SAE-FSC: A Siamese Attention Encoder–Based Few-Shot Cross-Domain Fault Diagnosis Framework for Bearings

Karkulali Pugalenthi[1*], Van Tung Tran[1], Ang Shiming[1], and Doan Ngoc Chi Nam[1*]

[1] *Digital Manufacturing Division, SIMTech, A*STAR, Singapore*
Karkulali_Pugalenthi@simtech.a-star.edu.sg
Tung_Tran@simtech.a-star.edu.sg
Ang_Shiming@simtech.a-star.edu.sg

[1*] *Corresponding author: doanncn@simtech.a-star.edu.sg*

ABSTRACT

Rolling element bearings are critical components in rotating machinery, where failures can cause severe downtime and safety risks. Existing fault diagnosis methods are predominantly supervised, requiring large amounts of labeled data across multiple operating conditions. However, in realistic industrial scenarios, such labeled datasets are scarce, and models trained on one regime often fail to generalize to others. To overcome this cross-domain generalization challenge, we propose a Siamese Attention Encoder–based few-shot cross-domain fault diagnosis (SAE-FSC) framework. The key novelty of this work lies in an attention-augmented Siamese encoder that extracts highly discriminative and transferable time-series features, coupled with a composite objective function that jointly optimizes supervised cross-entropy, pairwise binary cross-entropy, and domain adversarial loss. This combination enforces intra-class domain invariant feature learning across multiple operating conditions. Extensive experiments on the Case Western Reserve University (CWRU) dataset under leave-one-fault-out (LOFO) and leave-two-fault-out (LTFO) protocols demonstrate robust generalization across unseen fault types, load conditions, and fault severities, achieving a prediction accuracy of 87% for 5 shot learning.

1. INTRODUCTION

Rolling element bearings are critical components in rotating machinery, and their premature failures can lead to costly downtime and severe safety risks. Accurate and reliable fault diagnosis of bearings is therefore essential for predictive maintenance and safe operation of industrial systems. Over the past few decades, data-driven approaches have emerged as the dominant paradigm for bearing fault diagnosis (Jardine, Lin, & Banjevic, 2006; Randall & Antoni, 2011).

Traditional data-driven methods use time-series or image data for fault diagnosis and prognosis. For time-series data, vibration and related signals are preprocessed and converted into engineered features in the time, frequency, or time–frequency domains, which are then classified using traditional machine learning algorithms such as support vector machines (SVMs) (Mishra, Choudhary, Mohanty, & Fatima, 2021), hidden Markov models (HMMs) (Ocak & Loparo, 2005), or extreme learning machines (ELMs) (Ma, Yu, & Cheng, 2022). Techniques such as envelope analysis (Yassine, Bengherbia, Benyezza, & Ould, 2022), spectral kurtosis (Tian, Morillo, Azarian, & Pecht, 2015), and wavelet-based feature extraction remain effective for localized defect detection under controlled conditions (Kankar, Sharma, & Harsha, 2011). However, these pipelines depend heavily on expert feature design and labeled data from specific operating regimes. When conditions such as load, speed, sensor placement, or background noise change, their performance often degrades (Zhu, Xu, & Wang, 2023; Asutkar, Singh, & Tiwari, 2023), which limits their effectiveness in real industrial environments.

1.1. Shallow and Deep learning-based methods

With the success of deep learning, deep learning–based methods for bearing fault diagnosis have emerged, reducing reliance on manual feature design. Convolutional neural networks (CNNs) (Zhang, Li, Ding, & Li, 2019) , recurrent neural networks (RNNs) (Shao, Jiang, Lin, & Li, 2017), autoencoders (Chen, Li, & Sanchez, 2017), and more recently transformers (Chen, Li, & Sanchez, 2020) and graph neural networks (Zhang, Li, Ding, & Li, 2022) have achieved state-of-the-art performance on benchmark datasets like CWRU and IMS. Transfer learning has also been applied to reuse pretrained models across related domains (Lee, Kim, & Park,

2023). Nevertheless, most deep learning approaches implicitly assume that training and deployment data are drawn from the same distribution and that large amounts of labeled data are available. In reality, when domain shifts occur—due to variable speeds, loads, or different test rigs—the accuracy of these models drops significantly. Collecting extensive labeled data across all conditions is impractical in most industrial cases (Zhu et al., 2023).

1.2. Cross-domain adaptation

To address this, researchers have turned to cross-domain fault diagnosis. Domain adaptation methods reduce the discrepancy between source and target domains by minimizing measures such as maximum mean discrepancy (MMD) (Long, Cao, Wang, & Jordan, 2015) or CORAL (Sun & Saenko, 2016), or by using adversarial approaches like gradient reversal networks (Ganin & Lempitsky, 2016). Recent work has proposed multi-adversarial (Pei, Cao, Long, & Wang, 2018) and triple-classifier designs to handle complex shifts (Shao, Xia, Wan, Zhang, & Li, 2021). While these methods enhance robustness to domain shifts, they typically assume access to large amounts of unlabeled target-domain data and are limited to closed-set label spaces. As such, they cannot easily adapt to unseen fault types or severities, which are common in realistic industrial scenarios. Moreover, many existing pipelines transform signals into 2D time–frequency images (e.g., STFT scalograms or wavelet spectrograms), which incur computational overhead and may lose important temporal information (Wang et al., 2019).

1.3. Few-shot learning in fault diagnosis

In parallel, few-shot learning (FSL) has gained traction as a strategy for learning from scarce labeled data. Architectures such as Siamese networks (Koch, Zemel, & Salakhutdinov, 2015), prototypical networks (Snell, Swersky, & Zemel, 2017), and meta-learning frameworks (Finn, Abbeel, & Levine, 2017; Lee et al., 2023) have demonstrated the ability to generalize to novel classes with only a handful of labeled samples. Recent works have applied few-shot methods to bearing fault diagnosis (Shen, Liu, Zhang, & Zhao, 2023), showing that unseen fault classes can be recognized with limited labels. However, most of these studies focus on single-domain settings or controlled laboratory conditions. Very few explicitly integrate few-shot learning with domain adaptation, which is critical for realistic deployment where unseen fault types occur under varying loads, speeds, and severities.

1.4. Research gap and contributions of this work

From the above review, two main gaps can be seen:

i) Current domain adaptation methods often depend on unlabeled target data and find it difficult to handle new or unseen fault types.

ii) Most few-shot learning studies ignore domain shifts and are tested only within a single domain, which limits their use in real industrial settings.

Therefore, there is a need for a practical framework that (i) works across different domains without requiring many target labels, (ii) can identify unseen fault types from only a few labeled samples, and (iii) learns directly from raw time-series data without relying on expensive time–frequency image transformations.

To address this gap, we propose a Siamese Attention Encoder–based few-shot cross-domain fault diagnosis framework (SAE-FSC). The main contributions are:

- **Time-series Siamese Attention Encoder:** A novel architecture that learns domain-robust embeddings directly from raw vibration signals, avoiding image conversion and preserving temporal discriminative structure.
- **Composite training objective:** A loss function combining supervised cross-entropy, pairwise binary cross-entropy, and domain adversarial loss. This jointly enforces discriminative class separation, intra-class compactness, and domain invariance.
- **Extensive evaluation under cross-domain settings:** Experiments on the CWRU dataset under leave-one-fault-out (LOFO) and leave-two-fault-out (LTFO) protocols demonstrate robust generalization across unseen fault types, load conditions, and fault severities, achieving a prediction accuracy of 87% with 5-shot learning.

This work establishes a novel and practical methodology for time-series few-shot cross-domain fault diagnosis, overcoming the limitations of prior image-based and purely supervised approaches, and moving closer to deployable predictive maintenance solutions in industrial settings.

2. METHODOLOGY

The proposed framework is designed to learn fault-discriminative and domain-invariant representations directly from raw vibration signals, enabling few-shot cross-domain fault diagnosis. The methodology comprises three core components: (1) SAE for feature extraction, (2) a composite training objective for discriminative and domain-robust embedding learning, and (3) a cross-domain few-shot evaluation protocol.

2.1. Siamese Attention Encoder (SAE)

The encoder is responsible for mapping raw vibration segments into a low-dimensional embedding space where samples from the same fault class cluster together, while samples from different classes remain well separated. Formally, for an input time-series segment x, the embedding is defined as:

$$z = f_\theta(x), \quad z \in R^d \tag{1}$$

where $f_\theta(\cdot)$ is the encoder and d is the embedding dimension.

2.2.1 Network Architecture

- The encoder begins with 1-D convolutional layers that capture local temporal patterns in the vibration signal, such as periodic impulses caused by bearing defects.
- Residual connections are used to stabilize training and ensure that both shallow and deep features are preserved.
- The feature maps are then fed into two attention modules: channel attention and temporal attention.

2.2.2 Attention mechanisms

- *Channel attention* learns which sensor channels or feature maps are most informative. For example, drive-end signals may carry stronger fault signatures than fan-end signals under certain conditions. This is achieved by applying global average pooling (GAP) across time, followed by a gating function:

$$\alpha_c = \sigma(W_c \cdot \text{GAP}(Z) + b_c) \tag{2}$$

- *Temporal attention* identifies the most discriminative time segments within the vibration signal. For instance, transient impulses from localized faults are more diagnostic than stationary noise. This is modeled using global max pooling (GMP) across channels:

$$\alpha_t = \sigma(W_t \cdot \text{GMP}(Z) + b_t) \tag{3}$$

- The two attention weights are combined to re-weight the original feature maps:

$$z' = \alpha_c \odot \alpha_t \odot Z \tag{4}$$

where $\sigma(\cdot)$ denotes the sigmoid function, and $\odot$ represents element-wise multiplication.

2.2. Composite Training Objective

To ensure that that the learned embeddings are simultaneously class-discriminative, pairwise-structured, and domain-invariant, a composite objective function is employed:

$$\mathcal{L} = \mathcal{L}_{CE} + \lambda_1 \mathcal{L}_{BCE} + \lambda_2 \mathcal{L}_{DA} \tag{5}$$

where $\mathcal{L}_{CE}$ enforces class-level discrimination, $\mathcal{L}_{BCE}$ enforces intra-class compactness and inter-class separation, and $\mathcal{L}_{DA}$ promotes domain invariance. This ensures the embeddings are aligned with ground-truth fault classes. CE provides the discriminative structure needed to separate different fault types in the embedding space. Parameter weights λ_1, λ_2 control the trade-off between discriminative embedding learning, pairwise similarity preservation and domain invariance. Empirically, grid-search experiments over λ_1, λ_2 indicate that moderate weighting $\lambda_1 = 0.5, \lambda_2 = 0.1$ achieves a good balance between source accuracy and cross-domain generalization, demonstrating the robustness of the normalized composite loss formulation.

- *(a) Cross-Entropy Loss (CE)* - the cross-entropy term is applied to source-domain samples with known labels:

$$\mathcal{L}_{CE} = -\frac{1}{N} \sum_{i=1}^{N} y_i \log \hat{y}_i \tag{6}$$

- *(b) Binary Cross-Entropy Loss (BCE)* - to explicitly model pairwise relationships, we form siamese pairs of samples and apply binary cross-entropy on their similarity scores:

$$s = \frac{z_1 \cdot z_2}{\|z_1\| \|z_2\|} \tag{7}$$

$$\mathcal{L}_{BCE} = -\big[y \log s + (1-y) \log(1-s)\big] \tag{8}$$

where $y = 1$ if samples belong to the same class and $y = 0$ otherwise. This helps the model remain robust when encountering unseen classed with limited support samples.

- *(c) Domain-Adversarial Loss (DA)* - domain invariance is achieved using a domain discriminator with gradient reversal:

$$\mathcal{L}_{DA} = -\frac{1}{N} \sum_{i=1}^{N} d_i \log \hat{d}_i \tag{9}$$

where $d_i \in \{0, 1\}$ indicates whether a sample belongs to the source or target domain. The gradient reversal layer forces the encoder to produce embeddings that confuse the discriminator, i.e., embeddings that cannot reveal their domain of origin. This encourages the learned representations to be independent of operating conditions and instead focus on fault-specific characteristics.

2.3. Cross-Domain Few-Shot Evaluation Protocol

To validate cross-domain generalization, the framework is evaluated under leave-one-fault-out (LOFO) and leave-two-fault-out (LTFO) settings:

- *LOFO:* the model is trained on all but one fault type and tested on the held-out fault.
- *LTFO:* the model is trained on all but two fault types, with two unseen faults used for testing.

For each evaluation, only a small number of labeled support samples (k-shot, with k = 5, 7, 10, 20) are provided from the unseen classes. The encoder must classify query samples by

leveraging embeddings and the structured space learned via the composite objective.

3. RESULTS AND DISCUSSION

3.1. Dataset and Preprocessing

The proposed framework is evaluated on the CWRU bearing dataset, a widely used benchmark for fault diagnosis research. The dataset was collected on a 2-hp reliance electric motor equipped with a torque transducer and dynamometer as summarized in Table 1. Vibration data were recorded from accelerometers mounted at the drive end (DE), fan end (FE), and base (BA) of the motor housing.

- *Fault simulation* - Localized defects were introduced on the inner race, outer race, and rolling element using electrical discharge machining (EDM). Fault diameters of 0.007, 0.014, 0.021, and 0.028 inches were used to represent different severities.
- *Operating conditions* - The motor was tested under four load levels (0, 1, 2, and 3 hp), corresponding to speeds of 1797, 1772, 1750, and 1730 RPM, respectively. Signals were sampled at 12 kHz depending on the configuration.
- *Preprocessing* - Raw vibration signals were segmented into windows of 1024 samples with 50% overlap. Each segment was normalized to zero mean and unit variance. Labels were assigned based on fault type, fault severity, and load condition. To evaluate cross-domain generalization, we adopted the leave-one-fault-out (LOFO) and leave-two-fault-out (LTFO) protocols, where the model was trained on all but one or two fault types, respectively, and tested on the held-out faults.

3.2. Cross-Domain Representation Analysis under Different Loss Objectives

The prediction results for cross domain analysis under different loss objectives for 5 shots are shown in Fig.1. It is to be noted that LOFO and LTFO analysis were conducted for all shot settings (5,7,10 and 20) and across all fault types and loss objectiveness. For brevity, only the case of inner race faults is shown in the manuscript as representative example.

The t-SNE embeddings shown in Fig.1 reveal how different training objectives affect cross-domain generalization when the model is tested on inner race faults unseen during training.

- *CE only* - The embeddings show partial clustering of fault classes, but significant fragmentation remains between domains. Samples from the same fault type under different loads are dispersed, indicating that CE alone enforces class separation but does not align cross-domain distributions effectively.
- *CE + BCE* - The addition of binary cross-entropy tightens intra-class clusters, making same-fault samples more compact. However, inter-domain misalignment persists, as smaller fault severities remain closer to the normal class, and embeddings from different loads are not fully aligned.
- *CE + DA* - Incorporating domain-adversarial loss reduces the discrepancy between source and target domains. Embeddings from different load conditions overlap more closely, showing improved domain alignment. However, this comes at the expense of class discrimination—clusters of inner and outer race faults overlap, making fault types less distinguishable.
- *CE + BCE + DA (composite)* - The proposed composite loss produces the most balanced embedding space. Fault classes form compact and distinct clusters, while domain effects due to varying loads are minimized. Compared to single-loss models, the composite approach simultaneously preserves fault separability and enforces cross-domain invariance, which is critical for reliable few-shot generalization.

3.3. Cross-Domain Representation Analysis under Different Shot Counts

This analysis was carried out for all fault types, but for brevity, only results for the inner race fault are presented here. The t-SNE plots illustrate how the learned embeddings evolve as the number of support shots increases (5, 7, 10, and 20), providing insight into the effect of data availability on cross-domain generalization.

- *5-shot* - The embeddings form recognizable clusters, but boundaries between fault classes remain fuzzy as shown in Figure. 1(d). Some overlap is observed between outer race and normal samples, especially under varying load conditions. This reflects the challenge of achieving both class discrimination and domain alignment with very limited labeled support data.
- *7-shot* - Increasing the number of support examples strengthens intra-class compactness and improves separation between fault clusters as shown in Figure. 2(d). Domain-induced variations (e.g., load effects) are reduced compared to the 5-shot case, but smaller fault severities are still more prone to misalignment.
- *10-shot* - Fault classes become more distinctly separated, and cross-domain discrepancies are further minimized as shown in Figure. 2(b). The embeddings demonstrate improved robustness, with tighter clustering of minor fault severities and clearer separation from the normal class. This indicates that the model benefits significantly from the additional support samples.
- *20-shot* - The embedding space achieves the most stable and structured form, with compact clusters that are both well separated across fault types and aligned across domains as shown in Figure. 2(c).

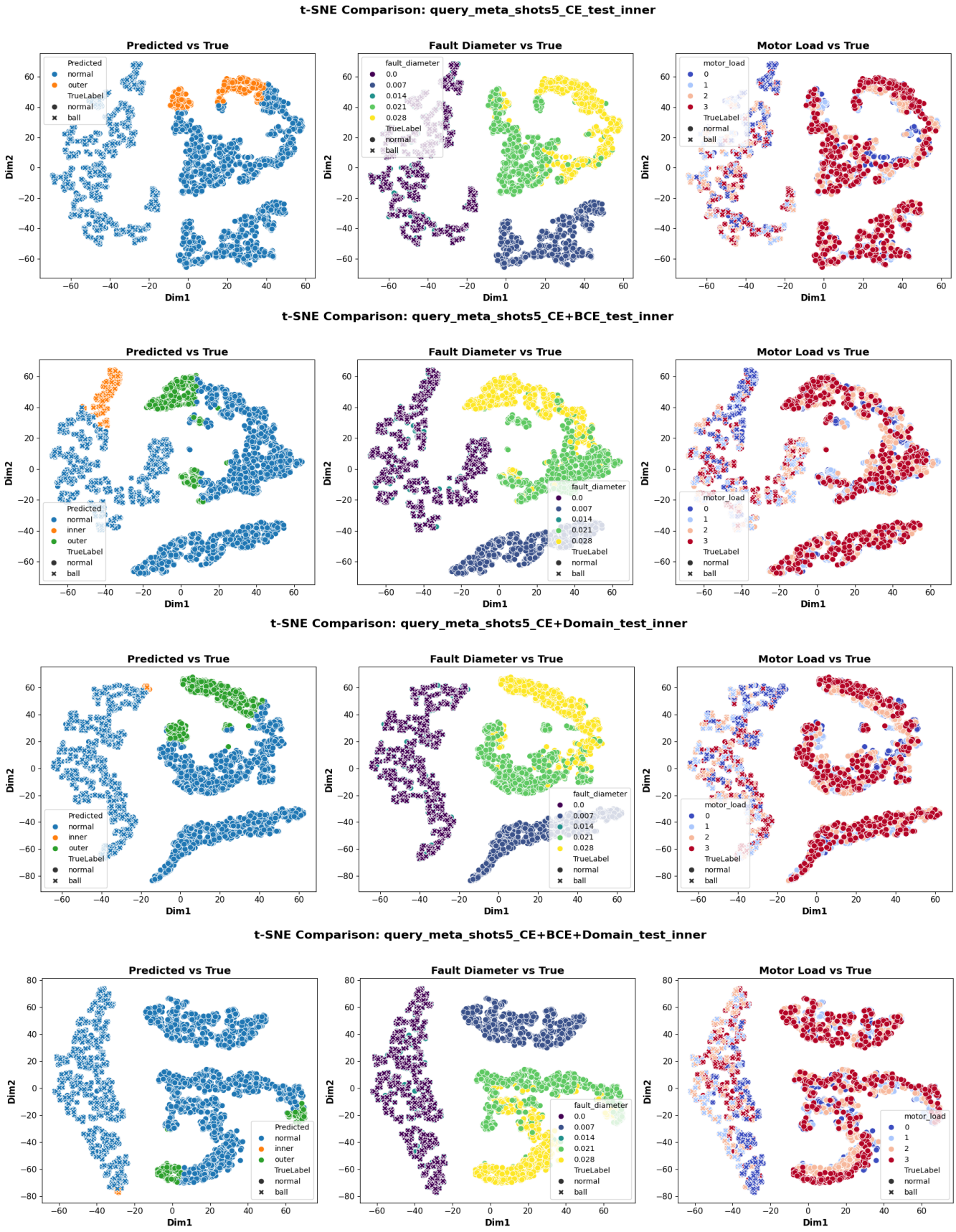

Figure 1. t-SNE visualizations of cross-domain feature representations under different loss functions for the 5-shot setting. Each panel shows clustering with respect to predicted vs. true labels, fault diameter, and motor load. Rows correspond to different loss objectives (e.g., CE, CE+BCE, CE+BCE+DA), highlighting how the choice of objective impacts separation, compactness, and domain invariance.

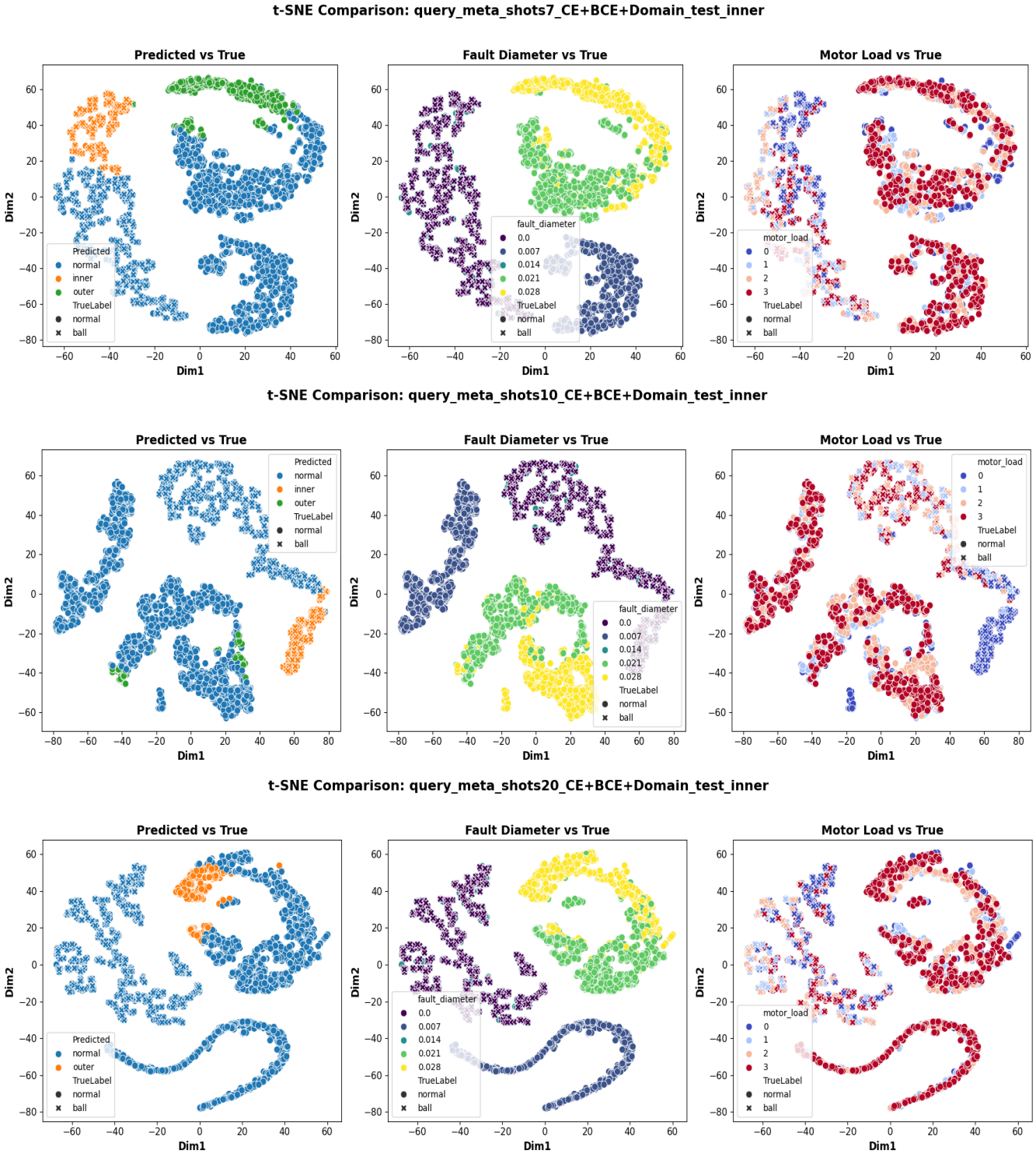

Figure 2. t-SNE visualizations of cross-domain feature representations under different shot counts (5, 7, 10, and 20). Each panel shows clustering with respect to predicted vs. true labels, fault diameter, and motor load. Increasing the number of shots improves intra-class compactness and inter-class separation, while also reducing domain-induced variations such as load effects. The comparison highlights how additional labeled support samples enhance both discrimination and domain invariance.

Load variations exert minimal influence, suggesting that the model has effectively learned fault-discriminative and domain-invariant features. Overall, these results show that increasing shot count consistently improves cross-domain representation quality. Even though the framework demonstrates reasonable clustering in extremely low-shot conditions (e.g., 5-shot), higher support counts (10 and 20) yield more reliable fault separation and domain invariance, strengthening classification performance under unseen fault conditions.

3.4. Accuracy Comparison across Shot Counts with Composite Loss

The prediction accuracy results across different shot counts as shown in Figure. 3 clearly demonstrate the strength of the composite loss in enabling robust cross-domain generalization. While low-shot settings such as 5-shot yield moderate accuracies (70–82%), the framework shows steady improvement with increasing shots, consistently surpassing 85% at 20-shot and reaching above 90% in several train→test fault transfer cases. Importantly, the composite loss balances fault discrimination and domain alignment, ensuring that representations remain stable across varying fault severities and load conditions. Even in the most challenging scenario—generalizing to unseen inner race faults—the model maintains competitive performance, highlighting its ability to transfer knowledge effectively across domains. These findings confirm that the composite loss function is particularly well-suited for cross-domain fault diagnosis with limited labeled data, making it a practical choice for real-world industrial applications. The expanded summary of cross-domain study is listed in Table 1.

4. CONCLUSION

In this work, we proposed a SAE-FSC framework for rolling element bearings. By integrating a composite training objective that combines cross-entropy, binary cross-entropy, and domain-adversarial losses, the framework learns representations that are simultaneously fault-discriminative and domain-invariant. Extensive experiments on the CWRU dataset under leave-one-fault-out (LOFO) and leave-two-fault-out (LTFO) protocols demonstrated that the method generalizes effectively to unseen fault types, severities, and load conditions. The t-SNE visualizations confirmed the ability of the composite loss to achieve compact clustering and cross-domain alignment, while the accuracy results showed steady improvements with increasing shot counts, achieving over 90% prediction accuracy at 20-shot. These findings establish the composite loss as a robust approach for cross-domain fault diagnosis, particularly in settings where labeled target data are scarce.

For future work, several directions are envisioned. First, the framework can be extended to other rotating machinery datasets (e.g., IMS, PRONOSTIA) to validate generalization across different rigs and sensor configurations. Second, multi-modal sensor fusion (vibration, current, and acoustic signals) can be explored to enhance robustness under noisy industrial environments. Third, incorporating temporal adaptation strategies such as recurrent or transformer-based modules could capture evolving degradation patterns beyond static fault detection. Finally, integration with real-time deployment pipelines will be pursued, ensuring that few-shot cross-domain learning can be translated into practical predictive maintenance solutions in industrial settings.

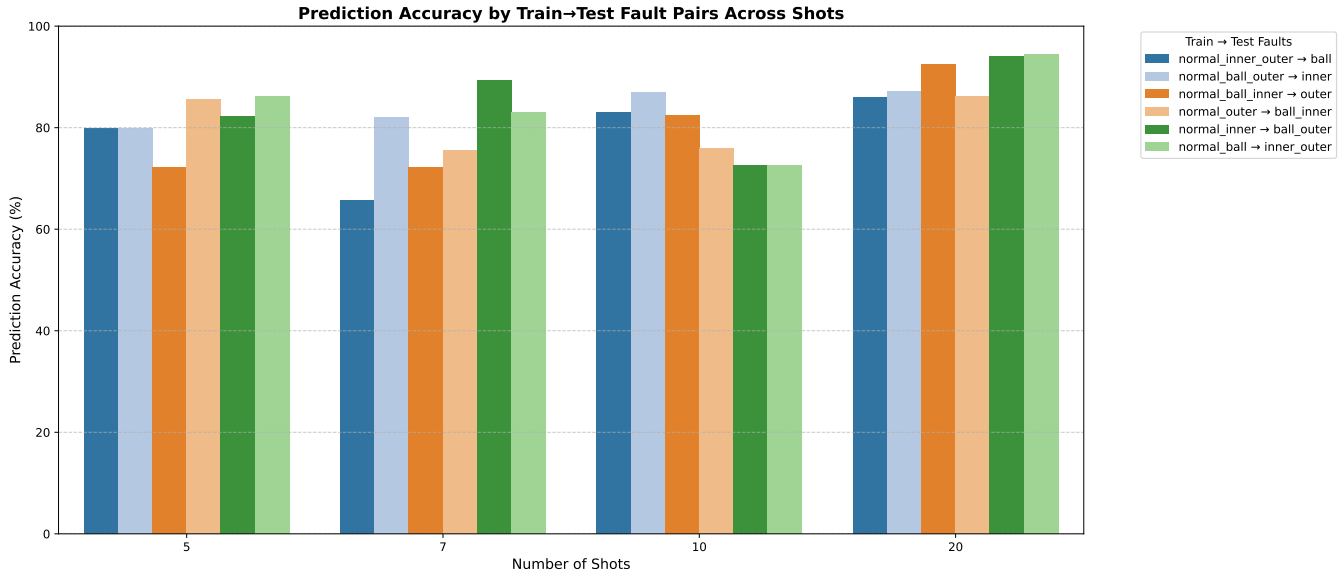

Figure 3. Prediction accuracy across different shot settings for train→test fault pairs. Bars represent accuracy values grouped by the number of shots.

Table 1. Expanded summary of cross-domain ablation and t-SNE metrics across shot settings.

Shots	Loss	Train Faults	Test Faults	Acc(%)	F1	Silhouette Coefficient	Davies-Bouldin Index
5	CE	normal+ball+inner	outer	67.396	0.326	0.105	2.127
5	CE	normal+ball+outer	inner	78.901	0.211	0.140	1.609
5	CE	normal+inner+outer	ball	82.755	0.172	0.101	1.790
5	CE+BCE	normal+ball+inner	outer	77.305	0.227	0.103	2.199
5	CE+BCE	normal+ball+outer	inner	69.422	0.306	0.109	1.637
5	CE+BCE	normal+inner+outer	ball	92.543	0.075	0.160	1.968
5	CE+Domain	normal+ball+inner	outer	77.380	0.226	0.101	2.331
5	CE+Domain	normal+ball+outer	inner	97.875	0.021	0.130	1.718
5	CE+Domain	normal+inner+outer	ball	83.192	0.168	0.142	1.738
5	CE+BCE+Domain	normal+ball+outer	inner	98.713	0.019	0.163	1.744
5	CE+BCE+Domain	normal+ball+inner	outer	98.879	0.018	0.158	1.725
5	CE+BCE+Domain	normal+inner+outer	ball	92.891	0.071	0.162	1.876
7	CE	normal+ball+inner	inner	78.660	0.213	0.107	1.681
7	CE	normal+ball+outer	ball	73.494	0.265	0.145	1.806
7	CE	normal+ball+outer	outer	67.365	0.324	0.101	2.098
7	CE+BCE	normal+ball+inner	inner	90.322	0.518	0.145	1.734
7	CE+BCE	normal+ball+outer	outer	67.565	0.326	0.132	2.099
7	CE+BCE	normal+inner+ball	ball	97.692	0.305	0.161	1.647
7	CE+Domain	normal+ball+inner	inner	89.003	0.109	0.141	1.834
7	CE+Domain	normal+ball+outer	outer	98.812	0.117	0.143	1.944
7	CE+Domain	normal+inner+ball	ball	92.566	0.074	0.141	1.797
7	CE+BCE+Domain	normal+ball+inner	inner	98.824	0.012	0.126	1.767
7	CE+BCE+Domain	normal+ball+outer	outer	98.683	0.016	0.127	1.776
7	CE+BCE+Domain	normal+inner+ball	ball	98.611	0.018	0.130	1.779
10	CE	normal+ball+inner	inner	59.392	0.433	0.423	1.590
10	CE	normal+ball+outer	outer	66.769	0.333	0.667	2.463
10	CE	normal+inner+ball	ball	59.324	0.432	0.422	1.580
10	CE+BCE	normal+ball+inner	inner	58.719	0.421	0.587	1.741
10	CE+BCE	normal+ball+outer	outer	58.719	0.421	0.587	1.741
10	CE+BCE	normal+inner+ball	ball	57.886	0.429	0.622	1.774
10	CE+Domain	normal+ball+inner	inner	77.153	0.224	0.775	1.440
10	CE+Domain	normal+ball+outer	outer	79.122	0.220	0.779	1.463
10	CE+Domain	normal+inner+ball	ball	92.539	0.176	0.925	1.556
10	CE+BCE+Domain	normal+ball+inner	inner	92.792	0.228	0.792	1.771
10	CE+BCE+Domain	normal+ball+outer	outer	92.699	0.227	0.791	1.769
10	CE+BCE+Domain	normal+inner+ball	ball	92.607	0.229	0.792	1.768
20	CE	normal+ball+inner	inner	77.754	0.295	0.775	1.707
20	CE	normal+ball+outer	outer	78.675	0.274	0.787	1.584
20	CE	normal+inner+ball	ball	79.091	0.209	0.742	1.836
20	CE+BCE	normal+ball+inner	inner	79.045	0.227	0.790	1.443
20	CE+BCE	normal+ball+outer	outer	79.045	0.227	0.790	1.443
20	CE+BCE	normal+inner+ball	ball	77.853	0.215	0.743	1.754
20	CE+Domain	normal+ball+inner	inner	92.759	0.295	0.933	1.721
20	CE+Domain	normal+ball+outer	outer	92.568	0.219	0.915	1.765
20	CE+Domain	normal+inner+ball	ball	92.525	0.219	0.913	1.753
20	CE+BCE+Domain	normal+ball+inner	inner	83.337	0.198	0.933	1.904
20	CE+BCE+Domain	normal+ball+outer	outer	83.411	0.212	0.834	1.413
20	CE+BCE+Domain	normal+inner+ball	ball	93.610	0.064	0.138	1.639

ACKNOWLEDGMENT

This research is supported by A*STAR under its RIE2025 Manufacturing, Trade and Connectivity (MTC) Industry Alignment Fund- Pre-Positioning (IAF-PP) (Award M23L4a0001).

REFERENCES

Asutkar, S., Singh, V., & Tiwari, R. (2023). Challenges and perspectives of deep learning-based fault diagnosis under variable working conditions. *IEEE Access*, *11*, 45621–45635.

Chen, Z., Li, C., & Sanchez, V. (2017). Multisensor feature fusion for bearing fault diagnosis using sparse autoencoder and deep belief network. *IEEE Transactions on Instrumentation and Measurement*, *66*(7), 1693–1702.

Chen, Z., Li, C., & Sanchez, V. (2020). Deep neural networks for vibration-based fault diagnosis of rotating machinery: A review. *Mechanical Systems and Signal Processing*, *133*, 106983.

Finn, C., Abbeel, P., & Levine, S. (2017). Model-agnostic meta-learning for fast adaptation of deep networks. In *International conference on machine learning (icml)* (pp. 1126–1135).

Ganin, Y., & Lempitsky, V. (2016). Domain-adversarial training of neural networks. In *Advances in neural information processing systems (neurips)* (Vol. 29, pp. 2096–2104).

Jardine, A. K., Lin, D., & Banjevic, D. (2006). A review on machinery diagnostics and prognostics implementing condition-based maintenance. *Mechanical Systems and Signal Processing*, *20*(7), 1483–1510.

Kankar, P. K., Sharma, S. C., & Harsha, S. P. (2011). Rolling element bearing fault diagnosis using wavelet transform. *Neurocomputing*, *74*(10), 1638–1645.

Koch, G., Zemel, R., & Salakhutdinov, R. (2015). Siamese neural networks for one-shot image recognition. In *Icml deep learning workshop.*

Lee, H., Kim, J., & Park, S. (2023). Meta-learning for cross-domain fault diagnosis with limited labeled data. *Mechanical Systems and Signal Processing*, *185*, 109729.

Long, M., Cao, Y., Wang, J., & Jordan, M. I. (2015). Learning transferable features with deep adaptation networks. In *Proceedings of the 32nd international conference on machine learning (icml)* (pp. 97–105).

Ma, J., Yu, S., & Cheng, W. (2022). Composite fault diagnosis of rolling bearing based on chaotic honey badger algorithm optimizing vmd and elm. *Machines*, *10*(6), 469.

Mishra, R. K., Choudhary, A., Mohanty, A. R., & Fatima, S. (2021). Multi-domain bearing fault diagnosis using support vector machine. In *2021 ieee 4th international conference on computing, power and communication technologies (gucon)* (pp. 1–6).

Ocak, H., & Loparo, K. A. (2005). Hmm-based fault detection and diagnosis scheme for rolling element bearings. *Journal of Vibration and Acoustics*, *127*(4), 299–306.

Pei, Z., Cao, Z., Long, M., & Wang, J. (2018). Multi-adversarial domain adaptation. In *Aaai conference on artificial intelligence* (Vol. 32).

Randall, R. B., & Antoni, J. (2011). Rolling element bearing diagnostics—a tutorial. *Mechanical Systems and Signal Processing*, *25*(2), 485–520.

Shao, H., Jiang, H., Lin, Y., & Li, X. (2017). A novel deep autoencoder feature learning method for rotating machinery fault diagnosis. *Mechanical Systems and Signal Processing*, *95*, 187–204.

Shao, H., Xia, M., Wan, J., Zhang, W., & Li, X. (2021). A novel multiple source domain adaptation method for intelligent fault diagnosis. *IEEE Transactions on Industrial Electronics*, *68*(5), 4374–4384.

Shen, T., Liu, X., Zhang, J., & Zhao, W. (2023). Few-shot learning-based intelligent fault diagnosis for rotating machinery: A survey. *Mechanical Systems and Signal Processing*, *181*, 109537.

Snell, J., Swersky, K., & Zemel, R. S. (2017). Prototypical networks for few-shot learning. In *Advances in neural information processing systems (neurips)* (Vol. 30, pp. 4077–4087).

Sun, B., & Saenko, K. (2016). Deep coral: Correlation alignment for deep domain adaptation. In *European conference on computer vision (eccv) workshops* (pp. 443–450). Springer.

Tian, J., Morillo, C., Azarian, M. H., & Pecht, M. (2015). Motor bearing fault detection using spectral kurtosis-based feature extraction coupled with k-nearest neighbor distance analysis. *IEEE Transactions on Industrial Electronics*, *63*(3), 1793–1803.

Wang, Y., Peter, W. T., Tang, B., Qin, Y., Deng, L., Huang, T., & Xu, G. (2019). Order spectrogram visualization for rolling bearing fault detection under speed variation conditions. *Mechanical Systems and Signal Processing*, *122*, 580–596.

Yassine, T., Bengherbia, B., Benyezza, H., & Ould, Z. M. (2022). Bearing faults diagnosis using envelope analysis and 1d convolutional neural network. *AINTELIA Science Notes Journal*, *1*(1).

Zhang, W., Li, C., Ding, X., & Li, Z. (2019). Few-shot bearing fault diagnosis with siamese networks. *IEEE Transactions on Instrumentation and Measurement*, *68*(6), 1503–1512.

Zhang, W., Li, C., Ding, X., & Li, Z. (2022). Graph convolutional networks for intelligent fault diagnosis of machines with limited data. *IEEE Transactions on Industrial Electronics*, *69*(1), 139–149.

Zhu, K., Xu, Y., & Wang, X. (2023). Domain generalization in intelligent fault diagnosis: A comprehensive re-

view. *Mechanical Systems and Signal Processing*, *187*, 109947.