Letters

Detecting Low-Intensity Cyber-Attacks in Electric Drives via Contrastive Few-Shot Learning

He Yang, *Student Member, IEEE*, and Jin Ye, *Senior Member, IEEE*

***Abstract*—This article proposes a novel anomaly detection framework for intelligent motor drives. The framework integrates contrastive learning (CL) and transfer learning (TL) into a convolutional neural network (CNN)-based architecture. The CL module enhances the discriminative capability of feature embeddings. It groups samples from identical operational states and enforces separation between samples collected under different conditions. This allows the model to effectively distinguish between normal and anomalous states. Meanwhile, TL enables the effective transfer of knowledge from abundant simulation data to limited real-world experimental data, improving generalization. Experiments on a motor drive testbed demonstrate that our method achieves 99.8% accuracy using the full training set and 99.69% accuracy with only 40% of the data—surpassing the TL-based model (99.0% and 89.0%,). In a challenging few-shot setting (12.5% of training data), it still maintains 91.6% accuracy, compared to 86.5% for TL. Moreover, our method yields a lower false alarm rate (2.1% versus 4.5% for TL), showing superior robustness and data efficiency in real-world deployment scenarios.**

***Index Terms*—Contrastive learning (CL), cyber-attacks, fault detection, intelligent motor drives.**

I. Introduction

INTELLIGENT motor drives are increasingly central to cyber-physical systems (CPSs), including industrial automation, smart grids, and electric vehicles [1]. However, their connectivity makes them vulnerable to cyber-attacks [2], which may cause operational disruptions or failure. Detecting such attacks—especially subtle ones—has become critical.

Convolutional neural network (CNN)-based supervised learning methods are widely used for fault detection in motor drives due to their strong feature extraction capabilities [3], [4]. Yet, these models struggle to distinguish low-intensity anomalies that closely resemble normal states. Such anomalies often manifest as small perturbations in current or voltage signals, making them difficult to detect with conventional classification.

Contrastive learning (CL), initially developed for computer vision [5], offers a promising solution by learning representations that bring similar samples closer and push dissimilar ones apart. However, its application to time-series data in CPS is nontrivial, and existing CNN approaches still degrade under data scarcity [3], [4].

Recent CL-based fault detection efforts face various limitations. Zhang et al. [6] required complex augmentations and dense labels. However, the method in [6] requires complex data augmentation and dense labeling, which limits its scalability in real-world motor drive systems. Similarly, Miao et al. [7] relied solely on voltage ripple features and required fully labeled datasets, which limit its applicability and reduces robustness against subtle or multimodal attacks. Chen et al. [8] needed severe faults and real-world supervision. Beyond these works, time-series CL tailored for control and industrial signals has also been explored; for example, a debiased contrastive formulation improves time-series representations and fault diagnosis performance in industrial settings [9].

Thus, a generalizable and efficient CL-based framework that works with limited data remains a challenge. To address this, we propose a CL-enhanced CNN architecture that integrates with existing testbeds without structural changes, leveraging a simulated dataset generated from a validated high-fidelity motor drive model [3] alongside limited real-world measurements. Experiments show our method improves detection performance over baseline CNN [3], using far less real-world data.

We propose a CL framework for low-intensity anomaly detection in motor drives, achieving high accuracy and robustness under limited data, with seamless integration into existing testbed systems.

The rest of this article is organized as follows. Section II reviews related work and introduces our method. Section III describes the experimental setup and data. Section IV presents results and analysis. Finally, Section V concludes this article.

II. Proposed CL-Enhanced CNN Methodology

We first apply transfer learning (TL) to leverage large-scale simulation data, followed by CL fine-tuning on limited experimental samples to improve sensitivity to subtle anomalies.

Received 1 July 2025; revised 20 August 2025; accepted 14 September 2025. Date of publication 22 September 2025; date of current version 13 November 2025. This work was supported in part by the U.S. National Science Foundation TI-PFI under Grant #2414706, in part by the DMS-AMPS under Grant #2318809, in part by the ECCS-EPCN under Grant #2102032, and in part by the NSF-SATC under Grant #2019311. *(Corresponding author: Jin Ye.)*

The authors are with the School of Electrical and Computer Engineering, University of Georgia, Athens, GA 30602 USA (e-mail: hy39566@uga.edu; jin.ye@uga.edu).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TPEL.2025.3613220.

Digital Object Identifier 10.1109/TPEL.2025.3613220

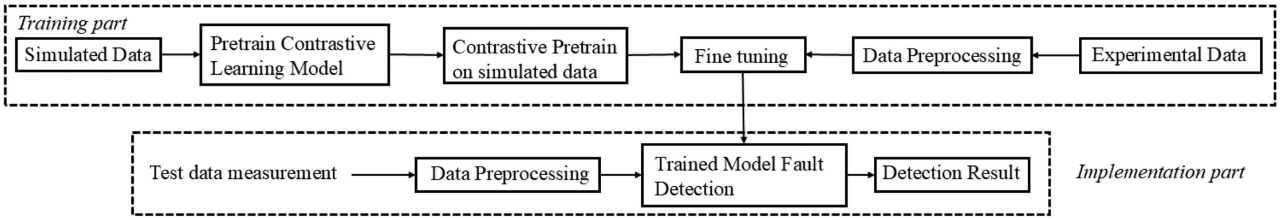

Fig. 1. Flowchart of the proposed CL-based detection framework. As shown, the architecture enables real-time deployability without hardware modifications, while ensuring consistent preprocessing for both simulated and experimental datasets.

Fig. 1 illustrates the complete workflow, covering signal acquisition, fast Fourier transform (FFT)-based preprocessing, CNN model training, and deployment. The architecture leverages TL to pretrain on abundant simulation data, followed by CL fine-tuning on limited experimental samples. This two-stage design enables accurate detection of low-intensity cyber-attacks while preserving real-time deployability and compatibility with existing testbed setups—without requiring additional hardware or modifications. By enforcing compact intraclass and well-separated interclass embeddings, the framework enhances feature robustness, making it both effective in controlled experiments and practical for industrial deployment.

The proposed method retains the original CNN structure but incorporates CL to enforce compact intraclass and distinct interclass embeddings, improving feature robustness.

Training follows a two-stage process. First, the CNN is pretrained on simulation data using binary cross-entropy loss

$$\mathcal{L}_{\text{cls}} = -\sum_{i} \left[y_i \log \hat{y}_i + (1 - y_i) \log(1 - \hat{y}_i) \right]. \tag{1}$$

Then, the model is fine-tuned on a mixed set of simulation and experimental data using a combined loss

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{cls}} + \lambda \cdot \mathcal{L}_{\text{cont}} \tag{2}$$

with contrastive weight $\lambda = 0.5$. The contrastive loss weight λ was set to 0.5 to balance the objectives, a value determined to be optimal through empirical tuning.

A. CL Framework

In CL, the network is trained with sample pairs—positive (same class) or negative (different class)—to learn discriminative embeddings. The contrastive loss is

$$\mathcal{L}_{\text{cont}}(x_i, x_j) = -\log \frac{\exp(\text{sim}(z_i, z_j)/\tau)}{\sum_k \exp(\text{sim}(z_i, z_k)/\tau)} \tag{3}$$

where x_i and x_j are inputs, z_i and z_j are their embeddings, $\text{sim}(\cdot)$ is cosine similarity, and $\tau = 0.1$ is the temperature parameter, which is a scaling factor applied to the cosine similarity before the softmax operation, controlling the smoothness of the similarity distribution. This loss function optimizes the feature embedding space by minimizing intraclass variance (bringing embeddings of samples from the same class closer together) and maximizing interclass separation (increasing the distance between embeddings of samples from different classes). This dual effect promotes compact clusters for identical classes and clear margins between distinct classes, thereby enhancing the discriminative capability of the learned representations.

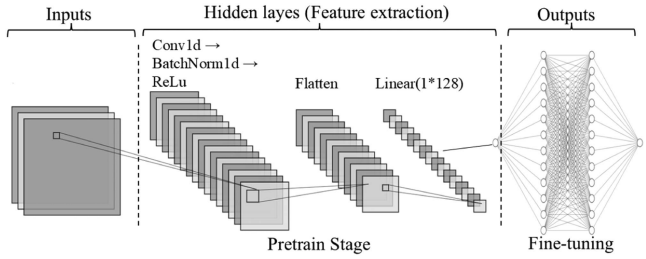

Fig. 2. Algorithm structure.

TABLE I
ALGORITHM ARCHITECTURE

Layer	Depth	Height	Width
Input	3	1	150
Conv1d (32 filters, kernel size = 3)	32	1	150
BatchNorm1d	32	1	150
ReLU	32	1	150
Conv1d (64 filters, kernel size = 3)	64	1	150
BatchNorm1d	64	1	150
ReLU	64	1	150
AdaptiveMaxPool1d (output size = 50)	64	1	50
Flatten	1	1	3200
Linear (128 neurons)	128	1	1
Classification Head			
Linear (32 neurons)	32	1	1
Linear (1 neuron, binary output)	1	1	1
Projection Head			
Linear (64 neurons)	64	1	1
Linear (64 neurons, embedding)	64	1	1

The CNN is first trained on simulated data, with layers for feature extraction and binary classification. It is then fine-tuned using limited experimental data and CL, which encourages the model to distinguish nuanced patterns. Fig. 2 and Table I summarize the architecture.

This integration of simulated and experimental data improves generalization while maintaining accuracy. The result is a robust, deployable solution for subtle attack detection in electric drives. Unlike conventional fine-tuning approaches that rely solely on classification loss, our method jointly optimizes classification and contrastive objectives, explicitly enhancing intraclass compactness and interclass separability. Moreover, it achieves high performance with mixed simulation and limited real-world data without requiring additional hardware or testbed modifications, making it both novel and deployment-ready.

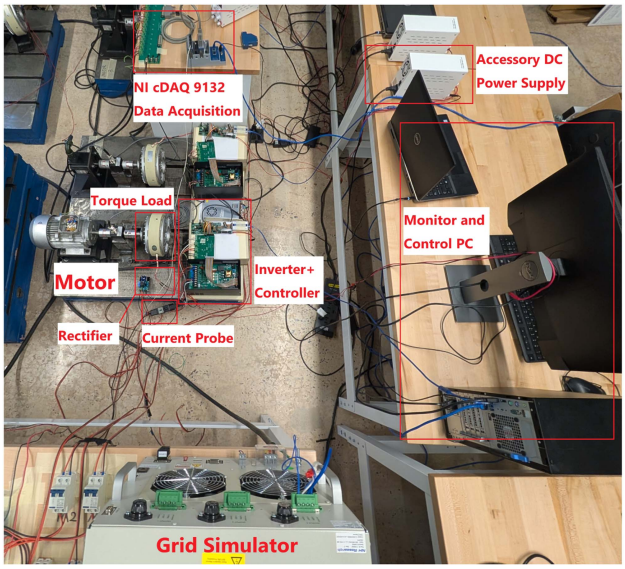

Fig. 3. Motor drive testbed used for data acquisition and cyber-attack emulation.

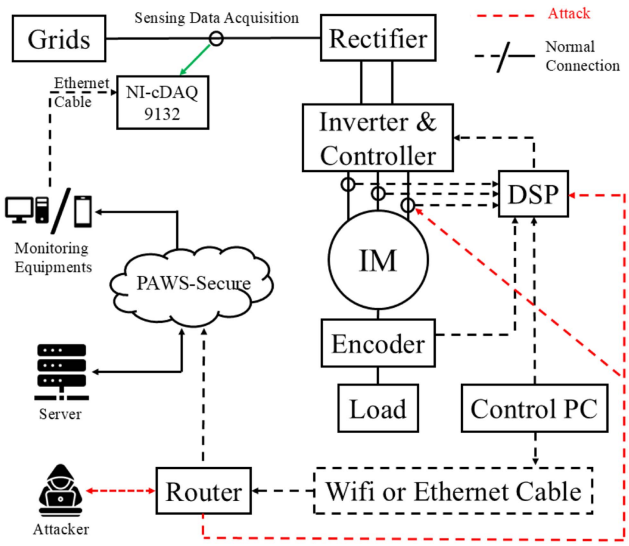

Fig. 4. System topology with attack points.

III. Experimental Setup

We built a prototype motor drive system with a grid simulator, induction motor, and digital signal processor (DSP) controller connected to a PC via Ethernet. This testbed supports safe, repeatable cyber-attack emulation using embedded backdoor routines.

Attacks include false data injection (FDI), replay, and low-level offsets. Data—including three-phase and dc-link currents—were sampled at 10 kHz. Experimental signals contain realistic disturbances such as noise, ripple, and jitter.

Only a small number of experimental samples were needed, as the model was pretrained on diverse simulation data. These real samples fine-tune the model to deployment conditions and improve real-world performance.

A. Testbed Configuration

To validate our CL-enhanced CNN, we developed a hybrid testbed combining simulation and physical hardware. The setup includes a three-phase induction motor, grid simulator, embedded sensors, and digital controller, forming a realistic environment for cyber-attack emulation.

The system continuously monitors three-phase currents and dc-link voltage, which are used as detection features. Network interfaces enable controlled injection of attacks. The overall motor drive testbed used for data acquisition and cyber-attack emulation is shown in Fig. 3.

B. Attack Injection Scenarios

To simulate realistic threats, we implemented low-intensity attacks that introduce subtle disturbances while avoiding system failure.

The attack has two following characteristics:

1) *False data injection (FDI):* It alters sensor values to mislead the controller.
2) *Offset-based attacks:* It applies persistent small-value disturbances to evade detection. Each is precisely triggered to replicate hard-to-detect conditions, motivating the use of CL.

These characteristics are particularly challenging for legacy detection systems because they intentionally operate at low magnitudes, producing signal perturbations that closely resemble normal operational noise. Traditional threshold-based or model-based methods often fail to detect such anomalies due to their limited sensitivity and reliance on coarse statistical features. By targeting subtle deviations—such as persistent offsets in sensor readings or minimally altered feedback signals—these attacks can bypass existing protection schemes without triggering alarms, making them ideal for evaluating the performance of advanced detection methods like our CL-enhanced CNN.

They are representative of low-strength cyber-attacks in motor drives, as they mimic realistic stealthy intrusion scenarios where attackers aim to remain undetected for extended periods while gradually degrading system performance.

The system topology with potential attack points is illustrated in Fig. 4.

C. Data Acquisition, Preprocessing, and Dataset Summary

Signals are sampled from simulation and hardware at 10 kHz. Time-series data are segmented into 500-point (50 ms) windows. To capture frequency-domain patterns from attacks, each segment is transformed via FFT. As shown in Fig. 5, attack conditions yield subtle spectral deviations, particularly elevated second-order harmonics. Traditional fault detection methods primarily focus on electrical, mechanical, and temperature-related features. In contrast, the subtle and persistent attacks considered in this work highlight the necessity of employing CL to enhance detection sensitivity.

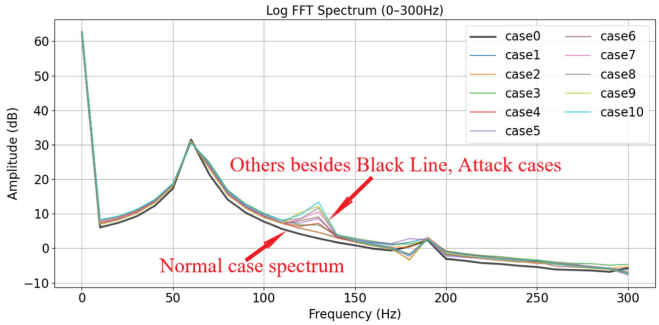

Fig. 5. FFT spectrum for normal and attack cases.

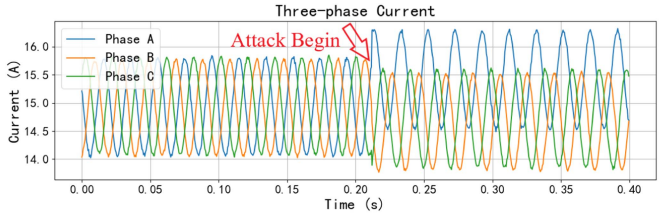

Fig. 6. Three-phase current waveforms under attack.

We retain the first 150 frequency bins from each of the three channels, forming a 3×150 input matrix per sample. All inputs are standardized. Fig. 6 displays representative current waveforms, showing minor waveform shifts during attack conditions.

We followed the data partitioning in [3], using 2000 labeled samples: 1000 normal and 1000 attack. These samples come from both simulated and experimental runs, each consisting of three-phase current segments of 500 time steps $\times$ 3 channels.

In total, 80% of the data (1600 samples) was used for training, and the remaining 20% (400 samples) for testing—ensuring consistent benchmarking with prior work. The simulated portion comes from a high fidelity simulation model; the experimental portion was acquired from the real testbed under the same conditions. The repeatability and robustness of the hybrid simulation–hardware testbed, including its ability to support varied attack scenarios, are described in detail in our related work [10]. In brief, the hybrid loop integrates a high fidelity model with real hardware components, enabling consistent reproduction of attack conditions and allowing controlled stress testing across a wide range of scenarios. This ensures that the evaluation is both systematic and representative of practical operating environments.

D. Training Procedure: CNN + CL

Training follows two stages. First, a baseline CNN is pretrained on simulated data with binary classification loss. Then, the model is fine-tuned using CL on a mixed dataset to improve generalization.

The CNN includes 1-D convolution layers, global pooling, and projection layers. During contrastive training, two augmented views of each sample pass through the encoder and projection head to generate embeddings, and contrastive loss encourages class-aware separation in the feature space.

TABLE II
COMPARISON OF MODEL ACCURACY UNDER VARYING TRAINING SET SIZES

Training Set Size	CNN Accuracy	TL Accuracy	CL Accuracy
Full (100%)	99.0%	99.6%	**99.8%**
Reduced (40%)	98.9%	99.0%	**99.69%**
Few-Shot (12.5%)	83.0%	89.0%	**91.6%**

CL: Contrastive Learning, TL: Transfer Learning, CNN: Convolutional Neural Network. CL consistently achieves the highest accuracy, even with limited data (40% or 12.5%).

Optimization uses the Adam optimizer [11] (learning rate 10^{-3}, batch size 16) for 20 epochs with gradient clipping and early stopping. Performance is evaluated using accuracy, precision, recall, F1-score, and confusion matrices, which clearly reveal false positives (FP) and false negatives (FN) for each method.

IV. RESULTS AND DISCUSSION

We evaluated all models using a fixed dataset of 2000 labeled windows. The test set remains 400 samples (200 normal and 200 attack), while the training set is subsampled to simulate full, reduced, and few-shot settings. In this article, "few-shot" refers to limited real-world experimental data used to fine-tune a simulator-pretrained CNN (12.5% of the training set); it does not denote meta-learning-style few/one-shot learning [12] for new classes. Few shots (12. 5%) denote fine-tuning with reduced real-world data. In this setting, we include multiple attack cases but merge them under a single "attack" label; for each attack case, only one small training subset is used (consistent with the 1–5 examples-per-label notion), yielding ten attack subsets in total. The number of normal samples is matched to the total number of attack samples.

Our model was designed to optimize the embedding space with greater interclass separability and intraclass compactness. To benchmark performance, we first tested a baseline CNN trained on simulation data and fine-tuned with experimental samples—without CL. While adequate under clear conditions, the baseline CNN struggled to detect subtle, low-intensity anomalies.

The CL-enhanced CNN showed notable accuracy improvements across all settings. By encouraging classwise embedding separation, CL enhanced the model's ability to distinguish normal and attack states—even under subtle disruptions.

As shown in Table II, CL consistently outperformed CNN and TL in all scenarios. With the full dataset (8000 simulated and 800 experimental samples from [3]), CL achieved 99.8% accuracy, exceeding TL (99.6%) and CNN (99.0%). Even with 40% of the data (3200 simulated + 320 experimental), CL reached 99.69%, outperforming both TL (99.0%) and CNN (98.9%).

In this study, attack samples are labeled as the positive class (1) and normal samples as the negative class (0). The model outputs attack probabilities, which are binarized using a 0.5 threshold. Based on the comparison between predicted and ground truth labels, true positives, FP, true negatives, and FN are determined for constructing the confusion matrix and calculating performance metrics.

CL's efficiency is further demonstrated in the few-shot case (12.5% data), achieving 91.6% accuracy, significantly better

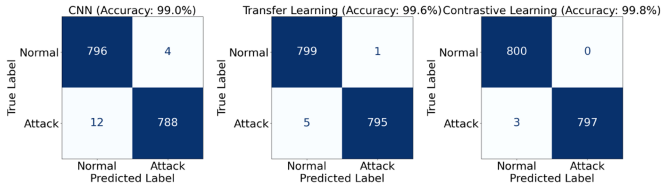

Fig. 7. CL achieves clearer separation of normal and attack cases: Confusion matrices for CNN, TL, and CL.

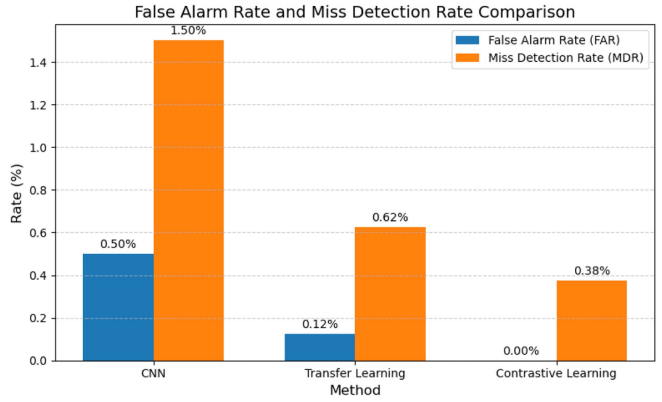

Fig. 8. Comparison of FAR and MDR showing CL's advantage in reducing errors.

than TL (89.0%) and CNN (83.0%). This confirms CL's robustness under low-label conditions. Even with sparse data, our method retained strong generalization, making it ideal for real-world deployments where label scarcity is common.

Confusion matrices in Fig. 7 show that CL reduces FP and FN more effectively than CNN and TL, especially under subtle attack conditions.

Fig. 8 compares false alarm rate (FAR) and miss detection rate (MDR), with CL achieving the lowest error rates across all metrics, reinforcing its superior reliability for stealthy anomaly detection.

The gains of CL stem from improved embedding representation. Unlike standard CNNs, which only optimize classification loss, CL explicitly enforces clustering of similar conditions and separation of dissimilar ones. In addition, by combining simulated and limited real-world data, CL generalizes well across domains, making it suitable for real-world deployment.

V. Conclusion

We proposed a CL-enhanced CNN framework for detecting low-intensity cyber-attacks in intelligent motor drives. By integrating contrastive loss into TL, our model achieves better feature separation and generalization, especially with limited labeled data.

Experiments showed that CL consistently outperforms CNN and TL across various training set sizes, reaching 99.8% accuracy with full data and 99.69% with only 40%. In few-shot settings (12.5% data), CL maintained 91.6% accuracy while significantly lowering both FAR and MDR.

Importantly, these results were achieved without requiring extra hardware or testbed changes, demonstrating that the method is both effective and practical for real-world industrial applications.

Overall, the proposed framework is generalizable and efficient, maintaining strong performance even under severe data scarcity, thus addressing key challenges in real-world anomaly detection. The proposed result delivers high detection accuracy with minimal computational burden, enabling real-time use in safety-critical industrial environments. This capability directly supports the resilience and trustworthiness of next-generation electric drive systems deployed in industrial automation, transportation electrification, and other critical infrastructure sectors.

References

[1] J. Ye et al., "Cyber–physical security of powertrain systems in modern electric vehicles: Vulnerabilities, challenges, and future visions," *IEEE Trans. Emerg. Sel. Topics Power Electron.*, vol. 9, no. 4, pp. 4639–4657, Aug. 2021, doi: 10.1109/JESTPE.2020.3045667.

[2] L. Guo, J. Ye, and B. Yang, "Cyberattack detection for electric vehicles using physics-guided machine learning," *IEEE Trans. Transport. Electrific.*, vol. 7, no. 3, pp. 2010–2022, Sep. 2021, doi: 10.1109/TTE.2020.3044524.

[3] B. Yang et al., "Enhanced cyber-attack detection in intelligent motor drives: A transfer learning approach with convolutional neural networks," *IEEE J. Emerg. Sel. Topics Ind. Electron.*, vol. 5, no. 2, pp. 710–719, Apr. 2024, doi: 10.1109/JESTIE.2023.3346802.

[4] B. Yang, J. Ye, S. Coshatt, W. Song, and F. Zahiri, "Data-driven approach for detection of physical faults and cyber attacks in manufacturing motor drives," in *Proc. IEEE Energy Convers. Congr. Expo.*, Detroit, MI, USA, 2022, pp. 1–6, doi: 10.1109/ECCE50734.2022.9948143.

[5] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton, "A simple framework for contrastive learning of visual representations," in *Proc. 37th Int. Conf. Mach. Learn.*, vol. 119, 2020, pp. 1597–1607.

[6] Y. Zhang, Z. Liu, and Q. Huang, "A contrastive learning-based fault diagnosis method for rotating machinery with limited and imbalanced labels," *IEEE Sensors J.*, vol. 23, no. 14, pp. 16402–16412, Jul. 2023, doi: 10.1109/JSEN.2023.3284044.

[7] J. Miao, Y. Liu, Q. Yin, B. Ju, G. Zhang, and H. Wang, "A novel soft fault detection and diagnosis method for a DC/DC buck converter based on contrastive learning," *IEEE Trans. Power Electron.*, vol. 39, no. 1, pp. 1501–1513, Jan. 2024, doi: 10.1109/TPEL.2023.3320878.

[8] J. Chen, D. Li, R. Huang, Z. Chen, and W. Li, "Contrastive learning-based feature-consistency distillation network for weak fault diagnosis of harmonic drive," *IEEE Trans. Instrum. Meas.*, vol. 74, 2025, Art. no. 3514210, doi: 10.1109/TIM.2025.3544384.

[9] K. Zhang, R. Cai, C. Zhou, and Y. Liu, "Debiased contrastive learning for time-series representation learning and fault detection," *IEEE Trans. Ind. Informat.*, vol. 20, no. 5, pp. 7641–7653, May 2024. [Online]. Available: https://ieeexplore.ieee.org/document/10443248

[10] H. Yang et al., "Real-world cyber security demonstration for networked electric drives," *IEEE Trans. Emerg. Sel. Topics Power Electron.*, vol. 13, no. 4, pp. 4659–4668, Aug. 2025, doi: 10.1109/JESTPE.2025.3550830.

[11] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. Int. Conf. Learn. Representations*, 2015. [Online]. Available: https://arxiv.org/abs/1412.6980

[12] L. Fei-Fei, R. Fergus, and P. Perona, "One-shot learning of object categories," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 4, pp. 594–611, Apr. 2006, doi: 10.1109/TPAMI.2006.79.