction-friendly features. Besides, in [18], a generative method based on GAN is proposed to solve the problem of zero-shot and few-shot fault diagnosis. Following these insights, zero-shot and few-shot learning approaches have emerged as promising research trends for addressing the challenges in practical industrial fault diagnosis applications [19]. In contrast, our work aims to improve the accuracy-efficiency tradeoff in ZSFD by introducing a lightweight multiscale semantic fusion strategy that facilitates effective attribute knowledge transfer with reduced computational cost.

III. Methodology

Here, we first introduce some notations and the problem definition of ZSFD. During the training process, the seen class samples, denoted as $S = \{(x, y)|x \in X^S, y \in Y^S\}$, are available, where X^S and Y^S represent the dataset and label set of the seen classes, respectively. Then, let $U = \{(x, y)|x \in X^U, y \in Y^U\}$ denotes the unseen class samples, where X^U is the dataset of unseen classes and Y^U is the label set of unseen classes. We define N_S and N_U as the number of categories for seen and unseen classes, respectively. Note that in the definition of ZSL, S and U are disjoint. The objective is to utilize the knowledge learned from seen classes to enable the recognition of samples belonging to unseen classes. To facilitate knowledge transfer from the seen classes to the unseen classes, we introduce the attribute matrix $A = \{a|a \in A^{S+U}\}$, where $a = \{a_1, \ldots, a_n\}$ represents an n-dimensional attribute vector, and A^{S+U} is the attribute set encompassing seen and unseen classes. The attribute matrix A can be used to construct a shared semantic space for both seen and unseen classes. Thus, we can establish a relationship between data and attributes of seen classes by learning the mapping $f(x) : x \to a$. When discerning the class of an unseen fault sample, we initially predict the attribute vector by aggregating each mapped attribute from $a = f(x)$. Thereafter, through the nearest neighbor search, we identify the fault class with attributes most similar to the sample. Finally, we realize the classification of fault samples of unknown class.

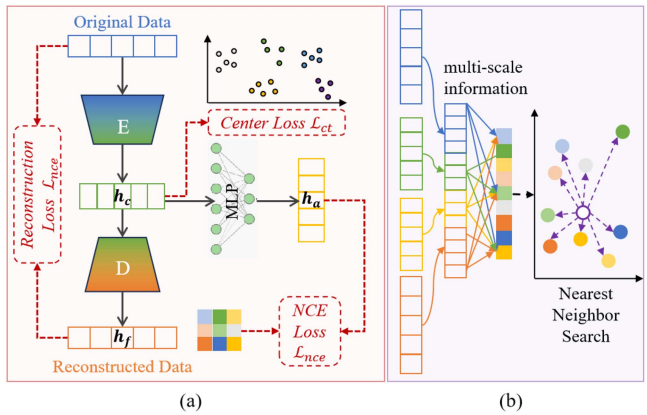

Fig. 2. Framework of **IF4FD**. Based on the original data, we enrich data representation from three perspectives: category knowledge h_c, attribute knowledge h_a, and feature knowledge h_f. Leveraging multi-scale information, we train binary classifiers for attribute prediction. (a) extraction of multi-scale information. (b) binary classification.

As shown in Fig. 2, the proposed **IF4FD** consists of two steps. The first step is the extraction of multiscale information, which focuses on extracting multiscale information to effectively represent data from multiple perspectives. By doing so, we gain valuable insights from low-dimensional data, which greatly contributes to our classification. Specifically, we first map the original data into an embedding space via an encoder E and guide the generation of discriminative features through the central loss $\mathcal{L}_{ct}$, thereby facilitating the extraction of category knowledge information h_c. Then, we employ a multilayer perceptron (MLP) to embed the category knowledge information h_c into a space with the same dimensionality as the attribute through the NCE [15] loss $\mathcal{L}_{nce}$, guiding the representation of attribute knowledge information h_a in the latent space. Finally, to enrich the representation of the feature knowledge information h_f, we reconstruct the category knowledge information h_c via a decoder D through the reconstructed loss $\mathcal{L}_{rc}$. To integrate multiscale semantic information, we fuse category h_c, attribute h_a, and feature h_f knowledge by concatenating them along the channel dimension. This approach preserves the unique characteristics of each knowledge type while promoting beneficial interactions between them. The resulting comprehensive representation is then processed by multiple binary classifiers, and the final prediction is made via nearest-neighbor search.

A. Multi-Modal Semantic Encoding for Data Processing

Attribute knowledge representation is fundamental to ZSFD, as it encapsulates the essential information needed to distinguish and identify various fault types. Therefore, the primary objective of data processing is to extract and construct attribute knowledge from diverse datasets. To address the unique characteristics of different datasets, we employ customized processing strategies that effectively capture and represent their attribute information.

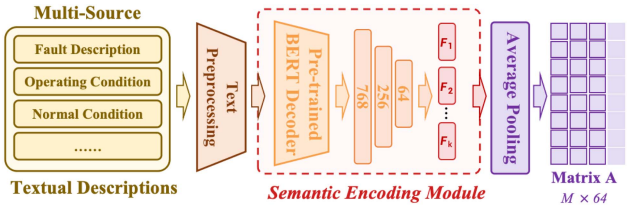

Fig. 3. Framework of the semantic encoding module. Multisource textual descriptions are encoded using BERT and projected into a 64-D semantic space, where average pooling is applied to form the attribute prototype matrix for each fault category.

Datasets with attribute: Existing publicly available benchmark datasets for ZSFD, such as three-phase transmission system (TPTS) [3] and Tennessee Eastman Process (TEP) [20], provide precise and comprehensive fault-attribute relationship matrices. These predefined attribute representations can be directly applied to ZSFD tasks. In this study, we utilize the attribute matrices provided by these datasets and convert them into a format suitable for model input, thereby ensuring both the integrity and semantic consistency of the attribute knowledge information.

Multimodal semantic construction: In real industrial settings, fault datasets rarely come with predefined attributes. To solve this problem, we introduce a multimodal semantic encoding approach that builds semantic representations directly from various unstructured text sources using pretrained language models [21]. As shown in Fig. 3, our framework follows three steps. 1) We collect text from multiple sources, including fault descriptions, operating conditions, and maintenance records, drawn from research papers, manuals, and open-source datasets. All text is cleaned and standardized to ensure consistency. 2) A pretrained BERT [22] encoder converts the text into contextual embeddings. These are then projected from 768 to 64 dimensions via a linear layer, making them compatible with our diagnosis model:

$$F_{\text{text}} = \phi_{\dim}(\text{Encoder}_{\text{text}}(T)) \tag{1}$$

where T represents the input text, and $\text{Encoder}_{\text{text}}$ denotes the text encoder of the language model, which produces the final text feature representation F_{text} through a linear transformation. This step reduces dimensionality, improves alignment with visual features, and helps prevent overfitting. 3) When multiple text sources describe the same fault, we average their embeddings into a single attribute prototype:

$$F^i_{\text{fault}} = \frac{1}{K}\sum_{k=1}^{K} F^i_k \tag{2}$$

where F^i_{fault} denotes the fused feature representation of the ith fault case, where K represents the number of information sources associated with the fault, and F^i_k corresponds to the feature representation of the kth information source. Finally, all fault prototypes are then organized into an attribute knowledge matrix:

$$A = [F^1_{\text{fault}}; F^2_{\text{fault}}; \ldots; F^M_{\text{fault}}] \tag{3}$$

where $A \in \mathbb{R}^{M\times 64}$ represents the fault-attribute knowledge matrix, M denotes the total number of fault types, and 64 is the dimensionality of the attribute features. These continuous-valued attributes capture richer semantic relationships than traditional binary attributes. During training, a cosine similarity loss aligns sample features with their corresponding attribute prototypes in a shared semantic space. Our method, **IF4FD**, converts unstructured industrial text into structured attribute representations. It preserves nuanced fault relationships and supports more interpretable diagnosis. In Section V, we demonstrate its application to the Cranfield Multiphase Flow Facility (CMFF) [23], a real-world dataset without predefined attributes, showing how our approach automatically constructs attribute knowledge from available text sources.

B. Multiscale Information Acquisition

Why acquire multiscale information? The faults encountered in industrial settings are often diverse, implying a necessity to address a variety of classification labels. However, limitations in equipment make it challenging to acquire multidimensional feature data, hindering our ability to describe a category from various perspectives. Motivated by this challenge, we aim to extract further informative knowledge from raw data to enhance the performance of the classifier.

Category knowledge representation: For multiclass classification tasks, imbuing the data distribution with discriminative qualities is the key to enhancing the classification performance. Thus, our first objective lies in acquiring category knowledge information that aids in classification. Formally, for the raw data x, we obtain its corresponding category knowledge embedding h_c through the encoder E, i.e., $h_c = E(x)$. In this process, we introduce the central loss $\mathcal{L}_{ct}$ to minimize intraclass distances, facilitating easier differentiation between embeddings of different classes. The central loss is defined as

$$\mathcal{L}_{ct} = \frac{1}{2}\sum_{k=1}^{K} ||h^k_c - c^k||^2_2 \tag{4}$$

where K is the number of categories in a batch. The h^k_c and c^k are the category knowledge embedding and center feature of class k, respectively. The center feature c^k is randomly initialized and updated in each batch.

Attribute knowledge representation: Maintaining the consistency of data and attribute space is a nonnegligible problem in ZSL [17]. Therefore, we aim to strengthen the alignment between data and corresponding attributes to achieve more effective knowledge transfer. To achieve this, we map the category knowledge embeddings h_c to a latent space of equal dimensionality as the attributes using an MLP, which is well-suited for industrial sensor data that lacks spatial locality and requires full connectivity to capture global feature correlations. Compared with sequential models such as LSTMs [24] or transformers [25], the relatively simple structure of MLPs better satisfies the real-time requirements of industrial intelligent applications and helps mitigate the risk of overfitting to some extent. The NCE loss $\mathcal{L}_{nce}$ which is introduced to maximize the mutual information

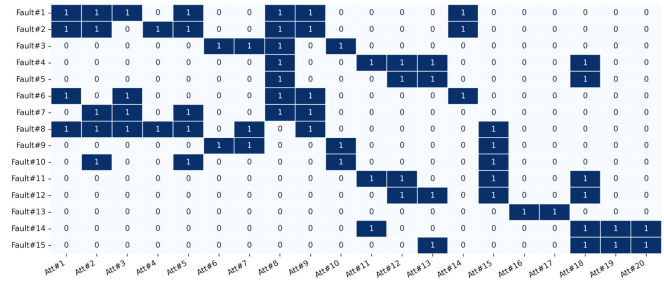

Fig. 4. Attribute matrix of TEP, where 0 denotes the absence of a particular attribute for this fault class and 1 denotes its presence.

between features and attributes:

$$\mathcal{L}_{nce} = log\frac{\exp\left(h_a^i \otimes a^i\right)}{\sum_{a^j \in A^S} \exp\left(h_a^i \otimes a^j\right)} \tag{5}$$

where $\otimes$ is matrix multiplication. For attribute knowledge embedding h_a^i of seen class i, the attribute a^i of the same class i can be regarded as a positive sample. Correspondingly, there are $N_S - 1$ negative samples. In this way, the NCE loss can also be regarded as a contrastive loss, which guides the model to distinguish between different classes while maintaining the semantic consistency of the embedding [15]. We adopt matrix multiplication instead of cosine similarity or Euclidean distance because it retains magnitude information and shows better empirical performance in our experiments. Furthermore, InfoNCE-based loss provides a suitable inductive bias for industrial fault diagnosis, where the semantic differences between fault types are often subtle. This contrastive learning framework helps the model learn more discriminative and semantically aligned representations, as supported by our ablation results in Section IV-C.

Feature knowledge extension: In addition to diversifying knowledge representation from various perspectives, we aim to extend feature knowledge based on the original data, thus expressing information within the data in higher dimensions. Therefore, via the decoder, we reconstruct the data h_f from the embedding h_c. Throughout this process, we introduce cosine similarity as a reconstruction loss $\mathcal{L}_{rec}$ to guide the decoder in generating diverse representations of the original data. This choice is motivated by cosine similarity's robustness to magnitude variations and its capacity to capture the directional consistency of feature vectors, which is crucial for preserving semantic content during reconstruction. Formally, the reconstruction loss $\mathcal{L}_{rec}$ is represented by

$$\mathcal{L}_{rec} = \sum_{i=1}^{n} \cos\left(x^i, h_f^i\right) \tag{6}$$

where n denotes the number of samples in a batch. Regarded as a form of data augmentation, incorporating reconstructed data as part of multiscale information fosters the expression of feature knowledge.

Joint optimization: After combining the category knowledge h_c, attribute knowledge h_a, and feature knowledge h_f with the original sensor data x, we simply concatenate them into a single 920-dimensional vector. This direct approach preserves all original information from each source and lets the classifier automatically learn which features are most useful. During training, we optimize the model using the following joint loss function $\mathcal{L}$:

$$\mathcal{L} = \alpha\mathcal{L}_{rec} + \beta\mathcal{L}_{nce} + \gamma\mathcal{L}_{ct} \tag{7}$$

where α, β, and γ are the tradeoff parameters.

TABLE I
FINE-GRAINED ATTRIBUTES OF THE TEP DATASET

No.	Attribute Description
$Att\#1$	Input A is changed.
$Att\#2$	Input C is changed.
$Att\#3$	A/C ratio is changed.
$Att\#4$	Input B is changed.
$Att\#5$	Related with pipe 4.
$Att\#6$	Temperature of input D is changed.
$Att\#7$	Related with pipe 2.
$Att\#8$	Disturbance is step changing.
$Att\#9$	Input is changed.
$Att\#10$	Temperature of input is changed.
$Att\#11$	Occurred at reactor.
$Att\#12$	Temperature of cooling water is changed.
$Att\#13$	Occurred at condenser.
$Att\#14$	Related with pipe 1.
$Att\#15$	Disturbance is random varying.
$Att\#16$	Model parameters are changed.
$Att\#17$	Disturbance is slow drift.
$Att\#18$	Related with cooling water.
$Att\#19$	Related with valve.
$Att\#20$	Disturbance is sticking.

TABLE II
DATA DIVISION OF THE TEP DATASET

Group	Seen Classes	Unseen Classes
A	2-4, 7-13, 15	1, 6, 14
B	1-3, 5, 6, 8, 9, 11-15	4, 7, 10
C	1-7, 9, 10, 13-15	8, 11, 12
D	1, 4, 6-15	2, 3, 5

IV. EXPERIMENT

A. Experimental Setup

Datasets: Our method is evaluated on two ZSFD benchmarks: TEP [20] and TPTS [3]. The TEP simulates 21 common industrial fault scenarios, of which the 15 well-documented faults are selected for evaluation. Each fault class contains 480 samples with 52 variables. Following FDAT [9], we adopt 20 fine-grained attributes (see Table I) to construct the fault–attribute correlation matrix (see Fig. 4). As shown in Table II, during the experiment, 12 of 15 fault classes are divided into training sets, and the other three classes are divided into test sets. The TPTS models a $110-$kV power system with four generators, containing six fault classes (one normal, five faulty) each described by six sensor variables. We define four phase-related attributes: $Att\#1$ (Phase A), $Att\#2$ (Phase B), $Att\#3$ (Phase C), and $Att\#4$

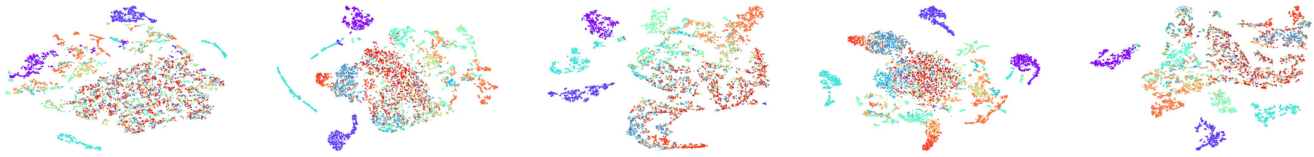

Fig. 5. T-SNE visualization of feature embeddings of 15 fault classes on the TEP. Our full **IF4FD** framework produces tighter clusters with clearer boundaries, demonstrating the effectiveness of our multiscale fusion strategy in enhancing both intraclass cohesion and interclass separation. (a) FDAT [11]. (b) SCE [5]. (c) IF4FD (Ours).

(Ground). The fault-attribute relationships are defined as follows: $Fault\#1$ has no phase involvement [0,0,0,0]; $Fault\#2$ involves Phase A and Ground [1,0,0,1]; $Fault\#3$ involves Phases B and C [0,1,1,0]; $Fault\#4$ involves Phases A, C and Ground [1,0,1,1]; $Fault\#5$ involves Phases B, C and Ground [0,1,1,1]; and $Fault\#6$ involves all attributes [1,1,1,1]. Following the SCE [3], $Faults\ \#2$, $\#5$, and $\#6$ are used as seen classes during training, while faults $\#3$ and $\#4$ are held out as unseen classes for testing.

Implementation details: Given the limited number of classes in the TPTS, we primarily use it for basic accuracy comparison, reserving more comprehensive evaluation for the richer TEP. Our experimental setup follows established practices from FDAT [9] and SCE [3]. For TPTS, models are trained on the original data, while TEP data are preprocessed using the FDAT before input. We set the loss weights in (7) as $\alpha = 0.05$, $\beta = 0.5$, and $\gamma = 0.5$. The reconstruction loss is lightly weighted to avoid distracting from the main classification task, while the central loss and NCE loss are balanced equally to jointly promote class separation and attribute alignment. These values were determined through empirical tuning and yielded stable convergence and strong performance. Our implementation is based on PyTorch.[1] The encoder and decoder each contain two linear layers with ReLU, batch normalization, and dropout. The encoder reduces a 400-D input to 100 dimensions, and the decoder reverses this process. The MLP uses two linear layers ($100 \rightarrow 100 \rightarrow 20$) with batch normalization, ReLU, and dropout. All models are trained with RMSProp at a learning rate of 0.01.

B. Comparative Results

Comparison with ZSFD methods: We compare **IF4FD** against seven recent ZSFD techniques. On the TPTS dataset (see Table III), **IF4FD** achieves top accuracy: 98.8% with Naive Bayes and 92.4% with Random Forest. This outperforms the strongest baseline (CycleGAN-SD) by 2.3% (NB) and 3.7% (RF), demonstrating improved generalization across classifiers. On the TEP dataset (Table IV), **IF4FD** consistently outperforms all baselines when using Naive Bayes. Compared to SCE and FDAT, it shows average accuracy gains of 5.1% and 13.0% , respectively, with a notable 12.5% improvement in Group B. To further analyze feature quality, we visualize embeddings using t-SNE (Fig. 5) for Group B, a challenging case with class overlap.

[1] Source code is available at https://anonymous.4open.science/r/ZSFD-B5B9/our_project_page.

TABLE III
COMPARATIVE RESULTS ON TPTS, WHERE NB AND RF MEAN THE EMPLOYMENT OF NAIVE BAYES AND RANDOM FOREST CLASSIFIERS

Method	Year	NB (%)	RF (%)	Average (%)
FDAT [9]	2021	68.8	69.6	69.1
FREE [26]	2021	91.0	81.8	86.4
SCE [3]	2022	95.3	90.4	92.9
FAGAN [18]	2022	86.4	86.6	86.5
SRWGAN [27]	2022	77.9	79.1	78.5
VAEGAN-AR [28]	2024	87.8	79.5	83.7
CycleGAN-SD [29]	2025	96.5	88.7	92.6
Ours	**2025**	**98.8**	**92.4**	**95.6**

Bold values indicate the best results.

TABLE IV
COMPARATIVE RESULTS ON TEP, WHERE NB AND RF MEAN THE EMPLOYMENT OF NAIVE BAYES AND RANDOM FOREST CLASSIFIERS

Classifier	RF (%)			NB (%)		
Method	FDAT	SCE	**Ours**	FDAT	SCE	**Ours**
Group A	88.9	**90.9**	90.4	80.3	89.5	**94.2**
Group B	52.0	57.3	**69.6**	62.6	78.1	**90.6**
Group C	48.8	55.6	**57.5**	59.0	62.4	**65.3**
Group D	66.3	**67.1**	66.9	72.4	76.0	**76.3**
Average	64.0	67.7	**71.1**	68.6	76.5	**81.6**

Bold values indicate the best results.

TABLE V
QUANTITATIVE EVALUATION OF EMBEDDING QUALITY USING SILHOUETTE, DAVIES-BOULDIN, AND CALINSKI–HARABASZ SCORES.

Method	Silhouette ↑	DB ↓	CH ↑
FDAT	0.324	1.26	5.32×10^2
SCE	0.234	1.52	7.52×10^2
IF4FD(ours)	**0.445**	**0.927**	$\mathbf{1.88}\times10^3$

Bold values indicate the best results.

While FDAT and SCE produce mixed clusters with weak separation, **IF4FD** forms compact, well-separated clusters. These observations are quantified in Table V using three metrics: 1) Silhouette Score (higher = better separation), 2) Davies–Bouldin Index (lower = tighter clusters), 3) Calinski–Harabasz Score (higher = denser clustering). **IF4FD** achieves the best scores across all metrics, confirming its ability to learn discriminative embeddings that support robust zero-shot diagnosis.

Comparison with ZSL methods: We also compare our **IF4FD** with various ZSL models on the TEP dataset. These include

TABLE VI
COMPARATIVE RESULTS WITH ZSL METHODS ON TEP DATASET (%)

Method	Year	Group A	Group B	Group C	Group D	Average
DAP [30]	2009	54.2	62.6	40.1	55.5	53.1
IAP [30]	2009	55.5	60.7	45.0	36.6	49.5
DEVISE [31]	2013	50.6	74.2	32.9	51.5	52.3
ESZSL [32]	2015	57.2	33.3	39.5	39.7	42.4
SJE [33]	2016	74.6	33.1	34.0	63.9	51.4
SAE [34]	2017	45.7	74.3	33.5	65.6	54.8
FAGAN [18]	2021	84.5	76.9	62.5	74.6	75.0
FREE [26]	2021	81.0	75.6	**71.5**	**78.8**	76.3
SRWGAN [27]	2022	79.2	77.6	69.4	77.4	75.2
SSB-ZSL-1DCNN [35]	2023	84.3	76.9	62.1	59.7	70.8
AFT [36]	2023	66.8	41.4	40.7	36.8	46.4
GLA-ZSL [37]	2025	**97.4**	86.2	54.8	77.7	79.0
Ours	2025	94.2	**90.6**	65.3	76.3	**81.6**

Bold values indicate the best results.

classic embedding-based methods (DAP [30], IAP [30], DEVISE [31], ESZSL [32], SJE [33], SAE [34]) and newer generative approaches (FAGAN [18], FREE [26], SRWGAN [27], AFT [36], GLA-ZSL [37]). As shown in Table VI, traditional ZSL methods perform poorly on industrial fault data—ESZSL and SJE achieve only 42.4% and 51.4% average accuracy, respectively. Newer models like FREE (76.3%) and SRWGAN (75.2%) perform better, with GLA-ZSL reaching the previous state-of-the-art at 79.0% . Our method outperforms all others, achieving 81.6% average accuracy and top results in Group B (90.6%). This shows that our multiscale fusion framework generalizes well, even with noisy industrial data. We also observe that all methods struggle in Group C, where faults #8, #11, and #12 are highly similar and exhibit complex coupling effects. While our method (65.3% in Group C) still beats GLA-ZSL (54.8%), distinguishing such similar faults remains challenging. Future work will focus on improving feature discrimination for highly similar fault types.

Comparison with supervised classification methods: We also compare our model with supervised classifiers, including linear support vector machine (LSVM), RF, NB, XGBoost, AdaBoost, K-nearest neighbor (KNN), gradient boosting machine (GBM), and Light GBM (LGBM), on TEP Groups A and C. Each classifier was trained using 1, 10, 50, 200, and 500 labeled samples per class (k-shot). For comparison, our results are obtained with zero-shot. As shown in Fig. 6, our zero-shot model achieves 94.2% accuracy on Group A, outperforming all supervised methods trained with 200 samples and matching their average performance with 500 samples. On the more challenging Group C, we reach 65.3% , surpassing most 200-shot models and even exceeding several 500-shot classifiers such as LSVM, AdaBoost, and KNN. These results show that IF4FD performs strongly under realistic low-data conditions, often matching or exceeding conventional supervised methods without using any labeled samples from target classes.

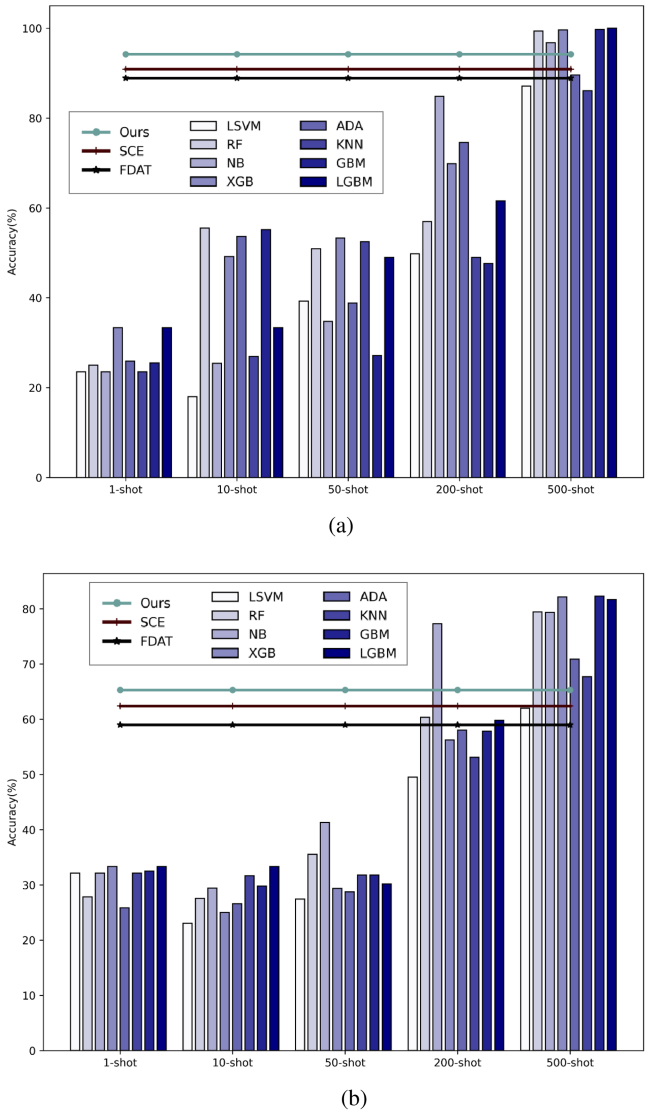

Fig. 6. Comparison of results with the supervised classification methods, k-shot means there are k samples available for training. (a) Group A. (b) Group C.

C. Ablation Experiment

We conduct ablation studies to evaluate three components: loss functions, fusion strategy, and knowledge integration.

Loss function: Table VII shows the impact of removing individual loss terms on TEP performance. Removing $\mathcal{L}_{ct}$ or $\mathcal{L}_{nce}$ yields similar results except in Group B, where data distribution is more compact. We also compare reconstruction losses: cosine similarity outperforms both MSE and L1 loss. While MSE and L1 minimize numerical deviations, cosine similarity maintains directional consistency in high-dimensional spaces, better preserving semantic relationships. Gradient analysis (see Fig. 7) also reveals that $\mathcal{L}ct$ and $\mathcal{L}nce$ cooperate to enhance discrimination, while $\mathcal{L}_{rec}$ operates independently to refine feature details. Furthermore, visualization [see Fig. 5(a)] confirms that $\mathcal{L}_{ct}$ enhances interclass separation. Removing $\mathcal{L}_{rec}$ causes the most significant performance drop [see Fig. 5(b)]. This complementary interaction supports effective multiscale optimization.

Fusion strategy: We also evaluate our concatenation-based multiscale fusion strategy against alternatives in Table VII, which shows that all alternative strategies yield inferior overall performance on the TEP dataset. Among them, weighted

TABLE VII
ABLATION STUDY OF LOSS FUNCTIONS AND FUSION STRATEGIES (%)

Component	Group A	Group B	Group C	Group D	Average
Ours (Full)	**94.2**	**90.6**	**65.3**	76.3	**81.6**
Loss Function Ablation:					
$w/o\ \mathcal{L}_{ct}$	87.4	58.6	62.2	74.9	70.8
$w/o\ \mathcal{L}_{nce}$	87.2	83.8	63.8	**76.9**	77.8
$w/o\ \mathcal{L}_{rec}$	82.0	70.3	63.2	74.4	72.5
$\mathcal{L}_{rec}$ w/ MSE	93.6	61.2	62.3	72.8	72.5
$\mathcal{L}_{rec}$ w/ L1	92.4	67.5	65.1	75.6	75.2
$\mathcal{L}_{rec}$ w/ Cosine Sim.	**94.2**	**90.6**	**65.3**	76.3	**81.6**
Fusion Strategy Ablation:					
Weighted Concat.	87.4	74.3	63.7	74.6	75.0
Attention Fusion	68.4	76.1	57.4	69.1	67.8
Element-wise Add.	69.2	41.7	56.7	63.3	57.7
Knowledge Ablation:					
$w/o\ h_a$	80.4	78.0	61.4	71.3	72.7
$w/o\ h_f$	88.9	64.0	62.9	74.0	72.3
$w/o\ h_c$	86.7	69.9	64.6	69.5	72.7
$w/o\ h_f\&h_c$	91.0	41.3	60.8	72.4	63.4
$w/o\ h_a\&h_c$	80.4	70.8	61.3	68.3	70.2
$w/o\ h_f\&h_a$	80.3	62.6	59.0	72.4	68.6

Bold values indicate the best results.

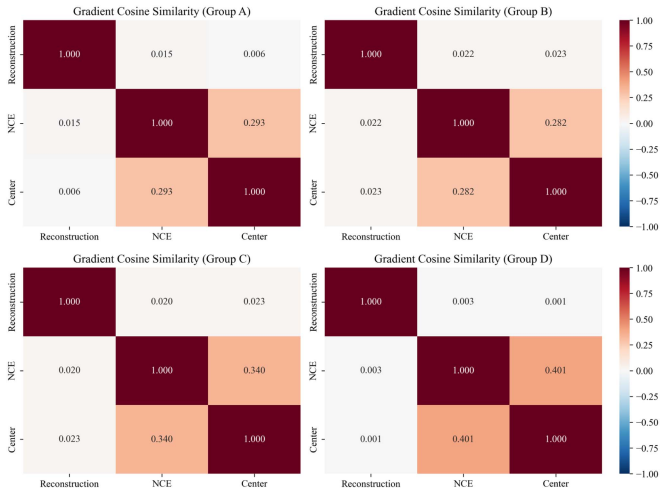

Fig. 7. Gradient similarity among loss terms $\mathcal{L}_{ct}$, $\mathcal{L}_{nce}$ and $\mathcal{L}_{rec}$. Red indicates cooperative optimization and blue indicates gradient conflict.

concatenation exhibits a slightly smaller performance drop with 75.0% , which indirectly indicates that preserving the original feature representations during fusion is beneficial. In contrast, attention-based fusion and elementwise addition introduce additional feature rescaling or aggregation operations, which may suppress fine-grained fault-relevant information and lead to noticeable performance degradation. These results demonstrate that the direct concatenation strategy effectively retains the complementary characteristics across different scales, thereby achieving superior diagnostic performance.

Knowledge integration: We conduct an ablation study on three knowledge (h_a, h_f, and h_c), comparing the performance when using each type individually, in pairwise combinations, and jointly. As shown in Table VII, the joint utilization of all three knowledge types achieves the best performance, indicating that their contributions are not merely additive. Specifically, category knowledge provides global discriminative cues, attribute knowledge introduces transferable semantic structure, and feature knowledge supplements fine-grained representations. Together,

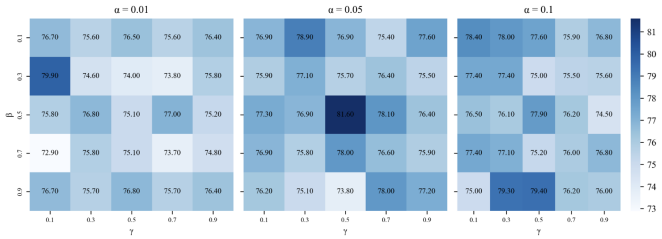

Fig. 8. Parameter sensitivity analysis on the TEP under different combinations of α, β, and γ.

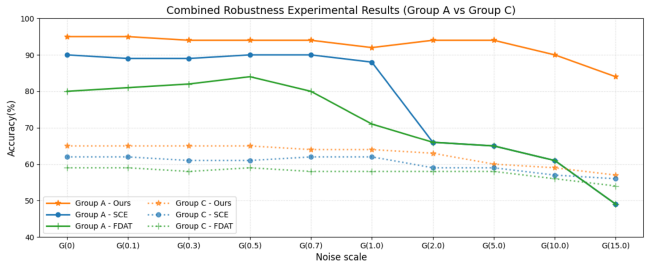

Fig. 9. Robustness experimental results of the proposed model.

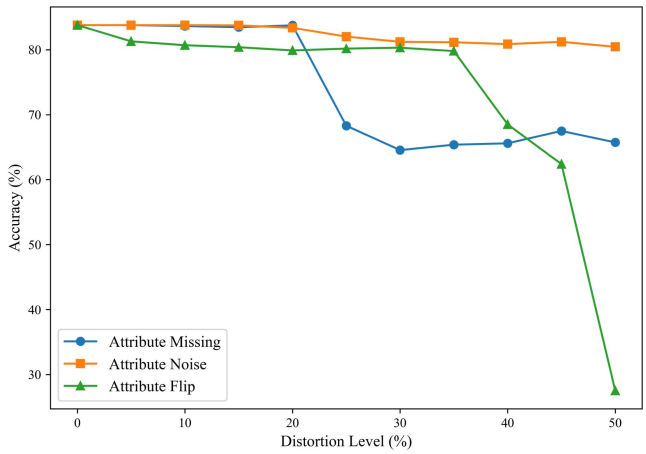

Fig. 10. Robustness analysis of model attribute disturbance.

these three knowledge types form a complementary synergy in industrial fault diagnosis scenarios.

D. Parameter Sensitivity Analysis

As shown in Fig. 8, we performed a parameter sensitivity analysis on the TEP dataset under different combinations of $\alpha \in \{0.01, 0.05, 0.1\}$ and $\beta, \gamma \in \{0.1, 0.3, 0.5, 0.7, 0.9\}$, evaluating model performance under different combinations of α, β, and γ in (7). Average accuracy across groups A–D is used as the evaluation metric. The model achieved optimal and most stable performance when $\alpha = 0.05$, $\beta = 0.5$, and $\gamma = 0.5$. In this configuration, the reconstruction loss ($\alpha = 0.05$) acts as a lightweight regularizer, while the central loss and NCE loss ($\beta = \gamma = 0.5$) are balanced to jointly promote category discrimination and semantic alignment. These values were determined

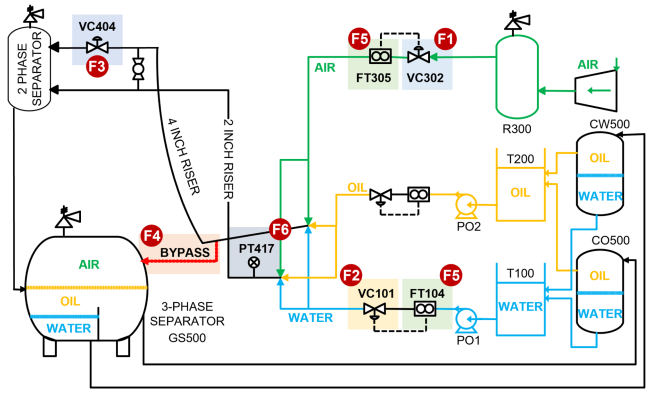

Fig. 11. Schematic of the process flow and fault distribution in the system of the Cranfield Multiphase Flow Facility dataset. Red circles (F1–F6) indicate the locations where faults occur, while the shaded areas represent the corresponding regions affected by each fault. Specifically, F1 corresponds to Air line blockage, F2 to Water line blockage, F3 to Top separator input blockage, F4 to Open direct bypass, F5 to Slugging conditions, and F6 to Pressurization of the 2-in line. This diagram is adapted from the original publication [23], with unrelated system components omitted to improve clarity and highlight the six primary fault types.

TABLE VIII
COMPUTATIONAL COMPLEXITY COMPARISON OF METHODS

Method	Parameter Complexity	FLOPs (Relative)	Model Size	Inference Time(ms)	Average (TEP %)
SCE	1k	1×	4 KB	0.05	76.5
AFT	2k	3×	8 KB	0.3	46.4
FDAT	37k	78×	148 KB	3.5	68.6
IF4FD (80-d)	271k	578×	1.08MB	2.7	69.4
IF4FD (920-d)	**271k**	**580×**	**1.08MB**	**2.1**	**81.6**
FREE	1.22M	19,877×	4.88MB	13.0	76.3

Bold values highlight the proposed method.

through empirical tuning and consistently delivered strong convergence and classification results.

E. Comparison of Model Complexity

As shown in Table VIII, we systematically evaluate the computational efficiency of **IF4FD** in terms of parameter count, FLOPs, model size, and inference time. While SCE, AFT, and FDAT maintain compact parameter sizes, their diagnostic accuracy remains limited. FREE, originally designed for image classification, incurs substantially higher computational costs. To preserve complete information from all knowledge sources, we concatenate the original sensor data (x, 400-d) with category (h_c, 100-d), attribute (h_a, 20-d), and feature (h_f, 400-d) representations, forming a 920-D fused vector (see Fig. 2). We also test a compressed version where x, h_c, and h_f are reduced to 20 dimensions each and concatenated with h_a (20-d), yielding an 80-D input. This compressed model shows similar computational costs but suffers a 12.2% accuracy drop on the TEP dataset, confirming that high-dimensional fusion better preserves discriminative information. Overall, the full **IF4FD** model achieves the highest accuracy (81.6%) with moderate resource requirements: 2.1-ms inference time and 1.08 MB model size. Compared to FDAT, **IF4FD** improves accuracy by 5.1% while maintaining lower latency and memory usage than larger models like FREE, demonstrating an effective balance between performance and efficiency for industrial deployment.

F. Robustness Check

Industrial sensor data often contains noise and uncertainty, making model robustness critical. We evaluate **IF4FD** under both Gaussian noise and semantic attribute perturbations. We firs test with additive Gaussian noise with mean of 0 and a variance of n $(0.1, 0.3, 0.5, 0.7, 1.0, 2.0, 5.0, 10.0, 15.0)$ on TEP Groups A and C (see Fig. 9). FDAT and SCE degrade significantly at $G(0.7)$ and $G(2.0)$, respectively, while **IF4FD** maintains stable performance until $G(10.0)$. Even at $G(15.0)$, **IF4FD** achieves 84.0% accuracy on Group A versus 48.7% for FDAT and SCE, demonstrating superior noise resistance through multiscale feature fusion. As shown in Fig. 10, we further evaluate model robustness under three types of semantic uncertainty, with perturbation levels ranging from 0 to 50% in 5% increments.

1) *Attribute missing:* Performance remains stable up to 20% missing ratio; maintains 65% accuracy even with 50% missing attributes.

2) *Attribute noise:* Average accuracy stays above 80% across all noise levels despite minor fluctuations.

3) *Attribute flipping:* Stable performance up to 35% flip ratio; preserves meaningful classification capability beyond 40% . These results confirm that **IF4FD** handles both sensor-level noise and semantic-level uncertainty effectively, making it suitable for real industrial environments where data quality and knowledge reliability vary.

V. CASE STUDY

This case study uses data from Cranfield University's Multiphase Flow Facility [23], which was collected from an operational research plant rather than simulated. This real-world origin gives the data greater practical relevance than the computer-generated datasets used in Section IV. The experimental setup, shown in Fig. 11, is built around a pipeline network with two main loops of different diameters. The 4-in loop has a 55 m downward-sloping (2°) pipe linked to a 10.5 m catenary riser, and the 2-in loop has a 40 m horizontal pipe connected to a 10.5 m vertical riser. Both loops join at a gas–liquid separator on a 10.5 m high platform. A final, ground-level separator (GS500, 11 m^3) then vents air and recycles the water. All data are recorded at 1 Hz by a DeltaV [38] Fieldbus based supervisory control and data acquisition system for reliable collection. As shown in Table IX, the resulting dataset includes one normal state and six different fault types—like blockages and operational errors—that mimic real industrial problems for fault diagnosis research.

1) *Fault Case #1. Air line blockage:* Simulates gradual obstruction in the air supply line by incrementally closing valve VC302 from fully open. System responses are recorded at defined intervals to evaluate dynamic characteristics under deteriorating airflow.

TABLE IX
SUMMARY OF FAULTS IN CRANFIELD THREE-PHASE FLOW DATASET

Fault	Fault Name	Location	Measured Variable	Samples
1	Air line blockage	VC302	Valve VC302 position (%)	9,734
2	Water line blockage	VC101	Valve VC101 position (%)	8,499
3	Top separator input blockage	VC404	Valve VC404 position (%)	8,499
4	Open direct bypass	BYPASS	Riser bottom pressure (MPa)	12,255
5	Slugging conditions	FT305 FT104	Air flow rate (Sm^3/s) Water flow rate (kg/s)	5362
6	Pressurization of 2-inch line	PT417	Pressure in 2-inch line (MPa)	4,870

Bold values indicate the best results.

TABLE X
GROUP SETTING OF ZERO-SHOT INDUSTRIAL FAULT DIAGNOSIS FOR CRANFIELD MULTIPHASE FLOW FACILITY DATASET

Group	Training		Target	
	Seen faults	Total	Unseen faults	Total
A	2,3,4,5	34615	1,6	14604
B	1,2,5,6	28465	3,4	20754
C	1,2,3,6	31602	4,5	17617
D	1,2,3,4	38987	5,6	10232

2) *Fault Case #2. Water line blockage:* Simulates pipeline obstruction by gradually closing water line valve VC101. While the procedure mirrors Case 1, system responses differ significantly due to the distinct physical properties of water versus air.
3) *Fault Case #3. Top separator input blockage:* Simulates obstruction at the top separator input by operating valve VC404. Unlike previous cases, this valve can be remotely controlled and its angular position precisely monitored.
4) *Fault Case #4. Open direct bypass:* Simulates a leakage at the riser bottom by opening the normally closed BYPASS valve. This diverts fluid directly to the separator, bypassing the main riser and creating insufficient flow to the test area.
5) *Fault Case #5. Slugging conditions:* Induced by reducing air and liquid flow rates (monitored via FT305, FT104) to accumulate liquid at the riser bottom. This causes periodic pressure buildup and release, leading to fluctuations in pressure and flow rate.
6) *Fault Case #6. Pressurization of the 2-in line:* Simulates abnormal pressurization in the normally isolated 2-in pipeline by opening the bridge valve connecting it to the 4-in main line. Pressure was monitored using transmitter PT417.

As shown in Table IX, the final dataset contains 77 600 records of normal operation and six distinct fault types. The normal data significantly outnumbers the fault data, which could cause a model to become biased and harm its ZSFD performance. To prevent this, we excluded normal samples from the ZSFD process. For the experiments, we treated two randomly chosen fault types as unseen classes and the other four as seen classes. This random selection was repeated in four independent trials. The specific setup and the division of training and test data for each trial are shown in Table X.

TABLE XI
COMPARISON WITH ZERO-SHOT FAULT INDUSTRIAL DIAGNOSIS METHOD ON CRANFIELD MULTIPHASE FLOW FACILITY DATASET

Method	Accuracy(%)				Average(%)
	Group A	Group B	Group C	Group D	
FDAT	66.7	56.9	69.6	20.7	53.5
SCE	68.2	**74.8**	78.0	57.2	69.6
Ours	**91.1**	72.8	**82.0**	**94.5**	**85.1(↑15.5)**

A. Zero-Shot Fault Diagnosis Implementation

After preparing the data, we apply the **IF4FD** to this real-world dataset. The first step is to create attribute descriptions for the different fault types, a process known as multimodal semantic encoding. Unlike standard datasets that come with predefined attributes, this is the first time the ZSFD method has been used on this dataset. Therefore, we have to build the attribute knowledge from scratch. Thus, we gather information from various sources, including official documentation and research papers, to create detailed text descriptions for each of the seven categories (one normal and six fault types). We then use the text encoder of pretrained BERT to convert these descriptions into numerical vectors. By processing the text and using an average pooling strategy, we generate a 7×64 fault attribute matrix. The size of 64 is chosen as it performed best in our tests. This matrix is different from the simple binary (0/1) attributes found in datasets like TEP. Our matrix contains continuous values that, while not directly readable by humans, allow the model to capture more complex and nuanced fault characteristics. After generating the attribute matrix, we perform the multiscale information extraction step. For the experiments, we set the α, β and γ in (7) as 0.05, 0.5, and 0.5, respectively, and conduct comparative experiments against FDAT and SCE.

B. Results and Comparative Analysis

Table XI presents a performance comparison of our proposed **IF4FD** method with the state-of-the-art ZSFD methods FDAT and SCE on the CMFF dataset, while Fig. 12 illustrates the optimal confusion matrices from four experimental groups. To ensure fairness of evaluation, all three methods employ identical Gaussian Bayesian classifiers for the zero-shot classification tasks. It should be emphasized that each set of accuracy metrics represents the average results of over ten independent experimental trials, which significantly increases the statistical reliability and robustness of our experimental conclusions. Regarding the evaluation framework, this case study compares the proposed method with two recently developed approaches specifically designed for ZSFD, namely FDAT and SCE, rather than with conventional ZSL algorithms. This selection is based on two primary considerations. First, traditional methods were originally developed for computer vision applications, and their underlying assumptions are not well aligned with the distinctive requirements of industrial fault diagnosis. Second, FDAT and SCE, as domain-specific approaches, have

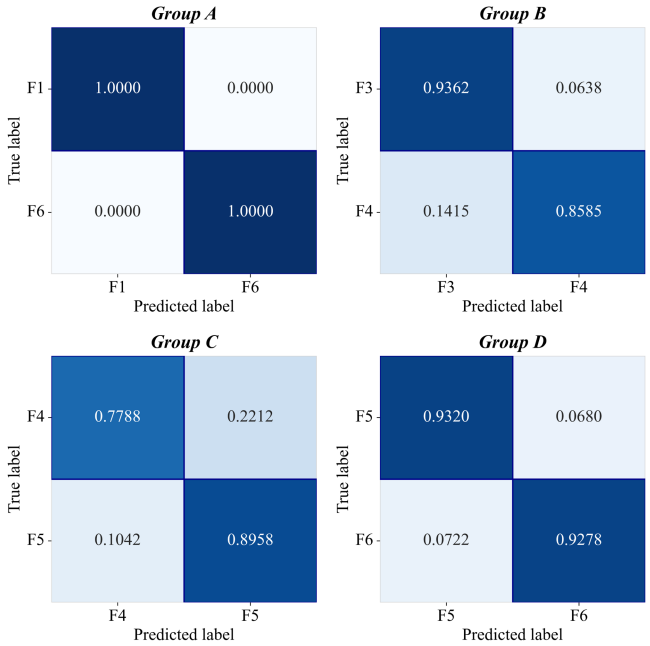

Fig. 12. Confusion matrices of the best results of the four groups for the Cranfield Multiphase Flow Facility Dataset.

demonstrated consistent superiority over conventional counterparts in their original studies. Therefore, benchmarking the proposed method against these specialized techniques enables a more rigorous and application-oriented evaluation of its practical effectiveness.

The **IF4FD** achieves the highest accuracy in three out of four groups (A, C, and D), with scores of 91.1% , 82.0% , and 94.5% , leading to a top average accuracy of 85.1% . This is a significant improvement over FDAT (53.5%) and SCE (69.6%). Although SCE had a slightly higher accuracy in Group B (74.8% vs. **IF4FD**'s 72.8%), our method was still much better than FDAT (56.9%). All methods struggle more in Groups B and C, where Fault Case #3 (top separator blockage) and Fault Case #4 (open bypass) are the unseen faults being tested. The difficulty arises because these faults are physically very different from the ones used in training: Fault Case #3 involves a precisely controlled valve, creating signal patterns unlike a simple blockage, while Fault Case #4 simulates a leak that reroutes flow, which is fundamentally different from an obstruction. This creates a large semantic gap, making it harder for models to generalize. Despite these challenges, **IF4FD** demonstrated strong generalization. Its success comes from its multiscale feature extraction framework and the use of multimodal semantic encoding. This allows the model to better understand the differences between fault types and identify key diagnostic features, leading to superior performance across diverse test scenarios.

VI. Conclusion

We propose a ZSFD framework, called **IF4FD**, based on multiscale information fusion. First, a structured semantic attribute knowledge graph is constructed through multimodal semantic encoding techniques. Then we augment the original data from the perspectives of category knowledge, attribute knowledge, and feature knowledge, thereby offering diverse representations of raw sensor data to aid knowledge transfer and fault classification. Experimental findings confirm that our proposed approach outperforms existing ZSFD and general ZSL methods, notably enhancing ZSFD tasks. Moreover, in scenarios with limited samples, our zero-shot-based approach surpasses supervised classification algorithms. By introducing noise to the data, we further demonstrate the model's robustness. Extensive experimental results unequivocally indicate the efficacy of employing the ZSFD method based on multiscale information for fault diagnosis in industrial production processes. Systematic validation on the real-world CMFF dataset highlights the method's strong potential for identifying novel and rare faults in practical applications.

Limitation and future work: While the proposed method demonstrates effectiveness across multiple datasets, certain practical considerations merit attention. The attribute extraction quality depends on the availability and diversity of data sources; in scenarios with limited textual information or imbalanced class distributions, more robust semantic encoding strategies may be beneficial. In addition, the feature concatenation approach increases dimensionality, which, although manageable in current applications, may pose scalability challenges for extremely high-dimensional industrial sensor systems. Computational efficiency, while adequate on standard hardware, could be further optimized for real-time critical applications requiring sub-second response times. In future work, our model can be extended to generalized ZSL settings that leverage unlabeled test data, developing cross-domain transfer learning mechanisms for multifacility deployment to reduce retraining costs, integrating temporal modeling approaches such as recurrent architectures to capture dynamic fault evolution patterns, and exploring multirate sampling strategies to handle sensors with varying sampling frequencies in complex industrial environments.

References

[1] C. Tang et al., "Attention-based early warning framework for abnormal operating conditions in fluid catalytic cracking units," *Appl. Soft Comput.*, vol. 153, 2024, Art. no. 111275.

[2] Y. Gui, M. Yi, H. Yin, P. Zhang, D. Zhao, and L. Cai, "A novel zero-shot fault identification based ontransfer learning," in *Proc. Chin. Intell. Syst. Conf.*, 2022, pp. 115–124.

[3] Z. Hu, H. Zhao, L. Yao, and J. Peng, "Semantic-consistent embedding for zero-shot fault diagnosis," *IEEE Trans. Ind. Inform.*, vol. 19, no. 5, pp. 7022–7031, May 2023.

[4] C. Tang et al., "Deep learning in nuclear industry: A survey," *Big Data Mining Anal.*, vol. 5, no. 2, pp. 140–160, 2022.

[5] W. Yu and C. Zhao, "Online fault diagnosis in industrial processes using multimodel exponential discriminant analysis algorithm," *IEEE Trans. Control Syst. Technol.*, vol. 27, no. 3, pp. 1317–1325, May 2019.

[6] X. Song et al., "Evolutionary multi-objective spiking neural architecture search for image classification," *IEEE Trans. Evol. Comput.*, to be published,doi: 10.1109/TEVC.2025.3528471.

[7] Y. Zhuo and Z. Ge, "Data guardian: A data protection scheme for industrial monitoring systems," *IEEE Trans. Ind. Inform.*, vol. 18, no. 4, pp. 2550–2559, Apr. 2022.

[8] C. Tang et al., "Rethinking generalized zero-shot learning: A synthesized per-instance attribute perspective," *IEEE Trans. Image Process.*, vol. 34, pp. 5847–5859, 2025.

[9] L. Feng and C. Zhao, "Fault description based attribute transfer for zero-sample industrial fault diagnosis," *IEEE Trans. Ind. Inform.*, vol. 17, no. 3, pp. 1852–1862, Mar. 2021.
[10] W. Wang, V. W. Zheng, H. Yu, and C. Miao, "A survey of zero-shot learning: Settings, methods, and applications," *ACM Trans. Intell. Syst. Technol.*, vol. 10, no. 2, pp. 1–37, 2019.
[11] Y. Gao et al., "Improving generalized zero-shot learning via cluster-based semantic disentangling representation," *Pattern Recognit.*, vol. 150, 2024, Art. no. 110320.
[12] C. Tang, J. Lv, Y. Chen, and J. Guo, "An angle-based method for measuring the semantic similarity between visual and textual features," *Soft Comput.*, vol. 23, pp. 4041–4050, 2019.
[13] Z. Zang, C. Lin, C. Tang, T. Wang, and J. Lv, "Zero-shot aerial object detection with visual description regularization," in *Proc. AAAI Conf. Artif. Intell.*, 2024, pp. 6926–6934.
[14] C. Tang, Y. Kuang, J. Lv, and J. Hu, "SAN: Sampling adversarial networks for zero-shot learning," in *Proc. Int. Conf. Neural Inf. Process.*, 2020, pp. 626–638.
[15] C. Tang, X. Yang, J. Lv, and Z. He, "Zero-shot learning by mutual information estimation and maximization," *Knowl.-Based Syst.*, vol. 194, 2020, Art. no. 105490.
[16] Z. Zhai, X. Li, and Z. Chang, "Center-vae with discriminative and semantic-relevant fine-tuning features for generalized zero-shot learning," *Signal Processing, Image Commun.*, vol. 111, 2023, Art. no. 116897.
[17] C. Tang, Z. He, Y. Li, and J. Lv, "Zero-shot learning via structure-aligned generative adversarial network," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 11, pp. 6749–6762, Nov. 2022.
[18] Y. Zhuo and Z. Ge, "Auxiliary information-guided industrial data augmentation for any-shot fault learning and diagnosis," *IEEE Trans. Ind. Inform.*, vol. 17, no. 11, pp. 7535–7545, Nov. 2021.
[19] J. Chen, G. Wang, J. Lv, Z. He, T. Yang, and C. Tang, "Open-set classification for signal diagnosis of machinery sensor in industrial environment," *IEEE Trans. Ind. Inform.*, vol. 19, no. 3, pp. 2574–2584, Mar. 2022.
[20] "A plant-wide industrial process control problem," *Comput. Chem. Eng.*, vol. 17, no. 3, pp. 245–255, 1993.
[21] C. Yu et al., "GPT-NAS: Neural architecture search meets generative pre-trained transformer model," *Big Data Mining Anal.*, vol. 8, no. 1, pp. 45–64, Feb. 2025.
[22] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in *Proc. Conf. North Amer. chapter Assoc. Comput. linguistics: Hum. Lang. Technol.*, 2019, pp. 4171–4186.
[23] C. Ruiz-Cárcel, Y. Cao, D. Mba, L. Lao, and R. Samuel, "Statistical process monitoring of a multiphase flow facility," *Control Eng. Pract.*, vol. 42, pp. 74–88, 2015.
[24] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.
[25] A. Vaswani et al., "Attention is all you need," in *Proc. Adv. Neural Inform. Process. Syst.*, 2017, Art. no. 30.
[26] S. Chen et al., "Free: Feature refinement for generalized zero-shot learning," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2021, pp. 122–131.
[27] L. Feng, C. Zhao, and X. Li, "Bias-eliminated semantic refinement for any-shot learning," *IEEE Trans. Image Process.*, vol. 31, pp. 2229–2244, 2022.
[28] L. Shao, N. Lu, B. Jiang, and S. Simani, "Feature generating network with attribute-consistency for zero-shot fault diagnosis," *IEEE Trans. Ind. Inform.*, vol. 20, no. 5, pp. 7787–7796, May 2024.
[29] W. Liao, L. Wu, S. Xu, and S. Fujimura, "Cycle-consistent generating network based on semantic distance for zero-shot fault diagnosis," *IEEE Trans. Instrum. Meas.*, vol. 74, 2025, Art. no. 3518413.
[30] C. H. Lampert, H. Nickisch, and S. Harmeling, "Learning to detect unseen object classes by between-class attribute transfer," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2009, pp. 951–958.
[31] A. Frome et al., "Devise: A deep visual-semantic embedding model," in *Proc. Adv. Neural Inform. Process. Syst.*, 2013, Art. no. 26.
[32] B. Romera-Paredes and P. H. S. Torr, "An embarrassingly simple approach to zero-shot learning," in *Proc. 32nd Int. Conf. Int. Conf. Mach. Learn.*, 2015, pp. 2142–2151.
[33] Z. Akata, S. Reed, D. Walter, H. Lee, and B. Schiele, "Evaluation of output embeddings for fine-grained image classification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 2927–2936.
[34] E. Kodirov, T. Xiang, and S. Gong, "Semantic autoencoder for zero-shot learning," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 3174–3183.
[35] S. Zhang, H.-L. Wei, and J. Ding, "An effective zero-shot learning approach for intelligent fault detection using 1 d CNN," *Appl. Intell.*, vol. 53, no. 12, pp. 16041–16058, 2023.
[36] L. Fan, X. Chen, Y. Chai, and W. Lin, "Attribute fusion transfer for zero-shot fault diagnosis," *Adv. Eng. Informat.*, vol. 58, 2023, Art. no. 102204.
[37] H. Tang, W. Jing, D. Tang, Z. Yang, X. Yang, and W. Xie, "Global-local attention-aware zero-shot learning for industrial fault diagnosis," *IEEE Trans. Instrum. Meas.*, vol. 74, 2025, Art. no. 3517916.
[38] E. E, Co The deltav Digital Automation System, *Emerson*. [Online]. Available: https://www.emerson.com/en-us/automation/deltav. Accessed: Feb. 13, 2014.