Deep Learning-Based Cybersecurity-Enabled Fault and Attack Detection in Cyber-Physical Energy Systems: A Systematic Review

Remon Das[1], Akash Debnath[2], Abu Shufian[3]
[1]Renewable Infrastructure, Dominion Energy, VA, USA
[2]Operational Planning, Dominion Energy, VA, USA
[3]Department of Electrical & Electronic Engineering, American International University-Bangladesh
remondas4@gmail.com, akashaust11@gmail.com, shufian.eee@aiub.edu

***Abstract*—Deep learning (DL) has become a leading data-driven approach for cybersecurity-enabled fault and attack detection in cyber-physical energy systems. However, existing research remains fragmented across different domains. This paper presents a compact systematic review that unifies research across three major domains: power and smart grids, industrial control systems (ICS), and wind/renewable energy systems. A common DL-based cybersecurity monitoring, fault detection, and attack detection pipeline together with a threat taxonomy are established. Eight representative studies are reviewed, including approaches based on deep belief networks, ensemble learning, stacked autoencoders, CNN-based detectors, and CNN–LSTM architectures for fault and intrusion detection. An additional ten supporting studies published between 2022 and 2026 are included to provide broader context and recent developments. A cross-domain comparison of models, attack types, datasets, and performance shows that most detectors achieve over 95% accuracy on normal data, but their performance drops sharply, often to near zero, under targeted backdoor and adversarial attacks. Key challenges include limited data, weak robustness against attacks, real-time deployment constraints, and poor transferability across domains. A future roadmap is outlined toward federated, physics-informed, explainable, and cybersecurity-aware AI systems for critical energy infrastructure.**

***Index Terms*—Deep learning, Cybersecurity, Fault detection, Attack detection, Smart grid security, Industrial control systems security, Renewable energy security**

I. Introduction

Modern energy infrastructure increasingly operates as a cyber-physical system, where supervisory control and data acquisition (SCADA) systems manage power grids and industrial processes, while phasor measurement units, smart meters, and turbine sensors continuously generate high-frequency operational data [1]. Physical or cyber faults in these systems can cause blackouts, plant shutdowns, and safety incidents. Deep learning (DL) has emerged as the leading approach for detecting both classes of fault, owing to its ability to learn discriminative features directly from raw multivariate time series [2].

However, the literature has evolved in three largely separate silos. In smart grids, research emphasizes false data injection attack (FDIA) detection and fault diagnosis from measurement streams [3], with a growing adversarial machine learning thread examining backdoor [4] and evasion [5] attacks against grid DL models. In industrial control systems (ICS), the focus is anomaly and intrusion detection over SCADA process data [6], with recent work questioning detector robustness under concealment attacks [7], [8]. In wind and renewable energy, DL is applied to SCADA-based condition monitoring and component fault diagnosis [9].

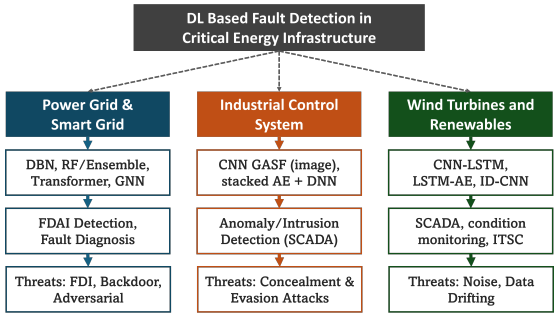

Fig. 1. Taxonomy of DL-based cybersecurity-enabled fault and attack detection across the three reviewed domains.

This review proposes a unified taxonomy and synthesized analysis of deep learning across domains, covering key studies, architectures, and comparative evaluations of attacks, datasets, and performance. It also identifies open challenges and outlines future research directions.

II. Background

A. Deep Learning Architectures for Cybersecurity-Enabled Fault and Attack Detection

Five major architecture families appear across the reviewed studies. Convolutional Neural Networks (CNNs) are widely used to extract local patterns from transformed time-series data for ICS detection. [6]. Recurrent neural networks (RNN), particularly LSTM and GRU models, are widely adopted to capture temporal dependencies in SCADA data. Additionally, autoencoders (AEs) model normal operating behavior and identify anomalies through reconstruction error, makes them effective, especially when labeled fault data are limited [10].

Extensions such as stacked AE–DNN frameworks further support multi-class attack classification [1]. Deep belief networks (DBNs) have also been applied to FDIA detection, often combined with metaheuristic optimization for feature selection and hyperparameter tuning [11].

B. Cybersecurity-Enabled Fault and Attack Detection Fundamentals

Fault detection can be performed using a supervised learning problem to classify faults (fault existence, fault class, fault location) or an unsupervised learning problem that aims to detect anomalies with regard to what has been learned about normal conditions. The classical workflow in Fig. 2 collects data via sensors or SCADA, applies preprocessing (normalization, feature selection, windowing), trains a DL model offline, and performs online fault detection that triggers alarms and operator responses [12].

C. Cyber Threats Against Deep Learning-Based Detection Systems

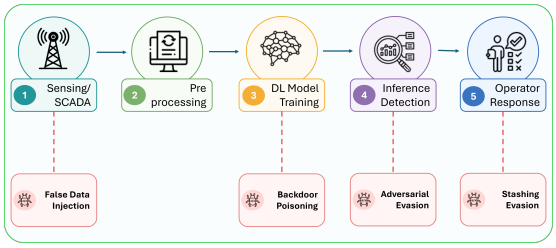

Fig. 2. DL fault-detection pipeline and attack injection points.

Recent physics-constrained variants generate poisoned samples that remain consistent with power-flow characteristics, making them more difficult to identify through standard data sanitization methods [4]. Adversarial evasion attacks modify input data during inference using methods such as FGSM, BIM, and DeepFool to alter fault predictions [5]. Concealment attacks target anomaly detection systems by manipulating observed process data to hide ongoing attacks, and model-independent variants have shown transferability across different detector architectures [7], [8].

III. Domain-Specific Review

A. Smart Grid Cybersecurity and Fault Detection

Alrslani et al. [11] propose ARDL-FDIAR for FDI attack detection which a hybrid framework that integrates Z-score normalization, a modified Lemurs optimization algorithm for feature selection, and an improved deep belief network optimized using a cetacean optimization strategy. This approach achieves a testing accuracy of 98.58% for FDIA detection and outperforms DNN, RF, and LSTM baselines. Complementing this metaheuristic-based direction by Li et al. [13] introduce a Transformer-based detector integrated with federated learning and Paillier homomorphic encryption, enabling utilities to collaboratively train FDIA models on IEEE 14- and 118-bus systems without sharing raw measurements. Han et al. [3] further model the spatial–temporal structure of power grids using a modified temporal multi-graph convolutional network. As a classical machine-learning benchmark, Redwan et al. [2] demonstrate that a tuned Random Forest ensemble trained on voltage and current features can still achieve 99.67% binary fault-detection accuracy and 99.81% fault-type accuracy, while also providing interpretable feature importance, serving as a useful baseline against which the added complexity of deep models should be justified.

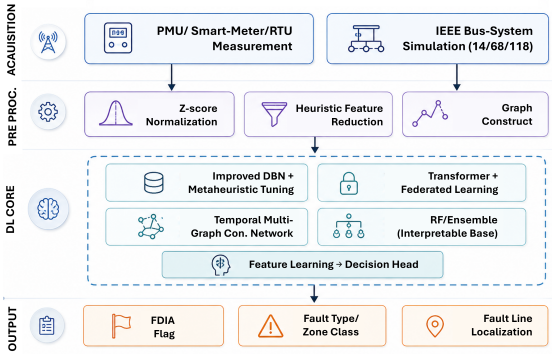

Fig. 3. Generalized smart-grid DL fault-detection architecture synthesized from the reviewed studies

Attacks on DL-based smart-grid fault-detection models reveal that strong performance does not necessarily translate into deployment robustness. Bai et al. [4] introduced the first physics-constrained backdoor poisoning attack for real-time fault-line localization, where triggers are designed to satisfy underlying power-system constraints. They achieved $> 98.4\%$ attack success on the IEEE 68-bus benchmark with only 0.04–9.31% clean accuracy loss across FCNN, CNN, and MSVM, making poisoned models hard to detect.

B. Industrial Control System Cybersecurity and Intrusion Detection

Wang et al. [1] introduce a stacked deep learning IDS for SCADA systems in which multiple deep models are combined to learn hierarchical representations from both power system and gas pipeline traffic. The method achieves 97.36% binary accuracy, 97.38% three-class accuracy, and 95.52% multi-class accuracy and consistently outperforms twenty standalone shallow and deep baselines, indicating the benefit of ensemble-based deep representations for heterogeneous ICS attack scenarios. On the other hand, Barbhaya et al. [6] focus on real-time deployment through a two-stage framework. Despite these advances, robustness remains limited. Erba and Tippenhauer [7] demonstrate that generic concealment attacks requiring no knowledge of the deployed detector can evade multiple model-free process-based anomaly detectors and that attacks designed for neural models transfer across different architectures. Fung et al. [8] further show that commonly used pointwise evaluation metrics such as point F1 can overestimate the practical effectiveness of reconstruction-based ICS detectors.

TABLE I
SUMMARY OF CYBERSECURITY-ENABLED FAULT AND ATTACK DETECTION LITERATURE

Ref.	Year	Domain	DL Model / Method	Dataset / Testbed	Key Reported Result
[4]	2025	Smart Grid	Improved DBN + MLOA Feature Selection + COA Tuning	FDIA benchmark (power system measurements)	98.58% accuracy and 98.57% recall for FDIA recognition
[3]	2026	Smart Grid	Random Forest / Ensemble Learning (DL-adjacent baseline)	Simulated power-system fault data (V/I features)	99.67% accuracy, 0.9995 ROC-AUC (binary), and 99.81% fault-type classification
[5]	2022	Smart Grid	Transformer + Federated Learning + Paillier Encryption	IEEE 14-bus and 118-bus systems	Privacy-preserving FDIA detection across distributed edge nodes
[6]	2023	Smart Grid	Modified Temporal Multi-Graph Convolutional Network	Smart grid measurement graphs	Enhanced FDIA detection through spatio-temporal graph modeling
[7]	2022	Smart Grid	Physics-Constrained Backdoor Poisoning vs. DL Fault Localizers	68-bus (86-location) fault localization system	Attack success rate >98.4%; clean accuracy reduction only 0.04–9.31%
[8]	2024	Smart Grid	FGSM/BIM/DeepFool Adversarial Attacks against CNN Fault Predictors	Fault zone and fault-type prediction tasks	BIM ($\epsilon = 0.04$) reduced prediction accuracy from 71.3% to approximately 5%
[1]	2022	ICS	Stacked Deep Learning (AE-Based Ensemble IDS)	Power system and gas pipeline SCADA datasets (MSU/ORNL)	97.36% binary, 97.38% three-class, and 95.52% multi-class detection accuracy
[9]	2025	ICS	TEDA Anomaly Detection + GASF Time-Series-to-Image CNN Classifier	Laboratory-scale ICS testbed	Real-time two-stage anomaly detection and attack-type classification
[10]	2022	ICS	Concealment Attacks vs. Model-Free Anomaly Detectors	SCADA process datasets (BATADAL-class and similar)	Transferable attacks successfully evade multiple detector architectures
[12]	2024	Wind	CNN-LSTM Multi-Fault Classifier	Wind farm SCADA data (four fault families)	90.06% overall accuracy; gearbox 97.39%, pitch 96.46%, and yaw 94.09%
[14]	2026	Wind	1D-CNN with Temporal Feature Engineering	Multi-source, multi-farm SCADA datasets	Early anomaly prediction across heterogeneous wind farms
[15]	2026	Wind	Unsupervised Autoencoder Trained on Healthy Data	Aeroelastic–electromechanical DFIG simulation	Detects medium- and high-severity stator ITSC faults; detection latency increases as fault severity decreases

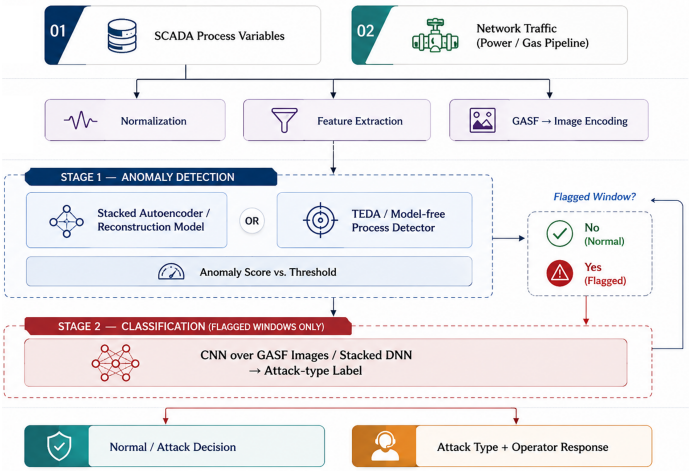

Fig. 4. Generalized ICS/SCADA DL detection architecture synthesized from the reviewed studies

C. Renewable Energy Cybersecurity and Fault Detection

Recent studies on wind turbine fault diagnosis highlight two dominant directions: supervised deep learning and unsupervised anomaly detection. Alagha et al. [14] identify SCADA-based deep learning as the prevailing approach while emphasizing data imbalance as a major challenge. Using SCADA time-series data, Qi et al. [15] and Zhang et al. [12] employ CNN–LSTM architectures for detecting multiple turbine faults and bearing failures, achieving strong diagnostic performance. For settings with limited labeled fault data, Espinoza et al. [10] propose an autoencoder trained only on healthy DFIG operation data, where faults are detected through reconstruction error, demonstrating reliable detection for moderate and severe faults but increased delay for low-severity cases. Figure 5 consolidates into a unified dual-paradigm framework that combines supervised (CNN–LSTM/1D CNN) and unsupervised (autoencoder) branches over shared preprocessing for integrated fault detection and early warning, while Table I indicates that high reported accuracies may not directly translate to real-world security performance.

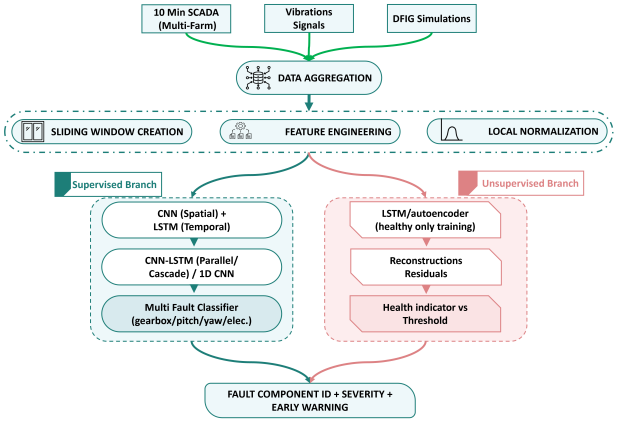

Fig. 5. Generalized wind/renewable DL detection architecture synthesized from the reviewed studies

Beyond physical component faults, renewable energy systems are increasingly exposed to cybersecurity threats targeting SCADA networks, remote monitoring platforms, sensor

measurements, and communication channels. Data manipulation, sensor spoofing, false data injection, and adversarial attacks may compromise fault detection performance and operational reliability. Despite significant progress in CNN-LSTM and autoencoder-based fault diagnosis, cybersecurity-oriented studies in renewable energy systems remain limited, highlighting an important research gap.

IV. Comparative Analysis

A. Models and performance

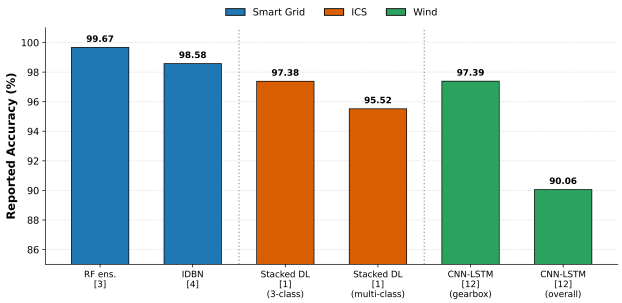

Fig. 6. Reported accuracy of representative detectors (values from the cited papers; tasks and datasets differ across bars)

Figure 6 highlights three trends: (i) all domains achieve $\geq 90\%$ benign accuracy, with smart-grid FDIA/fault detectors reaching 98–100% [11]; (ii) hybrid models (CNN-LSTM [15], stacked AE [1], DBN+metaheuristics [11]) outperform single architectures; and (iii) simpler ensembles [2] remain competitive on clean data, while deep models [10] mainly benefit raw or weakly labeled streams.

B. Cyberattack Impact on Deep Learning-Based Detection Systems

Figure 7 shows the contrasts of clean vs. under-attack accuracy where papers report both. Inference-time BIM evasion collapses grid fault classification from 71.3% to $\approx 5\%$ [5]; a physics-constrained backdoor flips $\geq 98.4\%$ triggered inputs while leaving clean accuracy within ≈ 0.04–9.31% the baseline [4] and generic concealment lets live ICS attacks pass model-free detectors entirely [7].

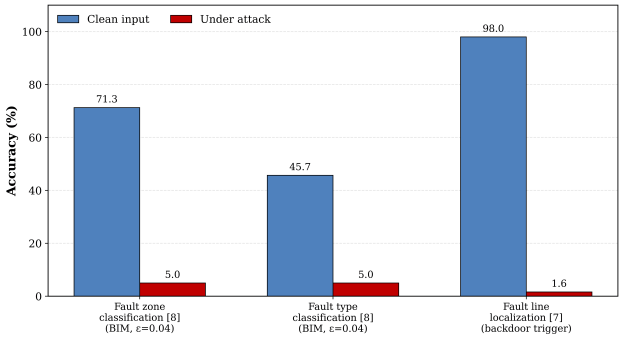

Fig. 7. Accuracy under attack vs. clean operation

C. Datasets

Table II shows that every domain leans on a small set of simulations and testbeds. Grid studies favor IEEE bus systems, which support physics-aware attack crafting [4], [13] and ICS work uses lab testbeds and the Mississippi State SCADA corpora [1], [6], [7]. Furthermore, wind studies use proprietary farm SCADA logs [9], [12], [15] or physics simulations [10], hindering reproducibility. No public benchmark spans more than one domain.

TABLE II
Datasets and Testbeds Used Across Reviewed Studies

Dataset / Testbed	Domain	Notes (used in)
IEEE 14-/118-bus	Smart Grid	FDIA detection with federated Transformer [13]
68-bus / 86 locations	Smart Grid	Fault line localization; backdoor study [4]
Simulated V/I fault data	Smart Grid	RF ensemble fault diagnosig [2]; FDIA corpus [11]
MSU/ORNL power & gas pipeline SCADA	ICS	Stacked DL IDS [1]
Lab-scale ICS testbed	ICS	TEDA + GASF-CNN real-time pipeline [6]
Water-system SCADA (BATADAL-class)	ICS	Concealment attack evaluations [8]
Wind farm SCADA logs	Wind	CNN-LSTM multi-fault [15]; bearing [12]; multi-farm 1D-CNN [9]
Aeroelastic DFIG simulation	Wind	Healthy-only AE for stator ITSC [15]

V. Challenges and Open Issues

TABLE III
Key Challenges and Mitigation Strategies

Challenge	Evidence and candidate mitigations
Data scarcity & class imbalance	Faults are rare; healthy-only AE training, GAN/augmentation, and transfer across farms mitigate but do not solve labeling cost.
Adversarial robustness	Backdoor, evasion, and concealment attacks defeat detectors; adversarial training, input sanitization, and physics-consistency checks are largely untested at scale.
Real-time / edge constraints	Two-stage designs (cheap anomaly stage + heavy classifier) [and compact 1D-CNNs trade accuracy for latency on resource-limited RTUs/PLCs.
Cross-domain gaps	No shared benchmarks; wind/RE detectors lack any robustness evaluation (Table II); metrics overstate utility.
Privacy & data sharing	Utilities cannot pool raw data; secure federated schemes add crypto/communication overhead.
AI Security & Trustworthiness	Backdoor attacks, data poisoning, adversarial examples, model theft, explainability, and secure AI deployment

Table III summarizes the main challenges found across the reviewed studies. The most important issue is evaluation. Most models are tested on clean data or simple replayed attack scenarios, while stronger attack methods assume more realistic and adaptive adversaries [4], [5], [7]. As a result, reported accuracy and F1 scores may not reflect real-world performance, and robustness testing is not yet a standard part of evaluation. Another key gap is in renewable energy systems. Although wind turbines and SCADA-based monitoring are widely studied, there is very limited work on how these models behave under adversarial attacks, especially for CNN-LSTM-based methods [9], [12], [15]. This creates an important

but largely unaddressed risk as these systems become more connected. Finally, real-time performance is often described in general terms rather than measured consistently.

VI. Future Directions

A. Federated deep learning

Future work should extend federated learning beyond FDIA to distributed ICS intrusion detection and wind-farm monitoring, while addressing robustness against poisoned clients and backdoor-style attacks [4], [13]. Physics-informed models can enforce system constraints to improve data efficiency and enable physics-based anomaly detection, though the same constraints may also be exploited for attacks [4], [16]. LLM-based tools offer promising support for monitoring, explanation, and mitigation via retrieval-augmented pipelines, but introduce new security risks requiring strong safeguards [17], [18]. Finally, improving explainability using methods like SHAP and attention remains critical, especially under evasion and concealment attacks where explanation reliability is still an open challenge [5], [7].

B. Trustworthy and Secure AI Architectures

Future fault-detection systems should adopt Zero Trust Architecture (ZTA), where every device, model, and communication channel is continuously verified rather than implicitly trusted. Secure federated learning can enable cross-operator collaboration while preserving data privacy and resisting poisoning attacks [11], [12]. Adversarially robust deep learning through adversarial training, certified defenses, and runtime validation should become standard to improve resilience against backdoor and evasion attacks [4], [11], [13]. In parallel, explainable security AI is needed to increase operator trust and support incident analysis, while AI governance frameworks should define requirements for transparency, auditing, accountability, and safe deployment of AI in critical infrastructure [17], [18].

VII. Conclusion

This review unified three previously siloed literatures on deep learning-based cybersecurity-enabled fault and attack detection in smart grids, industrial control systems, and renewable energy systems. Across eighteen studies from 2022–2026, we found a mature toolbox DBNs, stacked autoencoders, image-encoded CNNs, CNN-LSTM hybrids, Transformers, and graph networks delivering 90-100% accuracy on benign benchmarks, alongside mounting evidence that backdoor, adversarial, and concealment attacks reduce the same detectors to near-random behavior. The attack $\times$ domain matrix revealed that wind and renewable systems are essentially unstudied from a robustness standpoint and that no benchmark spans domains. We argue that the next phase of this field should treat robustness reporting as mandatory, build shared cross-domain benchmarks, and pursue federated, physics-informed, explainable, and carefully governed LLM-augmented architectures. For practitioners, the immediate lesson is simple: deploy DL-based cybersecurity and fault-detection systems behind physics-consistency checks and security-aware validation mechanisms.

References

[1] W. Wang, F. Harrou, B. Bouyeddou, S.-M. Senouci, and Y. Sun, "A stacked deep learning approach to cyber-attacks detection in industrial systems: application to power system and gas pipeline systems," *Cluster Computing*, vol. 25, no. 1, pp. 561–578, 2022.

[2] K. Redwan, M. M. R. Shamim, A. Sultana, D. H. Tuhin, N. Hannan, M. Ahmed, and A. Shufian, "Intelligent fault diagnosis in power systems using random forest and ensemble learning," *Measurement: Energy*, vol. 10, p. 100098, 2026.

[3] Y. Han, H. Feng, K. Li, and Q. Zhao, "False data injection attacks detection with modified temporal multi-graph convolutional network in smart grids," *Computers & Security*, vol. 124, p. 103016, 2023.

[4] J. Bai, R. Wang, and Z. Li, "Physics-constrained backdoor attacks on power system fault localization," in *2023 IEEE Power & Energy Society General Meeting (PESGM)*, pp. 1–5, IEEE, 2023.

[5] C. Ardito, Y. Deldjoo, T. Di Noia, E. Di Sciascio, F. Nazary, and G. Servedio, "Machine-learned adversarial attacks against fault prediction systems in smart electrical grids," *arXiv preprint arXiv:2303.18136*, 2023.

[6] M. Barbhaya, P. R. Dasari, S. K. Damarla, R. Srinivasan, and B. Huang, "A deep learning framework for cyberattack detection and classification in industrial control systems," *Computers & Chemical Engineering*, p. 109278, 2025.

[7] A. Erba and N. O. Tippenhauer, "Assessing model-free anomaly detection in industrial control systems against generic concealment attacks," in *Proceedings of the 38th annual computer security applications conference*, pp. 412–426, 2022.

[8] C. Fung, S. Srinarasi, K. Lucas, H. B. Phee, and L. Bauer, "Perspectives from a comprehensive evaluation of reconstruction-based anomaly detection in industrial control systems," in *European Symposium on Research in Computer Security*, pp. 493–513, Springer, 2022.

[9] M. M. Ata, S. Osama, M. R. Ibraheem, and A. R. Abas, "Early prediction of wind turbine anomalies using 1d-cnn and temporal feature engineering on multi-source scada data," *Scientific Reports*, 2026.

[10] V. Espinoza, C. Benalcázar-Parra, C. Tutivén Gálvez, and Y. Vidal Seguí, "Data-driven unsupervised stator interturn short-circuit fault detection for doubly fed induction generator wind turbines," in *Journal of physics. Conference series*, vol. 3224, Institute of Physics (IOP), 2026.

[11] F. A. Alrslani, M. A. Alohali, M. Aljebreen, H. Alqahtani, A. Alshuhail, M. Alshammeri, and W. S. Almukadi, "Enhancing cybersecurity via attribute reduction with deep learning model for false data injection attack recognition," *Scientific Reports*, vol. 15, no. 1, p. 3944, 2025.

[12] F. Zhang, Y. Zhu, C. Zhang, P. Yu, and Q. Li, "Abnormality detection method for wind turbine bearings based on cnn-lstm," *Energies*, vol. 16, no. 7, p. 3291, 2023.

[13] Y. Li, X. Wei, Y. Li, Z. Dong, and M. Shahidehpour, "Detection of false data injection attacks in smart grid: A secure federated deep learning approach," *IEEE Transactions on Smart Grid*, vol. 13, no. 6, pp. 4862–4872, 2022.

[14] N. Alagha, A. S. M. Khairuddin, Z. N. Haitaamar, O. Al-Khatib, and J. Kanesan, "Artificial intelligence in wind turbine fault detection and diagnosis: Advances and perspectives," *Energies*, vol. 18, no. 7, p. 1680, 2025.

[15] L. Qi, Q. Zhang, Y. Xie, J. Zhang, and J. Ke, "Research on wind turbine fault detection based on cnn-lstm," *Energies*, vol. 17, no. 17, p. 4497, 2024.

[16] B. Huang and J. Wang, "Applications of physics-informed neural networks in power systems-a review," *IEEE Transactions on Power Systems*, vol. 38, no. 1, pp. 572–588, 2022.

[17] S. Majumder, L. Dong, F. Doudi, Y. Cai, C. Tian, D. Kalathil, K. Ding, A. A. Thatte, N. Li, and L. Xie, "Exploring the capabilities and limitations of large language models in the electric energy sector," *Joule*, vol. 8, no. 6, pp. 1544–1549, 2024.

[18] J. Ruan, G. Liang, H. Zhao, G. Liu, X. Sun, J. Qiu, Z. Xu, F. Wen, and Z. Y. Dong, "Applying large language models to power systems: Potential security threats," *IEEE transactions on smart grid*, vol. 15, no. 3, pp. 3333–3336, 2024.