Real-Time Simulation of Convolutional Neural Network Detectors for False Data Injection Attacks Using Typhoon HIL

Timur Aminov, Hongyu Wu, *Senior Member, IEEE,* Bo Liu and Hang Zhang, *Member, IEEE*
The Department of Electrical and Computer Engineering
Kansas State University, Manhattan, KS, USA

2024 IEEE Kansas Power and Energy Conference (KPEC) | 979-8-3503-7240-3/24/$31.00 ©2024 IEEE | DOI: 10.1109/KPEC61529.2024.10676276

***Abstract*—Research for cyber-security defensive strategies to achieve enhanced power system security and attack detection against various cyber-attacks, such as the False-Data Injection (FDI) attack, are topics that could benefit from real-time digital simulation (RTDS). Recently, neural networks have gained traction within smart grid applications including cyber-attack detection. This paper performs RTDS of FDI attacks against traditional power system state estimation and simulates a convolutional neural network-based detector on the IEEE 14-bus system. This paper demonstrates the benefit of Typhoon hardware-in-the-loop (HIL) for evaluating the effectiveness of neural network models in identifying and mitigating FDI attacks in power systems. By integrating Typhoon HIL into the simulation environment, we provide a novel approach to assess the resilience of power systems under cyber-attack scenarios.**

Keywords—Real-Time Digital Simulation (RTDS), Cyber-Security, False-Data Injection (FDI) Attacks, Typhoon HIL, Weighted-Least-Squared (WLS) State Estimation (SE), Convolutional Neural Network (CNN).

I. Introduction

Power systems have become more efficient and controllable over the last two decades, but at the expense of vulnerability to cyber-attacks through both their hardware and software. Sources of these vulnerabilities include the integration of a large number of measuring sensors or meters, which utilize GPS and telemetry communication to transfer data to SCADA for monitoring of real-time measurement sampling technology. A class of cyber-attacks that is being utilized in major cyber security breaches that have graced the news is the False Data Injection (FDI) Attack, which is designed to manipulate the data sampled within a communication network of measurement and control devices. The Ukrainian blackout in 2015 was reported to have been caused by an FDI attack hijacking SCADA operations with malicious firmware installed after a reconnaissance of phishing emails. These emails allowed the attackers to obtain authorized credentials to the target's SCADA system network giving them access to implement malware to manipulate meter measurements and expedite open substation breakers remotely [1]. The 2010 Iran Stuxnet Worm attack, which altered the code of programmable logic controllers to create discrepancies between reported system behavior and the actual operations of nuclear centrifuges [2], exemplifies the severe impact of False Data Injection (FDI) attacks. These attacks have been instrumental in driving targeted systems to critical states, underscoring the urgency of further research in the field of cyber-security. The historical evidence of cyber-attacks on energy infrastructure [3], including these high-profile incidents, highlights the critical need for robust defensive strategies within the SCADA operations of power systems. Such strategies are essential to safeguard against and mitigate the effects of these sophisticated cyber threats.

Real-Time Data Simulators (RTDS) can prove to be a useful tool in various application, especially for propelling cyber-security and neural network research. RTDS can help validate and emulate system models and communication networks combined with a Hardware in the Loop (HIL) testbed platform. An RTDS HIL setup was utilized to study the impact of load loss due to an FDI attack that specifically targeted the under-frequency load-shedding scheme [4]. Another RTDS HIL system [5] simulated and optimized a three-phase asynchronous motor speed control system. Virtual HIL (VHIL) takes HIL directly to the user's PC and can provide useful insight into how a system will respond prior to RTDS or HIL testing [6,7,8]. In [9] and [10], neural network models were designed and implemented in a HIL environment for hybrid electric vehicle and cyber-security applications, respectively. The growth in neural network application and research can be attributed to faster computer chip processing, specifically the increased use of general-purpose graphical processing units, and the increased availability of training data for neural network models [11].

This paper is motivated by the need to effectively and accurately simulate FDI attacks and their countermeasures using RTDS. We conducted a detailed study simulating an FDI attack on the IEEE 14-bus system using the Typhoon HIL environment shown in figure 1 to address this. Central to our approach is the deployment of a Convolutional Neural Network (CNN) model, designed as a sophisticated defense mechanism. This model performs a multilabel classification task, aiming to detect data manipulation caused by FDI attacks and to identify the specific measurement data targeted by these attacks. The data in question is used by the Weighted-Least-Squared (WLS) State Estimation (SE) algorithm, a

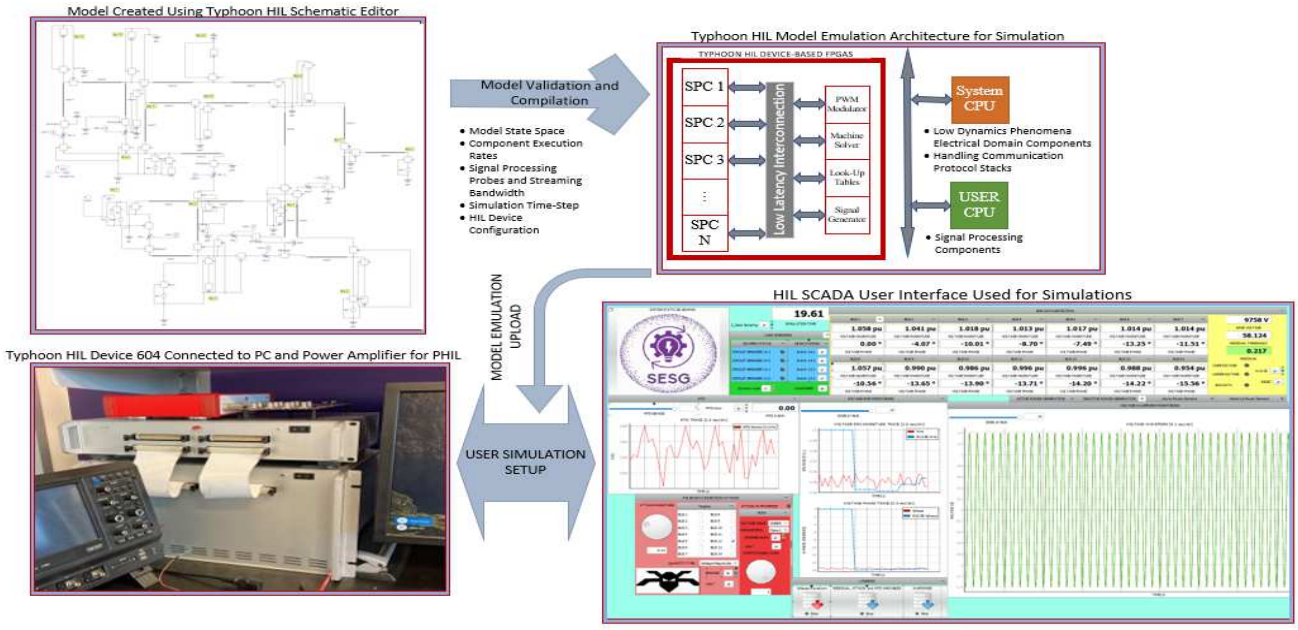

Fig. 1. Flow Chart containing Device that this paper's model schematic and SCADA interface used to achieve RTDS for the case studies performed and HIL setup with the Typhoon HIL 604 Device.

critical component for reporting system states. These states inform the optimal power flow algorithm, which is essential for generator dispatch in the system emulation. This paper highlights the potential of CNNs in enhancing the resilience of power systems against cyber threats, particularly in the complex and dynamic environment of smart grids.

II. Preliminaries: Cyber-Attack and CNN

Cyber-attacks are coordinated and require reconnaissance time to plan before executing. The objective of this attack model is to mislead the WLS SE algorithm implemented into the system's SCADA operations by manipulating the power measurements obtained and transmitted using an FDI attack. Assuming the worst-case scenario, the attacker has complete knowledge of the power system, such as distribution topology, electrical component parameters and setpoints, bad data detection (BDD) protocols, SE algorithms, and generator cost functions; and access to the measurement devices sampled by SCADA, to perform their FDI attack that manipulates the necessary data the operator utilizes for system operation or corrective actions.

A Convolutional Neural Network (CNN) leverages the use of the convolution operation within at least one layer of the network instead of general matrix operations. They are specialized in processing data that has a grid-like structure, such as time-series data and image data [15]. Given that an FDI attack targets this type of data utilized within power grid operations, CNN can help discern the inconsistency and co-occurrence dependency on the system's measurements during such an attack. Furthermore, after computationally expensive training of a CNN to learn the desired features of a given input data, they exhibit fast computation times for producing their output in real-time operations. The application of CNN to the electric grid is due to the fast changes in the safe operation of power systems.

A. The System Model and FDI Attack Scheme

The system's active power, P_i, and reactive power, Q_i, bus injections, active, P_{ij}, and reactive, Q_{ij}, branch power flows from bus i to bus j, and the slack bus voltage magnitude, V_1, quantities are sampled into a measurement vector, $Z = (z_1^t \cdots z_n^t)$ where n is the number of measurements utilized within the system. The targeted WLS SE models Z with the following equation:

$$Z = h(\hat{x}) + e$$

where e represents the Gaussian distributed noise error vector of each measurement, and $h(\hat{x})$ is the theoretical measurement vector. Considering each bus within the IEEE 14-bus system model as the set $i \in \{1,2,3, \ldots 14\}$ and $j \in \mathcal{N}_i$ as the set of all connected buses to bus i, $h(\hat{x})$ uses the equations(1-5) below of a two-port π-model of a network branch to construct the theoretical measurement vector, $h(\hat{x})$, from a given state vector, $\hat{x} = \left[\widehat{V_1} \quad \cdots \quad \widehat{V_{14}} \quad \widehat{\theta_1} \quad \cdots \quad \widehat{\theta_{14}}\right]^T$, representing the bus RMS voltage magnitudes, $\widehat{V_i}$, and phases, $\widehat{\theta_i}$.

$$\theta_{ij} = \theta_i - \theta_j \tag{1}$$

$$P_i = V_i \sum_{j \in \mathcal{N}_i} V_j (G_{ij} cos\theta_{ij} + B_{ij} sin\theta_{ij}) \quad (2)$$

$$Q_i = V_i \sum_{j \in \mathcal{N}_i} V_j (G_{ij} sin\theta_{ij} - B_{ij} cos\theta_{ij}) \quad (3)$$

$$P_{ij} = V_i^2 (g_{si} + g_{ij}) - V_i V_j (g_{ij} cos\theta_{ij} + b_{ij} sin\theta_{ij}) \quad (4)$$

$$Q_{ij} = -V_i^2 (b_{si} + b_{ij}) - V_i V_j (g_{ij} sin\theta_{ij} - b_{ij} cos\theta_{ij}) \quad (5)$$

It is assumed that the attacker has complete knowledge of the IEEE 14-bus system allowing for a stealthy attack to go unnoticed by the BDD scheme [12,13]. The FDI attack scheme adopted from [14] uses c_{inj} to represent the magnitude the attacker wishes to deviate a targeted state within $\hat{x}$. The targeted measurements within Z are dependent on which state of $\hat{x}$ is targeted based on the subgraph generated for a single bus attack as discussed in [14]. The goal of the attack model is to manipulate the output of $\hat{x}$ from the WLS SE by targeting the necessary measurements in Z while minimizing e with an attack vector a that is constructed as shown below.

$$a = h(\hat{x} + c_{inj}) - h(\hat{x}) \quad (6)$$

The attacked measurement vector, Z_a, for the WLS SE is calculated by the attack model using equation (7).

$$Z_a = Z + a \quad (7)$$

The attack model is performed under the assumption that the cyber layer of the system has been penetrated. The study focuses on the consequences of the system receiving falsified data. To achieve a more realistic setting within an HIL testbed environment, a time sequence shown in Fig. 2 was developed for simulating the Z sampling and WLS SE with scripts at consistent intervals. For the studies performed, the Z-sampling script executed every 4 seconds of the simulation time and the WLS SE Scripts executed 2 seconds after sampling of Z, allowing a 2-second window prior to the estimation of the states for the FDI attack.

B. Convolutional Neural Network Model and Training

The CNN model utilized for this study comes from [16], where a 1-dimensional CNN solves a classification of an FDI attack present within Z and whether each element within Z is compromised or uncompromised due to a present attack. With the BDD acting as the first line of defense against FDI attacks, the CNN classifier acts upon the same measurements that are passed through the BDD as the second line of defense. The CNN attempts to extract and analyze features of the FDI attack against the set measurements by determining the location of the compromised measurements within the system. For the study performed, the network architecture of the CNN consists of four 1-dimensional-convolutional layers, a flattening layer and a dense layer as shown in Fig. 3. The convolutional layers use the rectified-linear unit (ReLU) activation function and the dense layer uses the sigmoid activation function when passing their respective inputs through the layers. Details of the activation functions are discussed further in [16].

The training conducted on the CNN model used the mini-batch gradient descent method with a batch size of 100 and an epoch size of 100. The training data was obtained from the emulation of the IEEE 14-bus system having 8 buses (i = 4, 6, 9, 10, 11, 12, 13, and 14) experiencing a normally distributed dynamic load profile having a mean equal to that bus's base load and standard deviation equal to 1/6 of the base load with *Typhoon HIL Control Center* in real-time HIL mode using the typhoon device 604 to obtain 2,347 uncompromised samples of Z and 39,193 compromised samples from the FDI attack model targeting $\widehat{\theta_i}$ of bus i with a decrease of 0.026 radians. Following machine learning practice, the training data was split 7/10 for training the CNN and 3/10 for validating the model for each batch. Fitting used the initial learning rate of 0.001 and a patience of 5. The loss function for training shown in Fig. 4 was the cross-entropy function for each mini-batch.

For this study, the trained CNN model was implemented within the HIL emulation environment by having the CNN model perform its classification during the execution of the WLS SE within the simulation time sequence discussed prior. New load profiles were generated with similar base loads as the load profiles used during the emulation sampling session to obtain new Z vectors to pass through the CNN classifier.

III. Simulation Results

A. Simulation Setup: Validating Attack Model

To validate the success of the attack model discussed, the IEEE 14-bus power system was simulated with Typhoon HIL device 604 with constant loads at each bus where bus 14 was targeted. For the simulations, the user was able to specify which bus to target and quantity, its $\widehat{\theta_{bus}}$ or $\widehat{V_{bus}}$. *Data Loggers* and a *Stream Logger* were used within Typhoon HIL SCADA to sample the simulation data and emulated model output of the targeted bus voltage waveform every 250 milliseconds, respectively. The compiled model uses current and voltage phasor measurements at each bus to construct the system's Z matrix. A BDD module was designed to use WLS SE to give the system $\hat{x}$ vector, and compare the calculated r value to a threshold value, $\tau = 58.124$, calculated from a chi-square test with 95% confidence. Over- and under-voltage conditions were monitored for any $\widehat{V_{bus}}$ reported to be greater than 1.05 p.u. or less than 0.95 p.u., respectively, given the sampled Z.

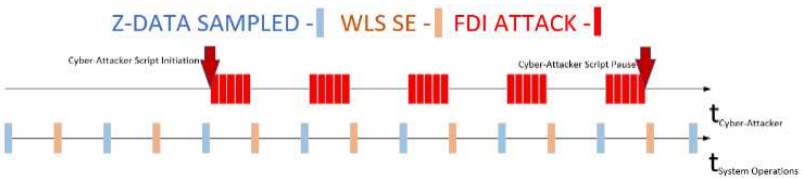

Fig. 2. Proposed Time Sequence for Executed SCADA System Operations with a simulated basic FDIA (top) during Variable Load conditions.

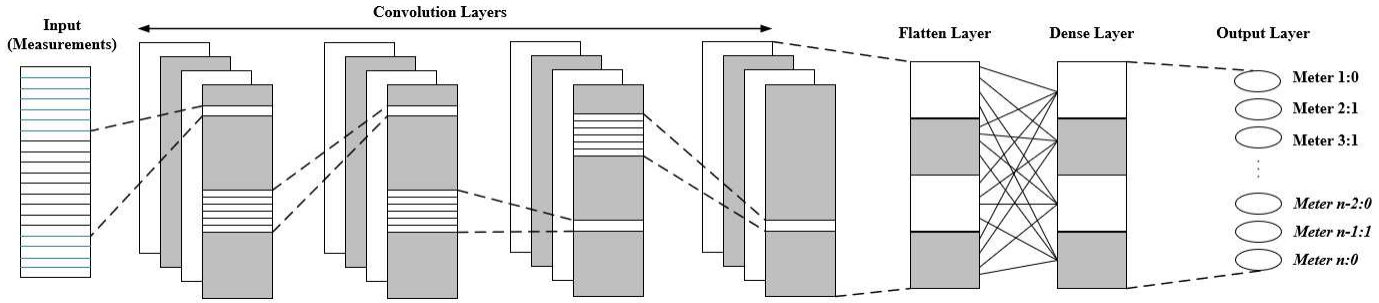

Fig. 3. Illustration of 1-Dimensional deep CNN architecture for proposed FDI attack classification.

The FDI attack against the WLS SE was simulated using steady-state operations of the power system. The cyber-attack model targeted $\widehat{\theta_{14}}$ with a $c_{inj} = 0.02$ radians increase. The WLS SE of bus 14's $\hat{x}$ and the reported residual before the attack are compared with their values after the attack in Table I. It can be seen that the attack successfully passes through the BDD criteria, and thus the attacker can remain stealthy, prior to CNN implementation, within reasonable c_{inj} values as the operator is required to maintain these quantities within a certain stable range.

B. *Simulation: Evaluating CNN Performance*

Using the same simulation setups as discussed in the prior section, the IEEE 14-bus system was modified using the dynamic load profiles obtained from a Western Electricity Conference Council (WECC) system. For the purposes of real-time emulation, the load profile was condensed to experience the hourly changes every ¼ second for a simulation time of 2190 seconds for each session performed. To obtain performance metrics of the CNN on the configured system and load profile, bus 12 experienced 470 attacks from the validated attack model to the WLS SE reported voltage phase angles with the same c_{inj} value in one emulation session that followed the time sequence defined in Fig. 1, but with the CNN executing at the same time as the WLS SE. The Attack Detection Probability (ADP) is defined as the number of detected attacks over the number of attacks performed during this experiment's simulations. Fig. 5 contains the ADP plot of the CNN after multiple emulation sessions where the c_{inj} varied for each session that bus 12 was attacked. To measure the false positive rate (FPR) of the

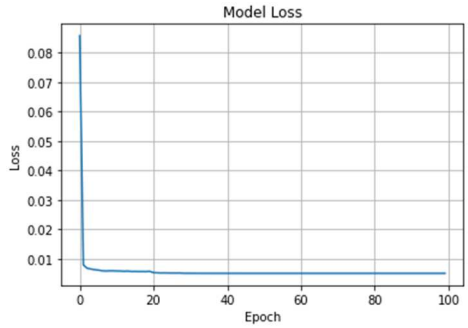

Figure 4. Plotted Loss Function for trained CNN model used for emulation in Typhoon HIL.

Table I. BDD Report of Bus 14 During $\widehat{\theta_{14}}$ Cyber-Attack

Quantity	Before FDIA	After FDIA
$\widehat{V_{14}}$	0.967 P.U.	0.967 P.U.
$\widehat{\theta_{14}}$	-14.85°	-13.71°
r	0.050	0.448

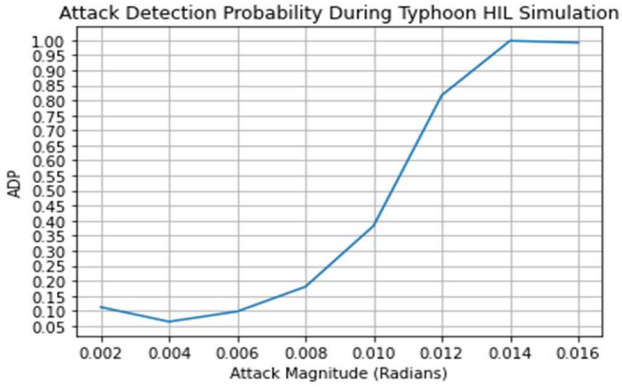

Figure 5. Attack Detection Probability Performance of the trained CNN model implemented into the IEEE 14-bus System with Typhoon HIL simulation.

CNN model, an emulation session was performed without attacking any bus and tracking the number of samples labeled as compromised by the model. The resulting FPR for this trained CNN model was 5.720% from a testing sample size of 472.

IV. Conclusion

Through the integration of both software and hardware in the Typhoon RTDS, we demonstrate its effectiveness in both the development and validation of attack models and defensive strategies, ensuring robustness before they are implemented in real-world scenarios. A noteworthy aspect of this paper is the practical application of a CNN in an RTDS environment. The CNN's effectiveness was evaluated under altered system conditions, employing a new set of readily available load profiles distinct from those used in the model's training phase. The simulation results are promising, particularly in the CNN's ability to detect larger FDI attack magnitudes, which are critical as their undetected presence could lead to significant consequences in the smart grid. This paper highlights the significant potential of the Typhoon HIL

testbed as a valuable tool in the realm of cyber-security for power systems.

REFERENCES

[1] G. Liang, S. R. Weller, J. Zhao, F. Luo, and Z. Y. Dong, "The 2015 Ukraine blackout: implications for false data injection attacks," IEEE Trans. Power Syst., vol. 32, no. 4, pp. 3317–3318, Jul. 2017.

[2] S. Karnouskos, "Stuxnet worm impact on industrial cyber-physical system security," in Proc. 37th Annu. Conf. IEEE Ind. Electron. Soc., Melbourne, VIC, Australia, 2011, pp. 4490–4494.

[3] Q. Zhao, X. Qi, M. Hua, J. Liu, and H. Tian, "Review of the recent blackouts and the enlightenment," IET Journals in CIRED 2020 Berlin Workshop, vol. 2020, iss. 1, 2020, pp. 312-314.

[4] M. Chlela, G. Joos, M. Kassouf, and Y. Brissette, "Real-time testing platform for microgrid controllers against false data injection cybersecurity attacks," IEEE Power and Energy Society General Meeting, Boston, MA, USA, pp. 1–5, Jul. 2016.

[5] L. Huiming and M. Rui, "Research and Implementation of Key Technology of HIL Architecture Design Based on xPC Real-Time Platform," IEEE 4th International Conference on Mechanical, Control and Computer Engineering, Hohhot, China, pp. 833-836, Oct. 2019.

[6] S. Vishnu, K. Pai, P. Krishna, N. S. Jayalakshmi, S. D. Suraj, and V. Prathimala, "Correlative Analysis of Dynamic Behaviour of Lithium-Ion Cell using MATLAB and Typhoon HIL," IEEE International Conference on Distributed Computing, VLSI, Electrical Circuits and Robotics, Nitte, India, pp. 225-230, Nov. 2021.

[7] A. W. Reis, F. G. K. Guarda, and C. C. Gastaldini, "Simulation of a Phasor Measurement Unit in Real Time Using Typhoon Virtual HIL," IEEE PES Innovative Smart Grid Technologies Conference – Latin America, Gramado, Brazil, Sept. 2019.

[8] H. Magnago, T. Guterres, F. Carnielutti, J. Massing, R. Vieira, and H. Pinheiro, "A Test Driven Design Approach to Benchmark Current Controllers for Grid-Tied Inverters," 20th Workshop on Control and Modeling for Power Electronics, Toronto, ON, Canada, Jun. 2019.

[9] M. E. M. Essa, J. V. W. Lotfy, and M. E. K. Abd-Elwahed, "Adaptive Neural Network Predictive Control Design for Hybrid Electric Vehicle with Hardware in the Loop (HIL) Verification," 17th International Computer Engineering Conference, Cairo, Egypt, pp. 118-123, Dec. 2021.

[10] B. Canaan, B. Colicchio, and D. Ould Abdeslam, "Experimental HIl implementation of RNN for detecting cyber physical attacks in AC microgrids," IEEE International Symposium on Power Electronics, Electrical Drives, Automation and Motion, Sorrento, Italy, pp. 958-963, Jun. 2022.

[11] N. Abroyan, "Convolutional and Recurrent Neural Networks for Real-time Data Classification," IEEE Seventh International Conference on Innovative Computing Technology, Luton, UK, Aug. 2017.

[12] H. Zhang, B. Liu, & H. Wu, "Net Load Redistribution Attacks on Nodal Voltage Magnitude Estimation in AC Distribution Networks," IEEE PES Innovative Smart Grid Technologies Europe, pp. 46-50, Oct. 2020.

[13] B. Liu, & H. Wu, "Optimal D-FACTS Placement in Moving Target Defense Against False Data Injection Attacks," IEEE Transactoins on Smart Grid, vol. 11, no. 5, pp. 4345-4357, Sept. 2020.

[14] J. Liang, L. Sankar, & O. Kosut, "Vulnerability Analysis and Consequences of False Data Injection Attack on Power System State Estimation." IEEE Transactions on Power Systems, vol. 31, no. 5, 3864–3872, Sept. 2016.

[15] M. Arjovsky, E. Brevdo, K. Divilov, E. Jensen, M. Mirza, A. Paino, M. Sayer, R. Stout, and W. Wu, "Convolutional Networks," in *Deep Learning,* T. Dietterich, Ed. Place of Publication: MIT Press, 2016, pp. 321-361.

[16] S. Wang, S. Bi, and Y. A. Zhang, "Locational Detection of the False Data Injection Attack in a Smart Grid: A Multilabel Classification Approach," IEEE Internet of Things, vol. 7, no. 9, 8218-8227, Sept. 2020.