26th International Congress on Modelling and Simulation, Adelaide, SA, Australia, 30 Nov–4 Dec 2025
mssanz.org.au/modsim2025

AI-driven detection of false data injection attacks in OLTC-controlled microgrids

Sebastian Rao[a], Asef Nazari[a], Mohammad Ghasemi[a], Dhananjay Thiruvady[a], Rosemary Van Der Meer[a], Fatemeh Shiri[b]

[a] School of Information Technology, Deakin University, Geelong, 3216, VIC, Australia, [b]School of Computing Technologies, RMIT
Email: asef.nazari@deakin.edu.au

Abstract: Microgrids are becoming a widely adopted solution in modern energy infrastructure as they enable localized power generation and distribution with minimal reliance on central grids. Despite the apparent novelties of microgrids, their heavy reliance on real-time data exchange and automated control systems leaves them vulnerable to a range of cyberattacks aiming to destabilize critical systems and cause harm. One such threat is False Data Injection Attacks (FDIA), in which threat actors manipulate system input or output data without triggering alarms. This study investigates artificial intelligence (AI) driven solutions for detecting FDIAs targeting load demand state estimation as well as On-Load Tap Changers (OLTC), which are critical for secure and stable microgrid operations. Conventional microgrid security measures such as secure communication protocols, residual-based state estimation and rule-based detection often fall short against FDIA due to the adaptive and subtle nature of FDI, calling for a more comprehensive solution. Four supervised learning models, namely Random Forest, Support Vector Machine, k-Nearest Neighbors and XGBoost using gradient boosted decision trees (GBDT), were trained and tested against a simulated OLTC dataset comprised of normal and FDI data. Soft voting ensemble techniques were successfully applied to models that underperformed to boost their detection accuracy, comparable to the higher-performing models. Model performance was assessed using metrics including but not limited to accuracy, precision and recall. Our study indicates that supervised learning models, specifically tree-based models, are an effective means for detecting stealthy FDIAs as they exhibited the highest level of performance across all evaluation metrics. Overall, this work contributes to the development of lightweight, scalable AI-driven solutions for securing smart grid infrastructure against sophisticated attacks that circumvent traditional security measures.

Keywords: *False data injection, Microgrid security, On-load tap changer, Machine learning, Ensemble methods*

1. INTRODUCTION

Microgrids are an emerging solution for distributed energy in modern energy systems. By situating power generation and distribution closer to the point of consumption, these grids minimize transmission losses and enhance energy reliability. Microgrids are invaluable due to their dual operational modes, running autonomously and independently in island mode or connected to the main grid, which bolsters system resilience against grid disturbances and outages (IBM 2024).

At the centre of this innovation lies complex and interconnected communication and control systems. Microgrids are heavily reliant on real-time data exchange between devices such as IoT sensors, OLTCs, Supervisory Control and Data Acquisition (SCADA) systems, etc., and whilst they are essential for grid stability and balancing load demands, the reliance on this interconnectivity introduces significant cybersecurity challenges. A most serious and emerging threat in this domain is the FDIA, by which attackers carefully forge sensor data that closely mimics regular sensor readings to avoid detection, yet deviates subtly enough to cause operational faults. Given the real-time nature of microgrids, such attacks can cause severe disruptions, cascading failures, and physical damage (Qu 2022). Research on the impacts of FDIA and relevant security solutions pertaining to OLTCs is lacking in existing studies. OLTCs are important as they dynamically adjust transformer taps to regulate voltage and are particularly susceptible to FDIAs. If OLTCs are disrupted and misled by false sensor readings, it could cause system-wide instability across the grid. Choeum et al. (2019) demonstrated the effects of FDI in a simulated IEEE 33-bus network, where optimized FDIAs bypassed Volt-VAR logic, resulting in abnormal voltage profiles that went undetected. The outcome of this simulation is indicative of inadequate security frameworks, emphasizing the need for enhanced detection algorithms.

In response to the increasing popularity of microgrid energy solutions and the significance of OLTCs in regulating and maintaining voltage profiles and grid stability, this research investigates the employment of supervised AI models for real-time FDIA detection to safeguard OLTC operations. The main focus of this work is placed on developing AI-based detection systems and evaluating their ability to classify and distinguish between normal operational data and attack data. Experimentation with supervised AI models consisted of standalone model testing, followed by ensemble model testing in an attempt to enhance the underperforming models. Key evaluation metrics were employed for assessing model performance, such as accuracy, F1 score, and false positive rates, amongst others. By focusing specifically on OLTC-specific vulnerabilities, this study provides a tailored and scalable solution for addressing this frequently overlooked aspect of microgrid cybersecurity.

2. LITERATURE REVIEW

FDIAs are particularly effective against microgrid systems as their dependence on real-time data exchange and control automation leaves them vulnerable to data poisoning. FDIAs work discreetly through the injection of adverse data that mimics genuine sensor readings or control signals. These attacks have the potential to diminish performance, cause voltage instability, and cascading system faults, all whilst remaining undetected by traditional security protocols (Reda et al., 2022). This section explores existing literature on microgrid-specific cyberattack vulnerabilities, current conventional security methods, and recent research into the adoption of AI-driven solutions to protect microgrids.

2.1. Nature and Impact of FDIA

FDIAs are stealthy and impose severe consequences if not addressed promptly. Liu et al. (2011) highlighted that attackers with insider knowledge regarding system configuration could forge false data values that appear legitimate to the system, evading residual-based estimation checks. This attack becomes even more critical in microgrids when we consider their decentralized nature and limited sensor redundancy. Sahoo et al. (2019) discovered that FDIAs can remain dormant and cause delayed implications on voltage stability and current sharing by maintaining regular voltage levels during early compromise, slowly degrading system health over time. Various other studies have gauged the impact of FDIA on microgrid systems. In Das (2024), even 10–30% FDI (low-level FDI) can increase energy cost and cause load mismatches. Jo et al. (2025) show that FDIAs targeting secondary control outputs in DC microgrids lead to voltage instability and unequal power distribution. In the study conducted by Cao et al. (2022), FDIAs on distributed controllers disrupt consensus values in voltage and reactive power. Collectively, current research has shown the devastating impacts of FDIA on microgrid systems, calling for overarching solutions.

2.2. Conventional Detection Methods and Limitations

Conventional FDIA detection and security measures notably comprise of residual-based state estimation and secure communication protocols. Residual methods compare sensor values to predicted model values, raising alarms when there are discrepancies between actual data and predicted data outputs. Despite this, Intriago et al. (2023) states that threat actors with system specific knowledge can inject false data that remains within residual-based thresholds, avoiding detection. Communication protocols such as the Distributed Network Protocol 3 (DNP3) and Modbus provide network-level defence, but they are inadequate due to notable weaknesses: (i) Early versions of DNP3 which could still be adopted in legacy systems, lacked authentication and encryption, allowing attackers to intercept and spoof commands. Even with up-to-date versions of DNP3, this protocol lacks data validation (Jamil et al., 2021). (ii) Modbus is utilized widely, yet it lacks native encryption and is susceptible to man-in-the-middle attacks (MITM). In addition, firewalls play a pivotal role in regulating network traffic, however its effectiveness is minimal in large scale systems. Jamil et al. (2021) indicate that commercial-grade firewalls aren't readily scalable, typically requiring hundreds of static rules. Static and complex configurations struggle to handle adaptive threats and rule abiding insider attacks. Whilst traditional methods provide some security, they fall short in dealing with subtle FDIAs designed to imitate genuine system operations.

2.3. AI and Machine Learning-Based Detection

Unlike conventional security protocols, AI-driven detection algorithms can analyze behavioural patterns in real time. Supervised models rely on labelled datasets to distinguish between safe or malicious inputs. These models are said to perform well in energy distribution systems if sufficiently labelled data is available (Sayghe et al. 2020). In instances where labelled data is scarce, a typical issue in FDIA scenarios, these models will struggle. Unsupervised models could be considerably effective in dealing with FDIA cases where labelled data is

unavailable, such as Fuzzy C-Means clustering, which can identify deviations in established patterns. This approach is suitable for in microgrid scenarios where confirmed attack data is unknown (Pinto et al., 2023).

Meanwhile, Deep Learning models such as Multi-Layer Perceptrons and Recurrent Neural Networks can learn nonlinear and temporal relationships in data. A drawback of these models is the need for large scale and diverse datasets which aren't readily available for simulated experiments (Chalapathy and Chawla, 2019). It is apparent that AI driven detection is apt for the dynamic and distributed nature of microgrids and adaptive FDIAs. In contrast to traditional static detection, the adaptability of AI in accordance with changing system behaviour and evolving FDIA patterns poses them as a suitable avenue for robust microgrid security.

2.4. Research Gaps and OLTC Vulnerability

While several studies have explored the application of AI-driven detection for FDIAs targeting state-estimation in AC and DC microgrids, existing literature overlooks FDIAs targeting OLTCs. As noted by De et al. (2017), OLTCs are essential for voltage profile preservation and ensures adequate system operation. If targeted with FDI, OLTCs may be deceived into making erroneous tap changes, causing disruptive voltage fluctuations that disturbs the operation of the entire grid. This critical gap in current research underscores the need for research dedicated to the impact of FDIA on OLTCs, and possible solutions to counteract it.

3. METHODOLOGY

This research employs a multi-method approach consisting of exploratory data analysis of a simulated OLTC dataset, simulated AI experiments, and quantitative performance evaluation. These three methods in conjunction, facilitated effective model selection and evaluation of performance in detecting FDIAs targeting OLTCs. The synthetic OLTC dataset simulated both normal and attack scenarios whilst preserving class imbalance by which normal data is more abundant than attack data, allowing for realistic simulation.

3.1. Research Design

The three-pronged research methodology is as follows: (i) Exploratory Data Analysis (EDA), (ii) Simulation-Based Experimentation, and (iii) Quantitative analysis. EDA is Correlation Matrices, Principal Component Analysis (PCA), and t-distributed Stochastic Neighbor Embedding (t-SNE) visualizations were employed to analyze the dataset's features and characteristics, providing informed model selection. Simulation-Based Experimentation are AI models were trained and tested against FDIA scenarios in a controlled simulation to emulate real-world detection performance. Quantitative analysis is defined as Performance assessment of supervised classification models using typical AI performance metrics such as accuracy, precision, recall, F1-score and false positive rate (FPR).

3.2. Dataset Description

The synthetic OLTC dataset was designed to closely mimic sensor data under both normal and FDIA-influenced scenarios. The dataset contains a total of 6000 samples, with 100 numerical characteristics to represent electrical measurements from OLTC operations. Each sample is labelled as "OK" (normal) or "FDI" (attack). Realistic imbalance found in real-world FDIA scenarios is present in the class distribution, with 5000 "OK" and only 1000 "FDI" samples. The data imbalance is crucial for realistic experimentation as FDIAs are subtle and discrete, only injecting small amounts of poisoned data to avoid detection. Whilst the features are synthetic and unlabelled, this dataset was designed to replicate the variability of voltage, current, and load parameters under different tap positions, as observed in real OLTC operations.

3.3. Evaluation Metrics

Several classification measures were used to assess model performance. These metrics are calculations derived from the confusion matrix where True Positives (TP) refers to correctly identified FDIA cases, True Negatives (TN) refers to correctly identified normal cases, False Positives (FP) refers to normal cases being incorrectly identified as FDIA, and False Negatives refers to FDIA cases being incorrectly identified as normal. The performance evaluation metrics are as follows:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (4)$$

$$FPR = \frac{FP}{FP+TN} \quad (5)$$

Where, accuracy refers to the overall proportion of correct predictions, while precision calculates the percentage of successfully predicted FDIA cases among all FDIA predictions. Recall is the proportion of true FDIA incidents that were accurately detected. F1-score is defined as the harmonic mean of precision and recall, which is particularly useful in skewed datasets, and the frequency whereby regular OLTC behaviour is misclassified as an attack is denoted by FPR.

3.4. Exploratory Data Analysis and Model Selection

Initial detection tests involved the use of a simple Logistics Regression classifier, which produced inadequate results. This indicated that the dataset is complex and non-linearly separable and that EDA was required to aid in appropriate model selection. Below are the EDA visualisations depicting the complexity and non-linearity of the dataset (see figures 1-3). Figure 1 represents the correlation matrix, explaining that most features have weak linear relationships, implying that there is a lack of strong redundancy or linear dependency. Figure 2 depicts the PCA visualization, which shows high overlap between normal and attack samples, indicating poor class separation in reduced linear dimensions. Figure 3 depicts the t-distributed Stochastic Neighbor Embedding (t-SNE) view, which also demonstrates significant intermixing of class labels and the absence of well-defined clusters, confirming the data's non-linear nature.

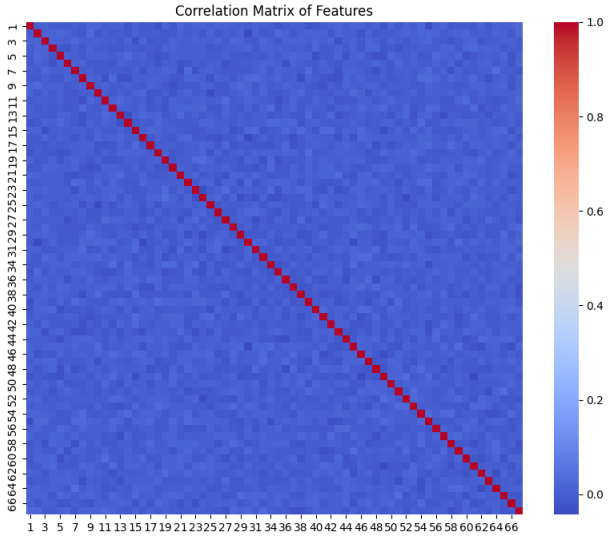

Figure 1: Correlation matrix showing minimal linear relationships between OLTC features

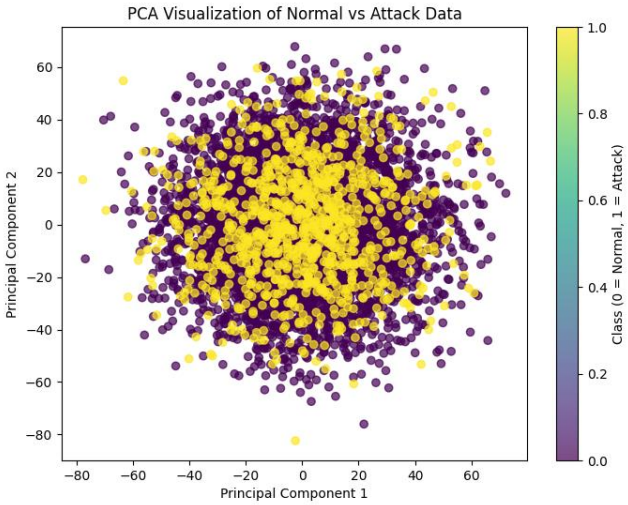

Figure 2: PCA projection showing overlap between normal and FDIA data, indicating weak linear separability

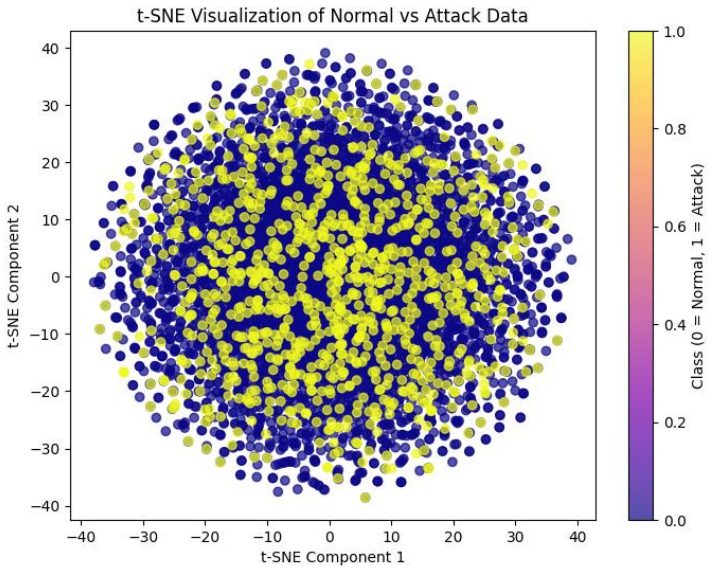

Figure 3: t-SNE visualization highlighting intermixing of normal and FDIA instances and non-linear complexity

These EDA results suggested that more models capable of capturing nonlinear patterns were required. Based on the EDA results, the following supervised learning models were chosen for experimentation: (i) Random Forest is a decision tree model that is resilient to noise and well-suited for high-dimensional data, (ii) GBoost **(**GBDT**)** is a gradient boosting approach that iteratively corrects faults and works well with imbalanced datasets, (iii) support vector machines (SVM) with radial basis function (RBF) Kernel transforms data into a higher-dimensional space to improve separation, and (iv) k-Nearest Neighbors (k-NN) identifies local structure and non-parametric behaviour. Underperforming models were later enhanced by integrating them into a soft voting ensemble, bolstering accuracy and robustness.

3.5. Model Implementation and Benchmarking

All models were developed using a consistent set of standards to ensure fairness and comparability. Label encoding was applied to convert class labels into a binary format. Feature scaling was performed using z-score normalization. The dataset was split into training and testing sets with a 70/30 ratio, and stratified sampling was used to keep the classes balanced. For class balancing, SMOTE was applied only to the training set to oversample the minority FDIA class, ensuring realistic testing conditions. Finally, validation checks included shuffle-label sanity tests to confirm that no data leaked between the training and testing datasets.

3.6. Model Configuration

Each individual model was modified to allow for the best possible performance and to facilitate consistent evaluation of each classifier under realistic FDIA settings. For Random Forest, the configuration used 100 estimators with class weight balancing. XGBoost was set with a log-loss metric and reduced verbosity for clarity. The SVM model employed an RBF kernel with default regularization. Finally, k-NN was configured with five neighbours and Euclidean distance.

4. EXPERIMENTAL RESULTS AND DISCUSSION

This section presents the experimental results of the chosen supervised learning models (Random Forest, XGBoost with GBDT, SVM, and k-NN), as well as the results for the ensemble SVM and k-NN models due to their underperformance as standalone models. These models were evaluated based on their ability to detect FDIA data in OLTC-controlled microgrid systems and were tested with an imbalanced dataset to simulate real-world class imbalance. Performance was measured using five key metrics: Accuracy, Precision, Recall, F1 Score, and FPR.

4.1. Benchmark Model Performance

Table 1 summarises the overall performance of the five standalone models. Observably, the top-performing models were the tree-based models (Random Forest and XGBoost with GBDT). They achieved superior results across all five performance metrics, such as accuracy ratings above 99.7% and FPRs below 0.21%. These results indicate that tree-based algorithms have a higher competency in handling complex, overlapping feature

distributions. In contrast, SVM achieved adequate results, with an accuracy of 97.33%; however, it has a FPR of 1% which isn't ideal in such a critical application. The k-NN model struggled the most with distinguishing normal data and attack data, with an accuracy of 58.17% and a FPR of 49.33%.

Table 1: Performance of Benchmark FDIA Detection Models

Model	Accuracy	Precision	Recall	F1	FPR
Random Forest	**99.83**%	**99.34**%	**99.67**%	**99.50**%	**0.13**%
XGBoost (GBDT)	99.78%	99.01%	**99.67**%	99.34%	0.20%
SVM	97.33%	94.68%	89.00%	91.75%	1.00%
k-NN	58.17%	27.95%	95.67%	43.26%	49.33%

These findings indicate that tree-based models are more suitable for this application as they achieved near-perfect classification, whilst SVM and k-NN models (more notably k-NN) could benefit from tuning or soft voting ensemble integration. The following subsection outlines the ensemble results for SVM and k-NN.

4.2. Ensemble Model Enhancement

To bolster the performance of the underperforming models (SVM and k-NN), they were integrated into a soft voting ensemble that combined them with the Random Forest and Logistic Regression classifiers. The soft voting ensemble works by training each model separately, then during testing, the predicted probabilities of each model are averaged to come to a final decision. Table 2 summarizes the performance of the ensemble integrations of the SVM and k-NN models. The SVM ensemble achieved a slight accuracy improvement, increasing the accuracy from 97.33% to 98.5%. More notably, the FPR dropped from 1% to 0.33%. As for the k-NN model, the ensemble version increased the accuracy from 58.17% to 98.17%, and the FPR decreased from 49.33% to 1.87% (not ideal, yet a significant improvement from the standalone model). Holistically, these ensemble integrations for both SVM and k-NN enhanced these models so much so, that they are closely comparable to the two tree-based classifiers (the top performers).

Table 2: Performance of Ensemble FDIA Detection Models

Model	Accuracy	Precision	Recall	F1	FPR
SVM	**98.50**%	**98.23**%	92.67%	**95.37**%	**0.33**%
k-NN	98.17%	91.33%	**98.33**%	94.70%	1.87%

These results indicate that soft voting ensemble algorithms may be a robust and effective solution for detecting FDIA accurately with low false positives in the context of OLTCs.

4.3. Key Insights

Amongst all tested models, Random Forest and XGBoost with GBDT displayed near-perfect classification consistently across all five metrics, showcasing their suitability for real-time FDIA detection. Their capacity to handle complex nonlinear feature correlations makes them ideal for detecting discrete attacks that closely mimic regular operations. Additionally, the ensemble configurations turned otherwise limited models into viable solutions for this application. The soft voting mechanism within the ensemble models played a pivotal role in lowering the FPR of the standalone versions, a primary concern for real-time security in smart grid systems, showcasing the benefits of hybrid AI configurations in FDIA detection for OLTCs.

4.4. Key Takeaways

- Random Forest and XGBoost were the top performers, achieving high accuracy and low FPRs.
- Hybrid AI models (specifically soft voting ensembles) significantly enhanced underperforming models (SVM and k-NN).
- Results show that ensemble models are a promising direction for developing robust and scalable security solutions for FDIA detection in OLTC microgrid systems.

5. CONCLUSION AND FUTURE WORK

This research explored the application of AI-driven detection for detecting the presence of FDIAs in microgrids, specifically focusing on OLTC operations. Through rigorous training, testing, and evaluation of four different models on a synthetic OLTC dataset, this study presents a viable solution for detecting emerging and adaptive FDIAs that threaten the safety of microgrids. Random Forest and XGBoost proved to be the most reliable models, achieving near-perfect classification, whilst soft voting ensemble methods for SVM and k-NN

confirmed the effectiveness of hybrid model approaches in boosting detection capabilities of weaker models. This study provides a targeted framework for detecting FDIA against OLTC systems, addressing a largely overlooked research direction in the context of microgrid security. It also demonstrates the capabilities of hybrid AI approaches in enhancing classifiers that perform inadequately in FDIA settings. Additionally, it employed a reproducible methodology for evaluating model performance against FDIA detection. Despite its contributions, this research has significant limitations. The use of a synthetic OLTC dataset, whilst closely reflecting real-world FDIA imbalance, can't fully replicate the variability and complexity of live OLTC data streams. Additionally, this study focuses specifically on the detection of FDIA and does not address the mitigation of FDIAs. Lastly, this work only focuses on four supervised learning models, overlooking the possible advantages of alternative supervised learning models and the exploration of unsupervised learning and anomaly detection algorithms. Future work should address these limitations by utilizing real-world OLTC data to evaluate the effectiveness of AI-driven detection within live microgrid environments. Furthermore, it's worth considering and exploring the application of unsupervised learning and anomaly detection as they could distinguish between attack and normal data without the need for labelled data (labelled data is scarce in FDIA scenarios). Ultimately, further studies should consider mitigation strategies that synergise with FDIA detection to form a complete threat detection and response framework.

REFERENCES

Cao G, Jia R and Dang J (2022) Distributed resilient mitigation strategy for false data injection attack in cyber-physical microgrids. Frontiers in Energy Research 10. https://doi.org/10.3389/fenrg.2022.845341

Chalapathy R and Chawla S (2019) Deep learning for anomaly detection: A survey. arXiv preprint arXiv:1901.03407. Available: https://arxiv.org/pdf/1901.03407 [Accessed: Aug. 11, 2025].

Choeum D and Choi D-H (2019) OLTC-induced false data injection attack on Volt/VAR optimization in distribution systems. IEEE Access 7, 34508–34520. https://doi.org/10.1109/access.2019.2904959

Das S (2024) Modeling and mitigating power grid vulnerabilities: A comprehensive analysis of renewable energy integration, cascading failures, and microgrid resilience. VCU Scholars Compass. https://doi.org/10.25772/Q6SF4M59

De J, Karady GG and Zhang Y (2017) Analysis and design of an electronic on-load tap changer distribution transformer for automatic voltage regulation. IEEE Transactions on Industrial Electronics 64(1), 883–894. https://doi.org/10.1109/tie.2016.2592463

IBM (2024) What is a microgrid? IBM. Available: https://www.ibm.com/think/topics/microgrid [Accessed: Mar. 24, 2025].

Intriago A, Liberati F, Hatziargyriou ND and Konstantinou C (2023) Residual-based detection of attacks in cyber-physical inverter-based microgrids. IEEE Transactions on Power Systems 39(2), 4020–4038. https://doi.org/10.1109/tpwrs.2023.3286019

Jamil N, Qassim QS, Bohani FA, Mansor M and Ramachandaramurthy VK (2021) Cybersecurity of microgrid: State-of-the-art review and possible directions of future research. Applied Sciences 11(21), 9812. https://doi.org/10.3390/app11219812

Jo S-B, Tran DT, Nguyen HX, Kim M and Kim K-H (2025) Hybrid control strategy for DC microgrid against false data injection attacks and sensor faults based on Lagrange extrapolation and voltage observer. Electronics 14(6), 1087. https://doi.org/10.3390/electronics14061087

Liu Y, Ning P and Reiter MK (2011) False data injection attacks against state estimation in electric power grids. ACM Transactions on Information and System Security 14(1), 1–33. https://doi.org/10.1145/1952982.1952995

Pinto SJ, Siano P and Parente M (2023) Review of cybersecurity analysis in smart distribution systems and future directions for using unsupervised learning methods for cyber detection. Energies 16(4), 1651. https://doi.org/10.3390/en16041651

Qu Z (2022) False data injection attacks. Encyclopedia.pub. Available: https://encyclopedia.pub/entry/21304 [Accessed: Mar. 24, 2025].

Reda HT, Anwar A and Mahmood A (2022) Comprehensive survey and taxonomies of false data injection attacks in smart grids: attack models, targets, and impacts. Renewable and Sustainable Energy Reviews 163, 112423. https://doi.org/10.1016/j.rser.2022.112423

Sahoo S, Mishra S, Peng JC-H and Dragicevic T (2019) A stealth cyber-attack detection strategy for DC microgrids. IEEE Transactions on Power Electronics 34(8), 8162–8174. https://doi.org/10.1109/tpel.2018.2879886

Sayghe A, Jin M, Kalsi K, Zhu H, Khan B and Liu Y (2020) Survey of machine learning methods for detecting false data injection attacks in power systems. IET Smart Grid 3(5), 581–595. https://doi.org/10.1049/iet-stg.2020.0015

