Enhanced Cyber-Attack Detection in Intelligent Motor Drives: A Transfer Learning Approach with Convolutional Neural Networks

Bowen Yang, Shushan Wu, Kun Hu, Jin Ye, Wenzhan Song, Ping Ma, Jianjun Shi, Peng Liu

***Abstract*—As networked digital control units become increasingly prevalent in intelligent motor drive systems, cybersecurity concerns have risen, leading to the development of various cyber-attack detection methods to improve system reliability. Although data-driven methods offer advantages over physics-based approaches, the requirement for extensive experimental data presents a significant challenge. This paper proposes a novel cyber-attack detection approach for motor drives using Transfer Learning based on Convolutional Neural Networks (CNN). The method initially pre-trains a CNN model with substantial simulation data and fine-tunes it using transfer learning with limited experimental data, achieving outstanding detection performance with 99.5% accuracy while reducing development costs, risks, and time. Additionally, the proposed model Maintains satisfactory detection accuracy of over 96% even when experimental training data is limited to 10% of original available data. The findings indicate that transfer-learned models exhibit faster convergence and better performance when limited experimental data is available compared to newly-trained models. The proposed approach substantially reduces the reliance on large quantities of experimental data during the development process, lowers costs and risks associated with cyber-attack detector development, strengthens the connections between simulations and experiments, and significantly shortens the development period by leveraging powerful simulation models.**

***Index Terms*—motor drives, cybersecurity, deep neural network, transfer learning, anomaly detection**

I. Introduction

Over the past decade, concerns about cybersecurity in intelligent power electronics systems have grown due to the widespread implementation of networked digital control units. Numerous studies [1], [2], [3], [4] have demonstrated the vulnerabilities and impacts of modern power electronics systems across various applications, such as photovoltaic (PV) systems, electric vehicles, and intelligent manufacturing systems. Recent research has focused on different detection approaches targeting diverse power electronics applications, including DC microgrids, PV farms, and industrial motor drives, to address these concerns and enhance the reliability of intelligent power electronics systems. The majority of recently proposed detection methods can be classified as either physics-based methods or data-driven methods.

Physics-based methods commonly detect cyber-attacks by analyzing pre-defined system performance metrics or residuals between predicted system variables and corresponding true measurements [5]. For instance, [6] proposed a cooperative vulnerability factor for each power electronics agent within the microgrid to detect stealthy cyber-attacks. [7], [8] characterized local voltage and frequency measurements into a 2D feature space, enabling the metrics to distinguish between cyber-attacks and physical faults in microgrids. [9] utilized signal temporal and time-frequency logic formalism to detect anomalies in microgrids. [10] developed a detection approach targeting microgrid false-data-injection attacks using the discordant element. [11] devised an attack detector for PV farms based on harmonic state space models. [12] proposed a residual-based detector for island microgrids using the harmonics state-space matrix and space phase model. [13] proposed an anomaly detection method for motor drives in electric vehicle powertrains using self-defined frequency-domain metrics. However, most physics-based methods rely on accurate physical models of the target systems, which are unavailable for most cyber-attack scenarios. In real-world applications, cyber-attacks are highly unpredictable, and their analytical impact models heavily depend on specific attack policies. These factors render the performance of most physics-based methods unreliable.

Recent research has begun to harness the power of data-driven methods to develop model-free detection methods in power electronics systems, reducing dependency on physical models. [14] adopted a specific type of recurrent neural network, namely a nonlinear auto-regressive exogenous model, to detect false data injection attacks in microgrids. [15] proposed an attack detection method by combining deep neural networks and wavelet singular value decomposition. [16] employed multi-class support vector machines to detect and localize false-data-injection and denial-of-service attacks in inverter-based systems. [17] proposed a detection and diagnosis method targeting data integrity attacks in solar farms using a multilayer long short-term memory network. [18] examined the effectiveness of various standard data-driven methods with micro-PMU data in detecting cyber-attacks in PV farms. [19], [20] developed anomaly detection methods for electric

Manuscript received XXX, 2023; revised XXX, 2023; accepted XXX, 2023; online XXX, 2023. (*Corresponding author: Jin Ye. Email: jin.ye@uga.edu)*

This research was partially supported by U.S. National Science Foundation ECCS-EPCN 2102032 and NSF-SATC-2019311, and U.S. Department of the Air Force FA8571-20-C-0017.

B. Yang, K. Hu, J. Ye, and W. Song are with College of Engineering, University of Georgia, Athens, GA, USA (e-mail: bowen.yang@uga.edu, kun.hu@uga.edu, jin.ye@uga.edu, wsong@uga.edu).

S. Wu and P. Ma are with Department of Statistics, University of Georgia, Athens, GA, USA (e-mail: shushan.wu@uga.edu, ping.ma@uga.edu).

J. Shi is with School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA, USA (e-mail: jianjun.shi@isye.gatech.edu).

P. Liu is with College of Information Sciences and Technology, Pennsylvania State University, State College, USA (e-mail: pxl20@psu.edu).

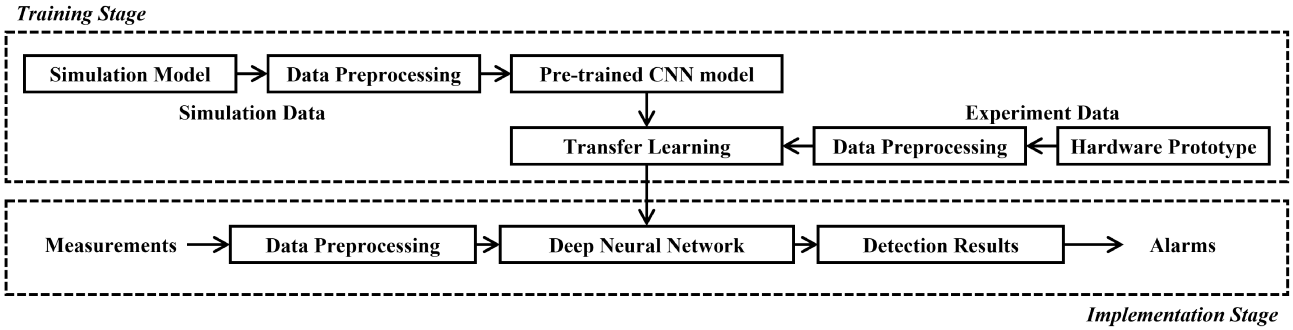


Fig. 1. General flow chart of the proposed method.

vehicle traction motor drives using a combination of support vector machines, random forests, k-nearest-neighborhood, and logistic regression. [21] employed supervised classification methods to differentiate cyber-attacks and physical faults in manufacturing motor drives.

Despite the advantages of recently developed data-driven approaches, a significant challenge in deep learning-based cyber-attack detection in power electronics systems is the requirement for large-volume training datasets. The model can only learn features incorporated in the training data, and the algorithm may fail when testing data contains different features [22], [23]. To address this issue, a large-scale training dataset is necessary to include similar data to the testing data. However, the computational cost due to the large volume of training data hinders deep learning model performance. Transfer learning techniques have been proposed to enable machine learning models to leverage knowledge from one domain to another [24], thus reducing the amount of required training data [25]. Deep transfer learning methods have been utilized in cyber-attack detection and fault diagnosis in intelligent machine systems. Some methods [26], [22], [27] employ deep adversarial models to achieve transfer learning by minimizing predicted domain labels, while others aim to minimize the discrepancy between learned features from the source and target domains [23]. Nevertheless, most existing deep-learning-based methods for intelligent power electronics systems rely on simulation data rather than hardware experimental datasets. Compared to experimental datasets, models trained on simulation data are more challenging to implement in real-world environments, particularly for mission-critical power electronics systems. Generating large amounts of experimental datasets faces two challenges: (1) experimental data generation requires substantial resources, and (2) cyber-attack experiments on intelligent power electronics systems pose risks and may cause damage to the environment and human health. Consequently, acquiring vast amounts of experimental data for training data-driven models is unrealistic in practice.

This paper aims to address this issue by combining simulation and experimental datasets using transfer learning, enabling the resulting motor drive attack detector to perform well even with limited experimental data. Fig. 1 shows a general flow chart of the proposed method. The method first trains a convolutional neural network (CNN) with extensive simulation data. Then, it uses transfer learning to fine-tune the pre-trained CNN model with limited experimental data. In other words, the proposed method first learns the primary characteristics of different cyber-attacks with simulation data. Then, it learns how to adapt the existing knowledge to real-world target prototype systems with limited experimental data. The contributions of the proposed method to existing data-driven attack detection research can be summarized as follows:

1) This work proposes a novel approach for cyber-attack detection in motor drives using Transfer Learning based on Convolutional Neural Networks (CNN).
2) The proposed method achieves outstanding detection performance with an accuracy of 99.5%, while reducing development time, costs, and risks.
3) The proposed transfer learning model effectively reduces convergence time during training in the presence of limited experimental data.
4) The proposed model maintains satisfactory detection accuracy of over 96% even when experimental training data is highly limited (10% of available data).

The rest of this paper is organized as follows: Section II will describe the details of the proposed method; Section III will elaborate on case studies, simulation, and experiment setups used to validate the proposed method; Section IV will discuss the validation results; and Section V will address the conclusions.

II. The Proposed Detection Strategy

Existing methods for detecting anomalies in power electronics and electric machines include physics-based and data-driven approaches. Physics-based methods rely on the expertise of the target systems. Common physics-based methods include using residuals between measured and predicted system variables and self-defined system metrics to detect anomalies. Such methods have been widely used in fault detection for power electronics and electric machines in recent decades. However, with the rising concerns of cyber-attacks in recent years, physics-based methods exposed several weaknesses:

1) There is no accurate and fixed physical model for predicting system behaviors under cyber-attacks.
2) Residuals and metrics used in physics-based methods are inflexible, while cyber-attacks are constantly evolving

and highly unpredictable. Therefore, such methods may not be effective in cyber-attack scenarios.

In recent years, with the emergence of data-driven methods for detecting cyber-attacks in power electronics and electric machines, the advantages of such approaches started to reveal:

1) Data-driven methods do not require an accurate physical model.
2) Data-driven methods can adapt to different systems without changing their original architecture.

However, the core requirements for all data-driven methods are the quality and quantity of the available training datasets. So far, the primary sources for such datasets in power electronics and electric machines are simulations and experiments. Each source has its strengths and weaknesses. On the one hand, simulations can generate a large amount of data with considerably low costs. However, they can only capture some of the details of target systems due to limited computational power. On the other hand, experiments can accurately reflect system behaviors, but the costs and risks of cyber-attack experiments in power electronics and electric machines are substantial. Therefore, this paper proposes a new approach to develop the cyber-attack detector using transfer learning based on CNN to combine the advantages of both simulation and experiment datasets (Fig. 1). The proposed method first uses extensive simulation datasets to train a CNN model. This pre-trained model includes a convolutional network and a fully-connected network. The convolutional network will map the input data to a feature space. The fully-connected network will map this feature space to the estimated probabilities for each category. After acquiring this pre-trained CNN model based on simulation data, the proposed method will use the transfer learning method to fine-tune this pre-trained CNN model with limited experimental data. The rest of this section will elaborate on the details of the proposed method.

A. Training Stage

The training stage includes five tasks: generating simulation data; collecting experiment data; preprocessing collected data; pre-training the CNN model; and fine-tuning the CNN model using transfer learning.

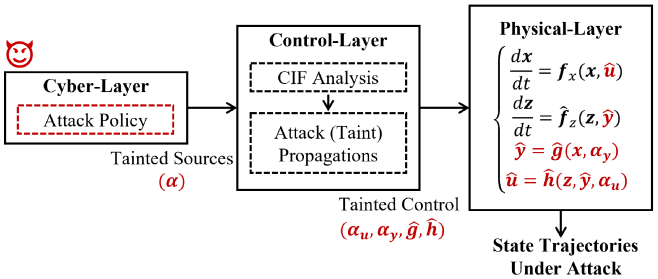


Fig. 2. Diagram of the simulation model, where α_u, α_y, $\hat{g}$, $\hat{h}$ are the attack coefficients and tainted control laws, $\hat{y}$, $\hat{u}$ are the resulted tainted feedback variables and control commands, and x, z, f_x, f_z are the state variables and equations for physical plants and digital controllers.

1) Simulation Data Generation: The proposed method uses simulation data to pre-train the CNN model. Therefore, the simulation data needs to reflect the significant impacts of cyber-attacks on motor drives and avoid trivial details like switching-caused ripples and measurement noise to lower the computational costs. Fig. 2 shows the adopted simulation model based on the motor drive average model and kill-chain-liked control information flow (CIF) model. Such a simulation model first maps different cyber-attacks to the CIF model, which reflects how attacks propagate in the motor drive controller. Fig. 3 shows an example of the CIF model for a permanent magnet synchronous motor (PMSM) drive with field-oriented control. The red paths in Fig. 3 denote the propagation paths of a false data injection attack targeting one of the offset variables for analog-to-digital-conversion (ADC) units. The resulting tainted control law for this example is shown in eq. (1) - eq. (8),

$$\frac{dz_m}{dt} = \Omega_m - \omega_m \tag{1}$$

$$\frac{d\hat{z}_d}{dt} = I_d - (i_d - \frac{2}{3}k_{adc} \cdot \cos\theta \cdot \alpha) \tag{2}$$

$$\frac{d\hat{z}_q}{dt} = I_q - (i_q + \frac{2}{3}k_{adc} \cdot \sin\theta \cdot \alpha) \tag{3}$$

$$I_q = K_{mp}(\Omega_m - \omega_m) + K_{mi} \cdot z_m \tag{4}$$

$$\hat{u}_d = K_{dp}(I_d - (i_d - \frac{2}{3}k_{adc} \cdot \cos\theta \cdot \alpha)) - K_{di} \cdot \hat{z}_d \tag{5}$$

$$\hat{u}_q = K_{qp}(I_q - (i_q + \frac{2}{3}k_{adc} \cdot \sin\theta \cdot \alpha)) + K_{qi} \cdot \hat{z}_q \tag{6}$$

$$\hat{v}_d = \hat{u}_d - \omega_m L_s(i_q + \frac{2}{3}k_{adc} \cdot \sin\theta \cdot \alpha) \tag{7}$$

$$\hat{v}_q = \hat{u}_q + \omega_m L_s(i_d - \frac{2}{3}k_{adc} \cdot \cos\theta \cdot \alpha) + \omega_m \lambda_{PM} \tag{8}$$

where z_m, z_d, z_q are the controller state variables, i_d, i_q, ω_m are the current and speed feedback variables, Ω_m, I_d, I_q are the references, v_d, v_q, u_d, u_q are the controller output variables, θ is the rotor position angle, L_s is the machine winding inductance, λ_{PM} is the flux linkage of the permanent magnet, k_{adc} is the ADC gain of the MCUs, K_x are controller parameters, and α is the attack coefficient.

The simulation model will then map the tainted control laws to the average model for the physical plant (i.e., motor drive) and solve the tainted dynamic equations to get system state trajectories under attack. Finally, such system state trajectories will be converted to measurement samples and organized as data sets for later training processes. Fig. 5 shows one sample of the measurement (motor line currents) waveform from simulation data sets.

2) Experiment Data Collection: The proposed method uses experimental data to fine-tune the pre-trained CNN model so that the final model can perform well in real-world target motor drives. As the pre-trained model covers most structures and features of different cyber-attacks, the experiment does not need to include many attack scenarios. Fig. 4 shows a diagram of the prototype motor drives used to generate experiment data sets. Such a prototype implements the control algorithms in a digital signal processor (DSP) connected to a host computer. The host computer is exposed to public networks where

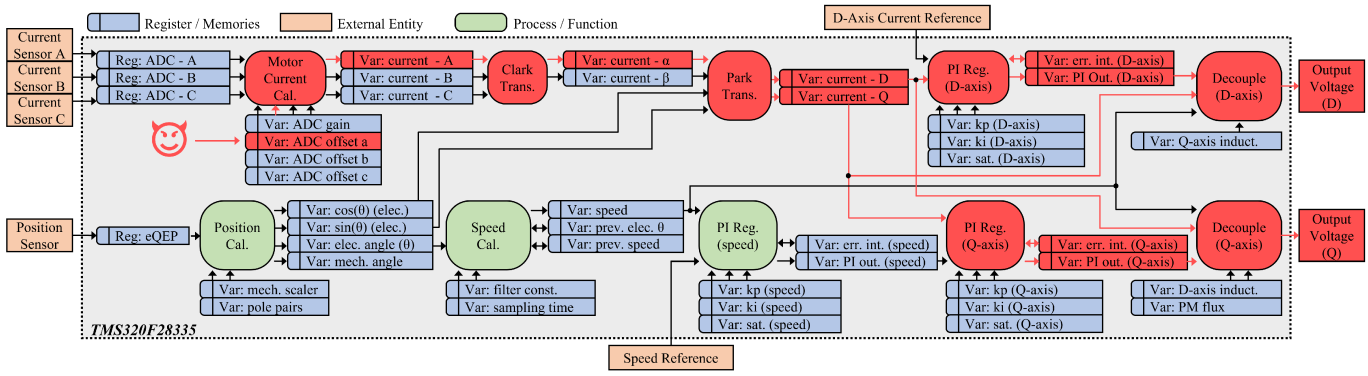


Fig. 3. An example of the kill-chain-liked control information flow model.

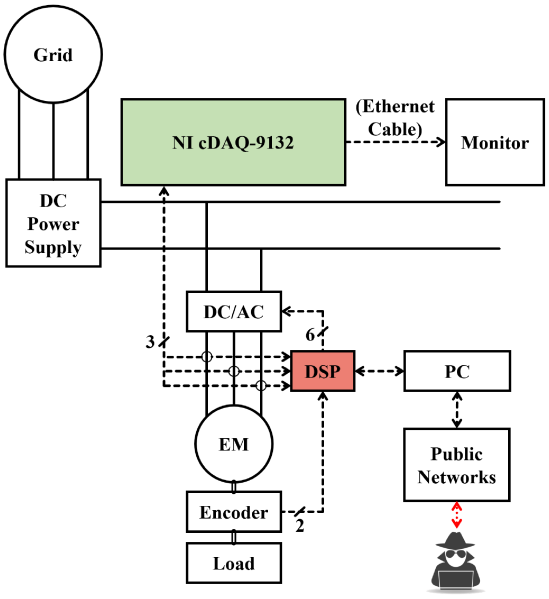


Fig. 4. Diagram of the prototype motor drives for generating experiment data.

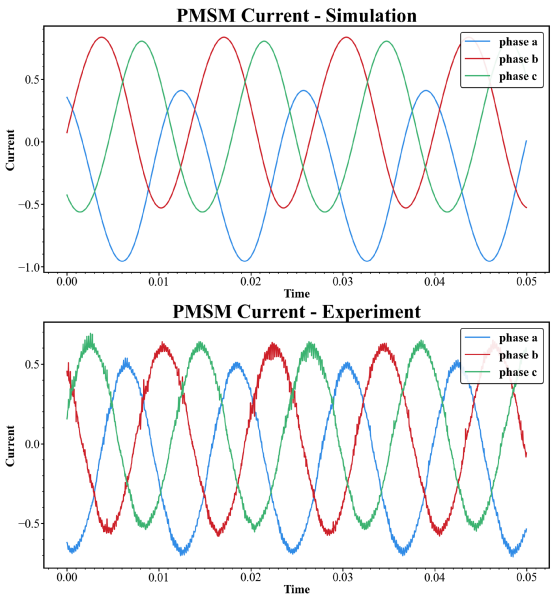


Fig. 5. Samples of simulation (top) and experiment (bottom) data sets.

the malicious hacker is trying to attack the prototype motor drive. As experimented cyber-attack scenarios need to be fully controllable, such prototype implants pre-defined malicious codes to the DSP and leaves some 'back-doors' to trigger these codes. Therefore, the hacker will explore these 'back-doors' and trigger the pre-defined cyber-attacks, reducing the risks of unpredictable impacts during experiments. Fig. 5 shows a sample waveform from experiment data sets. It shows that experiment data sets add real-world ripples and noises while the impact of cyber-attacks is similar to the simulation results.

3) Data Preprocessing: The primary objective of the proposed methodology is to enable continuous monitoring of the target system. Consequently, the method involves the periodic collection of a fixed quantity of samples (window size) from sensor signals (e.g., motor line currents) at a predetermined sampling frequency. Upon obtaining the raw measurement samples, the method transforms these samples into a specific format suitable for the input layer of the Convolutional Neural Network (CNN) model. Given that the majority of cyber-attacks targeting motor drives tend to induce undesired harmonics in motor line currents, these attacks become more discernible once the motor line currents are transferred to their frequency domains. As such, the proposed methodology initially employs the Fast Fourier Transform (FFT) to convert the sampled three-phase motor line currents into three distinct spectra. In this study, a sampling rate of 25 kHz and a window size of 1250 points (0.05 s) are used. Furthermore, since most harmonics resulting from cyber-attacks are concentrated in the lower frequency range, the first 100 elements (0 Hz to 2000 Hz) are chosen to reduce the size of input features.

4) Pre-training Process with Simulation Data: Utilizing the developed simulation model, it is possible to generate a substantial volume of data encompassing diverse attack scenarios and operating conditions. Since the preprocessing step employs the FFT to convert raw measurements into three sets of motor line current signature spectra, the input layer of the CNN model accepts three channels of 1-D arrays.

TABLE I
CNN STRUCTURE

Layer (type: index)	Output Shape	Number of Parameters
Input: 0	[:,100,3]	–
Conv1d: 1	[:,98,32]	320
ReLU: 2	[:,98,32]	–
MaxPool1d: 3	[:,49,32]	–
Conv1d: 4	[:,47,64]	6,208
ReLU: 5	[:,47,64]	–
MaxPool1d: 6	[:,23,64]	–
Conv1d: 7	[:,21,128]	24,704
ReLU: 8	[:,21,128]	–
MaxPool1d: 9	[:,10,128]	–
Flatten: 10	[:,1280]	–
Linear: 11	[:,128]	163,968
Linear: 12	[:,64]	8256
Linear: 13	[:,3]	195

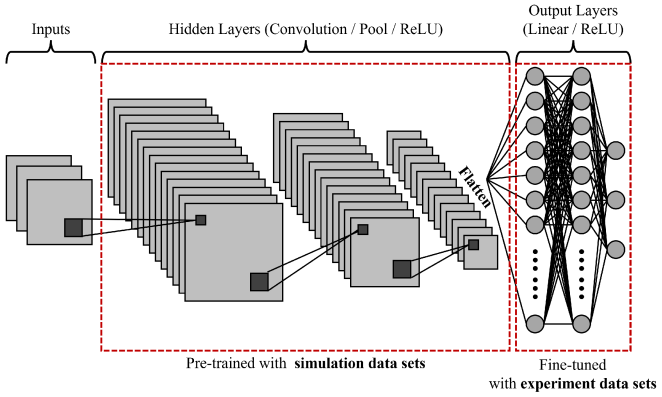


Fig. 6. Diagram of the CNN structure.

Subsequently, the model classifies these spectra into normal conditions and various attack scenarios. Table I and fig. 6 show the proposed method's general structure of the CNN model. As shown in table I, layer 0-9 is the backbone of the CNN model containing convolutional layers (Conv1d), nonlinear activation functions (ReLU), and maximum pooling layers (MaxPool1d). This first part of the CNN maps the input spectra images to a 10×128 feature space. Such feature space captures the primary structures of normal operating conditions and different cyber-attack scenarios. Then, layer 10 flattens the feature space to a 1×1280 array. Such an array is mapped to 3 possibilities corresponding to each scenario. (The example used in this paper includes two types of cyber-attack scenarios. Therefore, there are three possible classification outputs: normal condition, attack scenario 1, and attack scenario 2.) As shown in table I, there are 203,651 trainable parameters in this CNN model. These parameters are trained and validated by a large amount of simulation data from multiple scenarios.

5) Transfer Learning with Experiment Data: As indicated in Fig. 5, the primary patterns of cyber-attacks exhibit similarities between the simulated and experimental waveforms. Consequently, the objective of transfer learning is to incorporate subtle details (such as switching-induced ripples, mechanical disturbances, and measurement noise) into the pre-trained model, enabling the final model to accurately detect

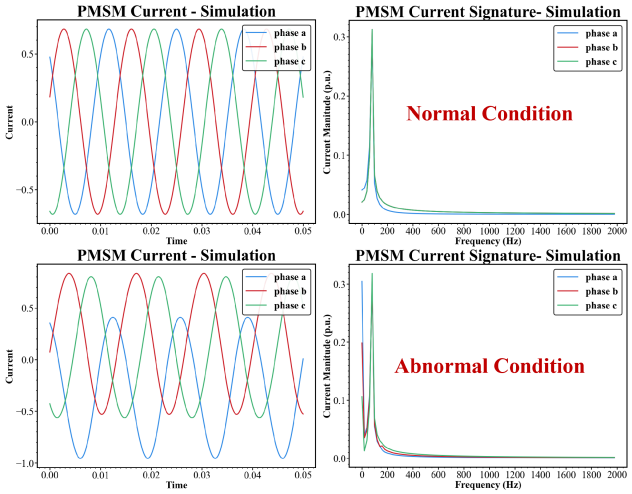


Fig. 7. Sample plots of simulation data set.

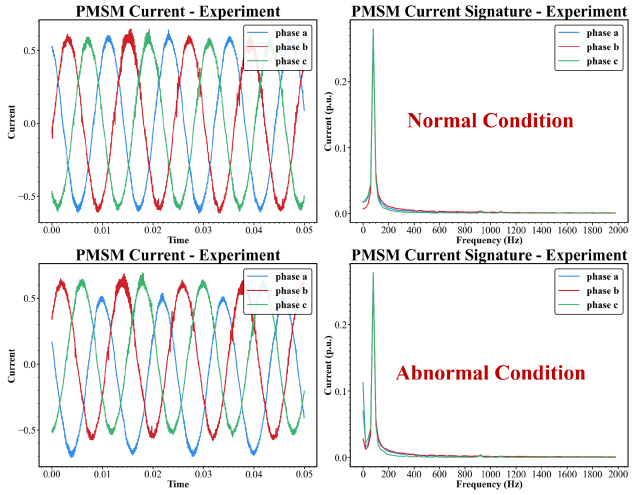


Fig. 8. Sample plots of experiment data set.

and distinguish normal conditions and various cyber-attack scenarios. Specifically, the proposed method freezes the parameters from layers 0-10 and sets the pre-trained parameters of layers 11-13 as initial values. The network is then re-trained with a limited amount of experimental data. Due to the well-developed pre-trained model, the re-training process requires only a minimal amount of experimental data. Simultaneously, as shown in Table I, the number of trainable parameters in the transfer learning model is reduced from 203,651 to 172,419, further decreasing the training time for transfer learning. Upon completing transfer learning, the resulting classifier accepts preprocessed signature spectra as inputs, outputs probabilities for each scenario, and generates detection results by selecting the scenario with the highest probability.

B. Implementation Stage

After generating the final classification model in the training stage, the implementation stage will continuously use the developed classifier to monitor the target motor drives. The

online monitoring system will follow a straightforward logic shown below:

- Step 1: Sample the measured current sensor signals with specified window size and sampling frequency.
- Step 2: Convert the raw measurement to motor line current signature spectra using the same parameters in the training stage.
- Step 3: Feed the spectra to the final classification model from the training stage.
- Step 4: Generate monitoring results based on the classifier output.
- Step 5: Update detection result and clear the workspace except for the preprocessing parameters and classification model.
- Step 6: Start a new monitoring cycle from Step 1.

It is essential to underscore the critical role that the safety and security of the monitoring signals play in the efficacy of the method proposed herein. Although various approaches exist to ensure the integrity of a limited set of signals, a comprehensive treatment of this topic lies beyond the scope of the present study. Therefore, for the purposes of this paper, we operate under the assumption that the safety and security of the monitoring signals have been assured.

III. Simulation And Experiment Case Study

This section forms a case study including two types of cyber-attacks targeting a 1.5 kW PMSM motor drive to test and validate the proposed method. The rest of this section describes details of the case study and simulation and experiment setups.

A. Case Study

The case study in this section includes typical false-data-injection (FDI)attacks targeting two control variables: the offset variable of the ADC unit and the calculated speed feedback variable. In addition, it is imperative to acknowledge that in practical operational environments, the nature of cyber-attacks often remains elusive to system architects and designers. Nevertheless, robust datasets can be synthesized through a combination of historical records and simulated adversarial engagements involving blue-red team exercises. In the present study, we have meticulously crafted attack scenarios to rigorously evaluate the performance of the proposed detection algorithms, thereby substantiating the novelty and efficacy of our methodological approach.

Another pivotal aspect to consider is the distinction between common physical faults and cyber-attacks. As elucidated in [19], [21], physical faults originate from tangible events and damages, whereas cyber-attacks manifest through digital controllers. The signature frequencies and event-response patterns associated with physical faults are intrinsically linked to the physical plants. For instance, the characteristic frequencies of electric machine bearing faults depend on both rotating speed and bearing structures, while inter-turn short circuit faults correlate with the number of turns in the shorted windings. Conversely, the response characteristics of cyber-attacks hinge on the controller configurations and control law implementations, making the current spectra under cyber-attacks dependent on variables such as controller bandwidth and sampling frequencies. These distinctions manifest in the motor current spectra, which we demonstrate in [19], [21] as distinguishable using data-driven methods, such as the CNN model employed in this study. Therefore, it is paramount to underscore that the principal aim of our proposed methodology is leveraging transfer learning techniques to overcome the data scarcity challenge in real-world cyberattack scenarios for data-driven detection methods. Our approach meticulously refines a baseline model, constructed from extensive simulation datasets, with a limited set of real-world data, establishing a robust attack detection framework. However, it is pertinent to note that differentiating between cyberattacks and physical faults falls beyond the scope of this study.

Meanwhile, in the realm of power converters and electric machine drives, a myriad of control laws exist. For the purpose of this study, we have selected field-oriented control, one of the most universally applied and representative control laws, as the testbed for demonstrating and validating the proposed methodology. It should be noted that the principles articulated herein are not limited to this specific control law. Indeed, analogous tainted control laws could be derived for alternative control algorithms by leveraging the control information flow models delineated in Fig.2 and Fig.3.

1) Scenario 1: FDI attack on the motor current offset variable: In practical motor drive controllers, current offset variables are critical to compensate for the current sensors' zero drift issues. Most controllers calculate such offset variables during the initialization process. After initialization, these offsets will maintain constant. However, as these offsets are stored in the memory data sections, multiple attacks could access these variables and maliciously modify them, such as buffer-overflow attacks and FDI attacks. This case study considers a scenario where the motor phase A current sensor offset variable is under an FDI attack. According to the CIF model in fig. 2, the tainted variable and attack policy is shown in eq. (9),

$$\hat{x}_{offsetA} = x_{offsetA} + \alpha \tag{9}$$

where $x_{offsetA}$ and $\hat{x}_{offsetA}$ is the original and attacked motor phase A current offset variables; α is the attack coefficient. The resulting tainted control laws for this scenario is shown in eq. (1) - eq. (8).

2) Scenario 2: FDI attack on the motor speed feedback variable: Besides current offset variables, the calculated speed feedback is also a vulnerable target of malicious attacks. For example, the Stuxnet worm compromised the industrial control

TABLE II
Parameters for operating conditions and attack coefficient

Operating Speed (rpm)	Load Torque (N*m)	Attack Coefficient
900	1.38	±0.3
1200	1.85	±0.275
1500	2.31	±0.25
1800		±0.225
		±0.2
		±0.175
		±0.15

system by manipulating the rotating speeds of industrial motor drives. Suppose the attack policy is the same as eq. (9), which is shown in eq. (10), where ω_m and $\hat{\omega}_m$ is the original and attacked motor speed feedback variables.

$$\hat{\omega}_m = \omega_m + \alpha \tag{10}$$

Then, the resulting tainted control law is shown in eq. (11) - eq. (18).

$$\frac{d\hat{z}_m}{dt} = \Omega_m - \hat{\omega}_m = \Omega_m - (\omega_m + \alpha) \tag{11}$$

$$\frac{dz_d}{dt} = I_d - i_d \tag{12}$$

$$\frac{d\hat{z}_q}{dt} = \hat{I}_q - i_q \tag{13}$$

$$\hat{I}_q = K_{mp}(\Omega_m - (\omega_m + \alpha)) + K_{mi} \cdot \hat{z}_m \tag{14}$$

$$u_d = K_{dp}(I_d - i_d) - K_{di} \cdot z_d \tag{15}$$

$$\hat{u}_q = K_{qp}(\hat{I}_q - i_q) + K_{qi} \cdot \hat{z}_q \tag{16}$$

$$\hat{v}_d = u_d - (\omega_m + \alpha)L_s i_q \tag{17}$$

$$\hat{v}_q = \hat{u}_q + (\omega_m + \alpha)L_s i_d + (\omega_m + \alpha)\lambda_{PM} \tag{18}$$

B. Simulation Setups

Based on the tainted control laws described by eq. (1) - eq. (8) and eq. (11) - eq. (18), the simulation encompasses various attack coefficients α for both scenarios. Simultaneously, the simulation also considers different operating conditions for the motor drives. Table II provides details of the operating parameters and attack coefficients. With these diverse operating conditions and attack scenarios, the simulation comprises 336 ($4 \times 3 \times 14 \times 2$) distinct scenarios. Subsequently, the simulation extracts 100 samples from each scenario, forming a simulation dataset consisting of 67,200 samples labeled as 'normal condition', 'attack type 1', and 'attack type 2'. Fig. 7 displays two samples from the simulation dataset.

C. Experiment Setups

Fig. 9 presents a photograph of the experimental prototype, while Table III provides a detailed overview of its specifications. The prototype features a 1.5 kW Permanent Magnet Synchronous Motor (PMSM) and shares the same structure illustrated in Fig. 4. Operating at a speed of 1000 rpm, the prototype experiences a mechanical load torque of approximately 2.77 N*m. Field-Oriented Control (FOC) algorithms, along with the malicious 'backdoor,' are implemented using a TMS320F28335 Microcontroller Unit (MCU) from Texas Instruments. Table IV outlines the attack coefficients employed in various attack scenarios during the experiment. From the 10 distinct scenarios, a total of 2000 samples are extracted, effectively capturing the diverse characteristics of each scenario.Fig. 8 showcases two representative samples extracted from the experimental dataset. It is important to emphasize that in the interest of experimental safety and reliability, cyber-attacks were manually embedded and controlled. This approach ensured the availability of accurate labels for both normal operations and attack scenarios. By incorporating these experimental samples into the training and validation process, the study aims to assess the proposed method's performance and generalizability in real-world situations. Ultimately, this experimental setup serves to bridge the gap between simulation and real-world implementation, ensuring the development of a more reliable and robust cyber-attack detection system for motor drives.

TABLE III
SPECIFICATIONS OF THE EXPERIMENT PLATFORM.

Rated Power	1.5 kW	Stator Resistance	0.4050 Ω
Rated Current	8.2 A	Stator Inductance	0.0024 mH
DC Bus Voltage	200 V	Magnet Flux Linkage	0.0599 Wb
Rated Frequency	250 Hz	Number of Pole Pairs	5
Control Frequency	10 kHz	Motor Inertia	3.10e-4 kgm^2

TABLE IV
PARAMETERS FOR ATTACK COEFFICIENTS IN EXPERIMENT

Scenario 1 (ADC Offset)	Scenario 2 (Speed Feedback)
0.1	0.05
0.2	0.1
0.3	0.15
-0.1	-0.05
-0.2	-0.1

IV. DISCUSSIONS ON VALIDATION RESULTS

The simulation data (67,200 samples) and experimental data (2,000 samples) are initially divided into 80% for training datasets and 20% for validation datasets. Subsequently, the simulation data is employed to pre-train and validate the CNN model. The 80% experimental training datasets are considered as the overall available experimental datasets for training. Thereafter, only a fraction of these experimental training data (10%, 20%, 30%, 40%, 50%, 100%) is used to train transfer-learned CNN models, as well as new CNN models with identical structures from scratch. It is essential to emphasize that when utilizing fewer than 50% of samples for the training dataset, special precautions have been taken to preserve representativeness. In lieu of a random sampling strategy, we adopted a stratified sampling methodology. This approach ensures a proportional representation of both normal operating conditions and diverse cyber-attack types within the reduced dataset. Consequently, this mitigates the potential for sampling bias and affirms that the performance metrics derived from this subset are genuinely indicative of the system's behavior under a wide array of conditions. This section discusses the outcomes of the training and validation processes based on the aforementioned settings.

TABLE V
VALIDATION RESULTS WITH EXPERIMENT VALIDATION SETS

	Overall Accuracy	False Alarm Rated
Simulation-Trained Model	47.75%	92.25%
Transfer-Learned Model	99.50%	0.01%

A. Discussion: Validation with Experimental Data Sets

Table V highlights the validation outcomes derived from independent experimental validation sets, which were deliberately excluded from both the training and transfer learning

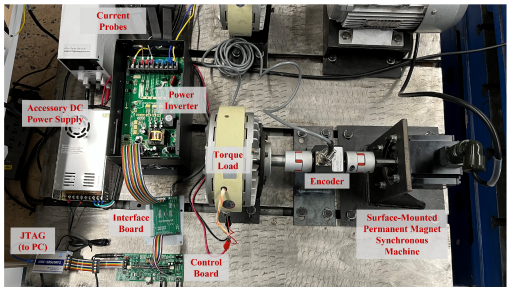


Fig. 9. Picture of the hardware experiment platform with a PMSM drive.

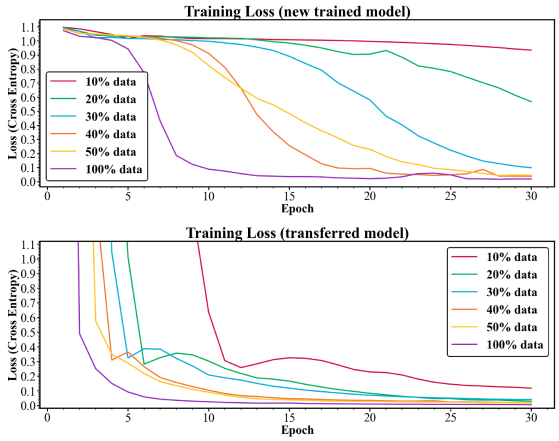


Fig. 10. Comparison of the training loss for the transfer-learned CNN model and newly-trained CNN model, utilizing varying sizes of available experimental data sets.

stages to ensure an unbiased evaluation. These outcomes reveal that adopting a simulation-trained model directly to a hardware experimental testbed would lead to notably poor performance, potentially compromising the effectiveness and reliability of the system. This evidence reinforces the notion that directly implementing simulation-trained models into real-world systems poses a significant challenge, primarily due to discrepancies between simulated environments and actual operational conditions. Factors such as unforeseen variables, noise, and non-ideal hardware characteristics can contribute to the observed performance degradation. In contrast, the transfer-learned model exhibits exceptional accuracy, surpassing 99%. This result emphasizes the value of transfer learning, which enables models to leverage pre-existing knowledge while adapting to new, real-world conditions. By refining the simulation-trained model using transfer learning techniques, the model's ability to generalize and perform accurately under real-world constraints is substantially improved. This demonstrates the importance of employing transfer learning when transitioning from simulated to real-world systems to ensure optimal performance and reliability.

B. Discussion: Training Loss

Fig.10 presents the training losses for both transfer-learned CNN models and newly-trained CNN models, employing varying sizes of available experimental datasets. As indicated by Fig.10, the initial losses for newly-trained models are smaller than those for transfer-learned models, suggesting that directly utilizing the parameters from the pre-trained model will yield higher losses compared to recalculating the initial parameters from available datasets. However, as the training process progresses, the losses of transfer-learned models rapidly converge to their minimum values, even when only 10% of the available experimental datasets are used. Conversely, the newly-trained models struggle to converge. Even with 100% of the available experimental training data, convergence still requires approximately 10 epochs, while transfer-learned models manage to converge after around 6 epochs. This finding supports the notion that the proposed transfer learning approach can reduce the convergence time of the training process when limited experimental datasets are available.

C. Discussion: Detection Accuracy

The detection accuracy is evaluated using experimental validation datasets, which are not involved in the training processes. Fig.11 displays the confusion matrices, and Fig.12 presents the overall accuracy for both transfer-learned CNN models and newly-trained CNN models, utilizing varying sizes of available experimental training datasets. As indicated by these results, when sufficient experimental training data is available, the classification accuracy exceeds 99% for both transfer-learned and newly-trained models. However, when the available experimental training data is limited, the classification accuracy of newly-trained models declines significantly, whereas transfer-learned models maintain a satisfactory accuracy of above 96%. In Fig. 11, the newly-trained models struggle to detect any cyber-attacks with only 10% of available experimental training data, while the transfer-learned model achieves a notable 96.75% detection accuracy. Concurrently, all anomalies successfully detected occurred within a span of two monitoring cycles. It is crucial to emphasize that although our study posits transfer learning as a promising avenue, it does not universally guarantee augmented performance across all contexts. Various optimization tactics can be deployed to increase the probability of successful application. Firstly, one can tailor the neural network architecture to match specific application scenarios. Secondly, regularization methods such as dropout and weight decay can be integrated to enhance the model's generalization performance, particularly when working with smaller target datasets. Lastly, custom logging mechanisms can be instituted for meticulous, real-time monitoring throughout the training process. It is also noteworthy that a significant discrepancy between the source and target datasets could render the application of transfer learning counterproductive. To mitigate this risk, it is imperative to quantitatively assess the similarity between datasets. Employing data augmentation techniques can be advantageous, especially when dealing with constrained target datasets, as this can counteract the tendency towards model overfitting. In this context, metric in [28] serves as a valuable tool for gauging the divergence between the source and target datasets by contrasting their out-of-distribution and in-distribution performance metrics. Our successful application of transfer learning corroborates that the

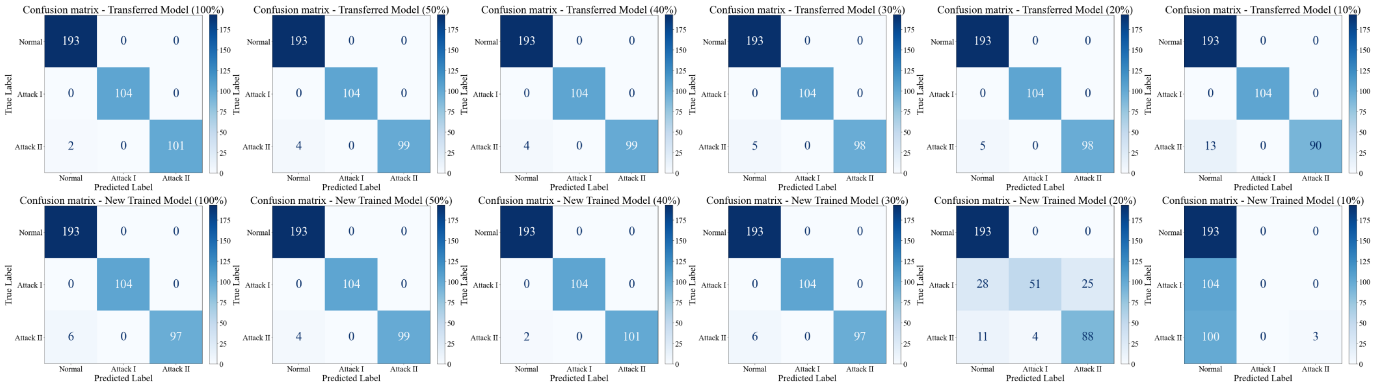
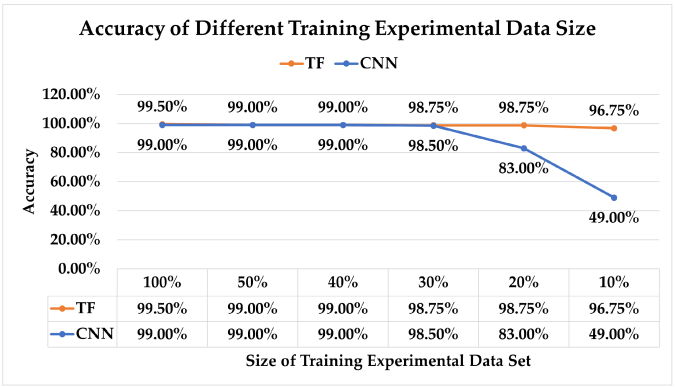


Fig. 11. Confusion matrices from experimental validation sets for the transfer-learned CNN model and newly-trained CNN model, utilizing varying sizes of available experimental data sets.



	100%	50%	40%	30%	20%	10%
TF	99.50%	99.00%	99.00%	98.75%	98.75%	96.75%
CNN	99.00%	99.00%	99.00%	98.50%	83.00%	49.00%

Fig. 12. Overall classification accuracy for the transfer-learned CNN model and newly-trained CNN model, utilizing varying sizes of available experimental data sets. (TF: transfer-learned model, CNN: newly-trained CNN model)

source and target datasets share a sufficiently high degree of similarity to justify this methodological choice. Furthermore, our approach advocates for the use of a robust baseline model, initially trained on a comprehensive simulated dataset, as a foundation. The transfer learning model subsequently adapts this baseline to address the idiosyncrasies of specific real-world scenarios. In light of this framework, the urgency of quantitatively measuring the dissimilarity between simulation and real-world datasets is mitigated.

V. Conclusion

In conclusion, this paper presents a novel development approach for cyber-attack detection in motor drives, employing transfer learning based on CNN. The proposed method initially pre-trains a CNN model using extensive simulation data and subsequently fine-tunes the model through transfer learning with a limited amount of experimental data. This approach achieves outstanding detection performance, with an accuracy of 99.5%, while significantly reducing the development period, costs, and risks associated with high-performance monitoring systems for modern motor drives. Our findings indicate that the initial losses for newly-trained models are smaller than those for transfer-learned models, suggesting that directly using parameters from the pre-trained model may result in higher losses compared to recalculating initial parameters from available data sets. However, as the training process advances, the losses of transfer-learned models swiftly converge to their minimum values, even with only 10% of available experimental data sets. In contrast, newly-trained models face difficulties in convergence. It takes approximately 10 epochs for them to converge when using 100% of available experimental training data, while transfer-learned models converge after around 6 epochs. These results support the notion that the proposed transfer learning approach can effectively reduce the convergence time of the training process in the presence of limited experimental data sets. When assessing detection accuracy with experimental validation data sets not used during training, our results show that both transfer-learned and newly-trained models achieve classification accuracy exceeding 99% when sufficient experimental training data is available. However, when experimental training data is limited, the classification accuracy of newly-trained models declines considerably, whereas transfer-learned models maintain a satisfactory accuracy of over 96%. With only 10% of available experimental training data, newly-trained models struggle to detect any cyber-attacks, while the transfer-learned model achieves a remarkable 96.75% detection accuracy. The substantial performance demonstrated by the proposed method contributes to the motor drive cyber-attack detection framework in several ways. It substantially reduces the reliance on large quantities of experimental data sets during the development process, lowers the costs and risks associated with cyber-attack detector development, strengthens the connections between simulations and experiments, and significantly shortens the development period by utilizing powerful simulation models.

References

[1] J. Ye, A. Giani, A. Elasser, S. K. Mazumder, C. Farnell, H. A. Mantooth, T. Kim, J. Liu, B. Chen, G.-S. Seo, W. Song, M. D. R. Greidanus, S. Sahoo, F. Blaabjerg, J. Zhang, L. Guo, B. Ahn, M. B. Shadmand, N. R. Gajanur, and M. A. Abbaszada, "A review of cyber–physical security for photovoltaic systems," *IEEE Journal of Emerging and Selected Topics in Power Electronics*, vol. 10, no. 4, pp. 4879–4901, 2022.

[2] D. Zhang, G. Feng, Y. Shi, and D. Srinivasan, "Physical safety and cyber security analysis of multi-agent systems: A survey of recent advances," *IEEE/CAA Journal of Automatica Sinica*, vol. 8, no. 2, pp. 319–333, 2021.

[3] B. Yang, L. Guo, F. Li, J. Ye, and W. Song, "Vulnerability assessments of electric drive systems due to sensor data integrity attacks," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 5, pp. 3301–3310, 2020.

[4] J. Ye, L. Guo, B. Yang, F. Li, L. Du, L. Guan, and W. Song, "Cyber–physical security of powertrain systems in modern electric vehicles: Vulnerabilities, challenges, and future visions," *IEEE Journal of Emerging and Selected Topics in Power Electronics*, vol. 9, no. 4, pp. 4639–4657, 2021.

[5] J. Giraldo, D. Urbina, A. Cardenas, J. Valente, M. Faisal, J. Ruths, N. O. Tippenhauer, H. Sandberg, and R. Candell, "A survey of physics-based attack detection in cyber-physical systems," *ACM Computing Surveys (CSUR)*, vol. 51, no. 4, pp. 1–36, 2018.

[6] S. Sahoo, S. Mishra, J. C.-H. Peng, and T. Dragičević, "A stealth cyber-attack detection strategy for dc microgrids," *IEEE Transactions on Power Electronics*, vol. 34, no. 8, pp. 8162–8174, 2019.

[7] K. Gupta, S. Sahoo, R. Mohanty, B. K. Panigrahi, and F. Blaabjerg, "Distinguishing between cyber attacks and faults in power electronic systems – a non-invasive approach," *IEEE Journal of Emerging and Selected Topics in Power Electronics*, pp. 1–1, 2022.

[8] V. B. Kurukuru, M. A. Khan, and S. Sahoo, "Cybersecurity in power electronics using minimal data–a physics-informed spline learning approach," *IEEE Transactions on Power Electronics*, vol. 37, no. 11, pp. 12 938–12 943, 2022.

[9] O. A. Beg, L. V. Nguyen, T. T. Johnson, and A. Davoudi, "Cyber-physical anomaly detection in microgrids using time-frequency logic formalism," *IEEE Access*, vol. 9, pp. 20 012–20 021, 2021.

[10] S. Sahoo, J. C.-H. Peng, A. Devakumar, S. Mishra, and T. Dragičević, "On detection of false data in cooperative dc microgrids—a discordant element approach," *IEEE Transactions on Industrial Electronics*, vol. 67, no. 8, pp. 6562–6571, 2020.

[11] J. Zhang, L. Guo, and J. Ye, "Cyber-attack detection for photovoltaic farms based on power-electronics-enabled harmonic state space modeling," *IEEE Transactions on Smart Grid*, vol. 13, no. 5, pp. 3929–3942, 2022.

[12] J. Zhang, J. Ye, and L. Guo, "Model-based cyber-attack detection for voltage source converters in island microgrids," in *2021 IEEE Energy Conversion Congress and Exposition (ECCE)*, 2021, pp. 1413–1418.

[13] B. Yang, L. Guo, and J. Ye, "Physics-based attack detection for traction motor drives in electric vehicles using random forest," in *2021 IEEE Applied Power Electronics Conference and Exposition (APEC)*, 2021, pp. 849–854.

[14] M. R. Habibi, H. R. Baghaee, T. Dragičević, and F. Blaabjerg, "Detection of false data injection cyber-attacks in dc microgrids based on recurrent neural networks," *IEEE Journal of Emerging and Selected Topics in Power Electronics*, vol. 9, no. 5, pp. 5294–5310, 2021.

[15] M. Dehghani, T. Niknam, M. Ghiasi, N. Bayati, and M. Savaghebi, "Cyber-attack detection in dc microgrids based on deep machine learning and wavelet singular values approach," *Electronics*, vol. 10, no. 16, p. 1914, 2021.

[16] A. A. Khan, O. A. Beg, M. Alamaniotis, and S. Ahmed, "Intelligent anomaly identification in cyber-physical inverter-based systems," *Electric Power Systems Research*, vol. 193, p. 107024, 2021.

[17] F. Li, Q. Li, J. Zhang, J. Kou, J. Ye, W. Song, and H. A. Mantooth, "Detection and diagnosis of data integrity attacks in solar farms based on multilayer long short-term memory network," *IEEE Transactions on Power Electronics*, vol. 36, no. 3, pp. 2495–2498, 2021.

[18] Q. Li, F. Li, J. Zhang, J. Ye, W. Song, and A. Mantooth, "Data-driven cyberattack detection for photovoltaic (pv) systems through analyzing micro-pmu data," in *2020 IEEE Energy Conversion Congress and Exposition (ECCE)*, 2020, pp. 431–436.

[19] B. Yang and J. Ye, "Data-driven detection of physical faults and cyber attacks in dual-motor ev powertrains," in *2022 IEEE Transportation Electrification Conference & Expo (ITEC)*, 2022, pp. 991–996.

[20] B. Yang, J. Ye, and L. Guo, "Fast detection for cyber threats in electric vehicle traction motor drives," *IEEE Transactions on Transportation Electrification*, vol. 8, no. 1, pp. 767–777, 2022.

[21] 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 *2022 IEEE Energy Conversion Congress and Exposition (ECCE)*, 2022, pp. 1–6.

[22] Y. Zhang and J. Yan, "Domain-adversarial transfer learning for robust intrusion detection in the smart grid," in *2019 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm)*, 2019, pp. 1–6.

[23] L. Wen, L. Gao, and X. Li, "A new deep transfer learning based on sparse auto-encoder for fault diagnosis," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 49, no. 1, pp. 136–144, 2019.

[24] B. Maschler and M. Weyrich, "Deep transfer learning for industrial automation: A review and discussion of new techniques for data-driven machine learning," *IEEE Industrial Electronics Magazine*, vol. 15, no. 2, pp. 65–75, 2021.

[25] Q. Li, J. Zhang, J. Ye, and W. Song, "Data-driven cyber-attack detection for photovoltaic systems: A transfer learning approach," in *2022 IEEE Applied Power Electronics Conference and Exposition (APEC)*, 2022, pp. 1926–1930.

[26] P. Liao, J. Yan, J. M. Sellier, and Y. Zhang, "Divergence-based transferability analysis for self-adaptive smart grid intrusion detection with transfer learning," *IEEE Access*, vol. 10, pp. 68 807–68 818, 2022.

[27] X. Li, W. Zhang, Q. Ding, and X. Li, "Diagnosing rotating machines with weakly supervised data using deep transfer learning," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 3, pp. 1688–1697, 2020.

[28] J.-C. Gagnon-Audet, K. Ahuja, M.-J. Darvishi-Bayazi, P. Mousavi, G. Dumas, and I. Rish, "Woods: Benchmarks for out-of-distribution generalization in time series," *arXiv preprint arXiv:2203.09978*, 2022.