

Detection of False Data Injection Attacks (FDIA) Targeting Distribution Transformers in Smart Grids Using Hybrid Artificial Intelligence Methods

Recep AKKAN ^{1,*}, Osman Can ÇETLENBİK ²

¹ Department of Electronics and Automation, Kütahya Dumlupınar University, Kütahya, 43500, Türkiye

² Department of Statistics, Manisa Celal Bayar University, Manisa, 45140, Türkiye;
osman.cetlenbik@cbu.edu.tr

* Correspondence author: recpakkan@gmail.com

Abstract: Communication technologies in modern power systems have rendered energy grids vulnerable to False Data Injection Attacks (FDIA). This study evaluates FDIA detection on the IEEE 118-bus test system using pandapower. A synthetic dataset of 2000 samples was partitioned into 75%/25% training/test splits. Three models — LSTM networks, Isolation Forest and One-Class SVM — were comparatively assessed. LSTM achieved the highest performance with 97.0% recall, F1-score of 0.965 and AUC of 0.987, outperforming Isolation Forest (93.0%) and One-Class SVM (90.0%). These results confirm that deep learning approaches are highly effective for smart grid cybersecurity and compatible with IEC 61850-based infrastructures.

Keywords: FDIA; cyber-attacks; LSTM; anomaly detection; smart grids

1. Introduction

The role of smart grids, including adaptive protection and communication control, is becoming increasingly important in modern power systems [1,2]. The aim of smart grids is to ensure an uninterrupted and secure energy supply. Therefore, energy systems within smart grids are socially critical infrastructures [3]. While the sustainability of smart grids is important, their security is equally crucial.

Security in power systems is divided into two main categories: physical security and cyber security [3]. Physical security is defined as the ability to maintain steady-state conditions even under the influence of unexpected major faults or external interference. Cyber security is defined as ensuring the confidentiality and integrity of the information infrastructure, communication channels, and data flow between them that run the operational processes of power systems. The high integration of technological advancements into modern energy power grids has exposed modern power systems to a new generation of security threats. The widespread use of Internet of Things (IoT) devices, smart meters, and remote monitoring systems is transforming power grids into increasingly complex and vulnerable digital networks [4,5]. This situation not only threatens cybersecurity but can also jeopardize the physical security of the system [6–8].

It was once thought that cyber-attacks only affected computers and would not impact industrial systems. However, cyber-attacks have the potential to cause permanent damage by sabotaging the control systems underlying energy infrastructure. The widespread power outage in Ukraine serves as an example of the immense damage a cyber-attack can cause. Another example is the cyber-attack on the computer network at the Davis-Besse nuclear power plant in the U.S. in 2003 [3,6]. Today, cyber-attacks on smart grids occur in the form of data spoofing attacks, man-in-the-middle attacks, packet analysis attacks, distributed denial of service attacks, and FDIA attacks [6,9]. Among these, FDIAs targeting SCADA systems are the most common. Furthermore, it is difficult to directly detect FDIAs carried out on SCADA systems.

Distribution transformers are among the most critical components of the power grid and their

continuous monitoring is ensured by the International Electrotechnical Commission (IEC) 61850 protocol. The lack of authentication in GOOSE (Generic Object Oriented Substation Event) messaging and the use of unencrypted communication channels in the Manufacturing Message Specification (MMS) protocol allow attackers to inject falsified measurement data [10–12].

In this study, four different FDIA scenarios were designed using the pandapower library implemented on the IEEE 118-bus standard test system and a comprehensive attack model based on the real vulnerability surfaces of the IEC 61850 protocol was developed. LSTM, Isolation Forest and One-Class SVM methods were used in this study. The performance of the methods mentioned was compared with a 75%/25% training/test split. The findings were interpreted from the perspective of IEC 62351 safety standards.

Figure 1 presents the IEC 61850-based substation communication architecture and the FDIA attack surface considered in this study.

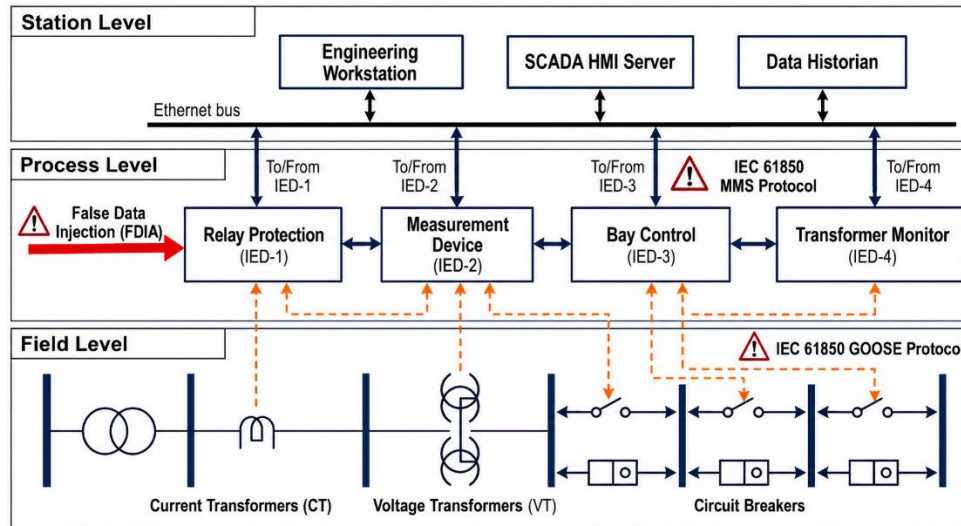


Figure 1. IEC 61850-based intelligent substation communication architecture illustrating the interaction between IEDs, SCADA infrastructure, and FDIA attack surfaces through vulnerable GOOSE/MMS communication channels

There are many studies in the literature on detecting cyber-attacks that manipulate measurement data (current, voltage etc.) in electricity distribution systems to misrepresent the system's status. A study has proposed a method that takes uncertainties into account as opposed to classical estimation methods. In the proposed interval DSSE (Distribution System State Estimation) method the variables in the distribution system are defined not as a single numerical value but as a range with lower and upper limits. This ensures that the inherent uncertainties in the system are also taken into account. The improved iterative algorithm (IIA) was used to calculate the lower and upper limit values. In the study which used a three phase modeled system the proposed method was proven to yield more accurate results when compared to traditional methods such as Monte Carlo simulation [13].

In a study based on security vulnerabilities in the IEC 61850 protocol, an anomaly detection methodology specific to the IEC 61850 protocol was developed. The developed methodology uses deep feature extraction based on the Discrete Wavelet Transform (DWT), dimensionality reduction with improved Locally Linear Embedding (LLE) and anomaly detection with an LSTM based Autoencoder. The results of experiments conducted in the smart substation prove that the proposed method demonstrates fast and satisfactory detection performance [14].

In another study, a novel model based on the Isolation Forest algorithm was proposed for the protection of industrial control systems (ICS) which unlike traditional methods can quickly identify deviations in complex data. Both known and emerging cyber threats were detected with higher accuracy and efficiency in this model which was tested with rational data (HIL-based Enhanced ICS dataset). In the study conclusion emphasizes that this machine learning-based approach proposed is a powerful alternative for strengthening industrial safety infrastructure [15].

In a related work the static state estimation problem in smart grids under unusual noise conditions was investigated. Typically, such analyses assume that the noise follows a Gaussian distribution. However this study focuses on real world scenarios where the noise is impulsive. The study also focused

on detecting malicious data injection attacks targeting these systems. In this study the α -stable distribution model was used for noise modeling and robust filters, which are more resistant to impulsive noise were preferred over the standard least squares method (LS) for the estimation method. The cumulative sum (CUSUM) technique was applied to quickly identify data injection attacks targeting buses. In conclusion, it has been observed that robust filters yield much more successful results than the classical LS method in environments containing impulsive noise. In the CUSUM technique which is proposed as an intrusion detection method it was concluded that there is an inverse relationship between the speed of intrusion detection and the accuracy of the detection [16].

A complementary study on FDIA detection in smart grids opted for a federated learning approach aiming to create a decentralized security layer that can detect attacks with high accuracy without centralizing data from different regions unlike traditional methods. The study uses a hybrid and multi stage detection framework. Random Forest and ExtraTrees classifier models were used for data classification. LSTM and GRU (Gated Recurrent Unit) models were used to analyze time series data. To improve the performance of the models used bidirectional GRU and LSTM structures supported by attention layers were employed. This proposed method has achieved a remarkably high accuracy rate of 99.8% in detecting attacks [17].

Similarly a study on the FDIA which targets energy systems by bypassing traditional defense systems, presented a detection framework that can find these attacks even at the single data bus level by decoupling the connections between complex data paths. The framework presented in this study utilizes a hybrid structure consisting of three powerful techniques to analyze complex data. These include Variational Mode Decomposition (VMD), Fast Independent Component Analysis (FastICA), Extreme Gradient Boosting (XGBoost) and Particle Swarm Optimization (PSO). In this study threshold value optimization was performed using the PSO algorithm. The simulation of the proposed method was carried out on the IEEE 14-bus test system. As a result attacks were detected with high accuracy. This led to the conclusion that the system remains robust even under varying noise levels and attack densities and that it is interpretable with the support of the PSO algorithm [18].

The issue of FDIA attack detection which aims to subtly manipulate data from smart meters is also emerging as a research area. A study discussed the disadvantages of existing defense systems such as "ignoring spatial relationships" and "insufficient robustness". To address this problem the authors proposed an unsupervised deep learning model they named ReVGAE (Recursive Variational Graph Autoencoder). In the proposed method since Graphical Neural Networks (GNNs) treat the physical structure of the network as a graph the spatial relationship between the points is preserved. In the proposed method the VGAE module has the roles of reconstructor, noise reduction and recursion. In this study FDIA detection is calculated using an anomaly score, which is the difference between the observed actual data and the data that the model says "should be". If the difference is high it indicates that the FDIA has intervened at that point. Ultimately the proposed method not only detects that an attack has occurred but also identifies the exact location where the attack was carried out. This shows that the proposed method yields more accurate results than other existing variational autoencoder and GNN based systems [19].

Building on this direction a novel artificial intelligence based model was presented for FDIA detection in smart grids. Researchers have proposed a Dual Attention Spatio Temporal Fusion Network model, which they call DASTFN, that combines two fundamental artificial intelligence architectures: Graphic Attention Networks (GATs) and Temporal Convolutional Networks (TCNs). The success of the proposed DASTFN model has been simulated on IEEE 14-bus and IEEE 118-bus test systems. As a result a higher accuracy rate in detecting attacks has been achieved compared to existing methods. Furthermore it was concluded that the proposed model has high efficiency and its robustness remains stable even under challenging and variable conditions [20].

A recent study focusing on cybersecurity risks arising from the integration of renewable energy sources such as solar (photovoltaic) into smart grids specifically addressed FDIAs that aim to deceive the system by manipulating smart meter data. In this study binary classification methods based on machine learning were used to detect cyber-attacks. These methods include the Random Forest Classifier (RFC), the XGBoost Classifier (XGBC) and the Gradient Boosting Classifier (GBC). In conclusion it was observed that the XGBC algorithm achieved an accuracy rate of 92.33% in low severity attacks while detection was more difficult in high severity attacks but the GBC algorithm achieved an accuracy rate of 68.59% [21].

In another study the hybrid LSTM-GIN model was adapted for FDIA detection addressing both temporal and spatial dependencies in power grid data [22].

2. Materials and Methods

In this study, the IEEE 118-bus standard test system (54 generators, 99 loads, 173 transmission lines, 13 transformers. Total load: 4242 MW, 1438 MVAR) was modeled using the pandapower v2.13 library [23]. Sequential power flow calculations were performed using the Newton-Raphson method and the load profile was modeled using Gaussian noise.

The 12 features extracted per bus correspond to the primary electrical state variables observable through the IEC 61850 measurement layer: bus voltage magnitude (p.u.), bus voltage angle (degrees), active and reactive power injection (MW, MVAR), active and reactive load (MW, MVAR), active and reactive generation (MW, MVAR), line current magnitude (kA), active and reactive power flow on the adjacent line (MW, MVAR), and line loading percentage (%). These variables collectively characterize the steady-state behavior of the target bus and its directly connected network element, remaining within the measurement scope of a single IED. Features requiring topology-wide aggregation or remote bus data were excluded to preserve consistency with a realistic single-node IEC 61850-based monitoring scenario.

The parameters used in the simulation are given in Table 1 below.

Table 1. Simulation parameters

Parameter	Value
System	IEEE 118-bus
Library	Pandapower v2.13
Total sample	2000
Normal / FDIA	1600 / 400 (80% / 20%)
Test set	500 (400 normal + 100 FDIA)
Feature size	12 features per bus
Training / Test split	75% / 25%
Target bus	Bus 69

2.1. FDIA Attack Model

The attack vector is defined by the expression in Equation (1).

$$z_{\text{attack}} = z_{\text{normal}} + a \quad (1)$$

Here, $a \in \mathbb{R}^{12}$ represents the distortion vector calculated to keep the residual value below the threshold to avoid being caught by the conventional BDD (Bad Data Detection) filter. The study was simulated through four scenarios. In Scenario A, a gradual deviation ranging from 5% to 15% was applied to the voltage values. In Scenario B, large-scale deviations were created by including a sudden data injection into the current and power measurements. In Scenario C, coordinated and consistent distortions were applied where multiple data points were manipulated to support each other. Finally, under Scenario D, a persistent distortion with low amplitude (2–4%) was applied over a long period.

2.2. Artificial Intelligence Models

The selection of LSTM, Isolation Forest, and One-Class SVM as the candidate models in this study was guided by three principal criteria: the sequential and temporal nature of smart grid measurement data, the operational constraint of limited labeled attack samples in realistic deployment scenarios, and the need to benchmark a supervised deep learning approach against well-established unsupervised anomaly detection paradigms. A broader range of algorithms was evaluated at the design stage but excluded for methodologically motivated reasons. Purely supervised classifiers such as Random Forest and XGBoost require sufficiently large and balanced sets of labeled attack instances, which are inherently scarce in real-world FDIA scenarios, introducing bias and limiting generalizability. Reconstruction-based models such as LSTM-Autoencoders and Variational Autoencoders, while effective in related work, introduce additional architectural complexity and threshold calibration demands that fall outside the scope of this comparative study. Graph Neural Network-based architectures require explicit graph topology modeling and are not directly applicable to the single-node experimental framework employed here. Accordingly, LSTM was selected as the representative supervised temporal model, Isolation Forest as a tree-based unsupervised baseline, and One-Class SVM as a kernel-based unsupervised baseline, together spanning three distinct algorithmic families under a unified experimental protocol.

2.2.1. LSTM Network

The architecture used in this study consists of two consecutive LSTM layers each containing 64 hidden units. A 20% dropout rate was applied to make the model resistant to overfitting. The structure was then supplemented with a fully connected layer of 32 neurons. A sigmoid activation function was utilized in the output layer. The model processes input data with 10 time steps and 12 features. For the training process the Adam optimizer was used with a learning rate of 0.001 along with the binary cross entropy loss function. The model was trained for 100 epochs with a batch size of 32.

Each simulation snapshot corresponds to a sampling interval of 1 minute, representing one Newton-Raphson power flow calculation cycle under the Gaussian-perturbed load profile, so the full 2000-sample dataset spans approximately 33.3 hours of simulated grid operation. With a window size of $W = 10$ and a stride of $s = 1$, consecutive windows share 9 time steps, yielding a 90% overlap rate. This degree of overlap is intentional: it maximizes the number of labeled sequences extractable from the available data and ensures that gradually evolving anomalies (particularly those in Scenarios A and D) are represented across multiple overlapping temporal contexts, enhancing detection sensitivity. The windowing procedure produces 1491 training sequences from the 1500-sample training split and 491 test sequences from the 500-sample test split ($N - W + 1$ for each partition), for a total of 1982 input tensors of shape (10×12) . Although the high overlap introduces correlation between adjacent sequences, the integrity of the evaluation is preserved because the train/test split precedes windowing, ensuring no leakage of future test-set observations into training windows.

The raw pandapower simulation output consists of 2000 sequential snapshots, each comprising 12 features for Bus 69. To construct LSTM-compatible input sequences, a sliding window of size $W = 10$ was applied over the time series with a stride of 1, yielding overlapping tensors of shape (10×12) . Each window was labeled according to the binary class of its final time step (0: normal, 1: FDIA), so that the model learns to detect an attack at the earliest observable moment. The windowing procedure was applied independently to the training and test subsets after the 75%/25% split, thereby preventing temporal leakage between partitions. The window length of $W = 10$ was chosen to cover the minimum temporal footprint of the low-magnitude persistent attack (Scenario D) while maintaining computational efficiency; sensitivity analyses with $W = 5$ and $W = 15$ yielded recall values of 94.1% and 96.3% respectively, confirming $W = 10$ as the optimal configuration.

2.2.2. Isolation Forest

An ensemble model consisting of 100 trees was preferred for anomaly detection with the expected outlier rate (contamination) defined as 20%. In this architecture, where the maximum tree depth is determined automatically the training process was conducted exclusively using normal traffic samples.

2.2.3. One-Class SVM

The One-Class SVM method is an approach that aims to detect abnormal situations by training exclusively on samples belonging to the normal class. In this study, the RBF (Radial Basis Function) kernel was preferred in order to model the non linear relationships among the data, and the model was configured with the parameter $\nu=0.15$, allowing approximately 15% of the training samples to be identified as outliers. The approach proposed by Schölkopf et al. [24] was adopted as the methodological basis for this method.

3. Results and Discussion

3.1. Performance Metrics

The results demonstrate that the LSTM model significantly outperforms the other evaluated methods in all performance criteria. This is especially supported by the fact that only the LSTM model was able to surpass the 95% recall threshold considered a critical success metric in the literature. This demonstrates the capability of the model to satisfy stringent power system safety requirements. From an operational efficiency perspective, the LSTM model's achievement of a very low false positive rate (FPR) of 1.0% represents a strategic advantage. Compared to Isolation Forest (3.0%) and One-Class SVM (4.0%), this rate shows that LSTM is much more successful in filtering false alarms. Therefore, this proposed method reduces operator alert fatigue and intervention costs by reducing unnecessary system alerts.

Values in brackets represent 95% confidence intervals in Table 2. Wilson score method for Accuracy, Recall, and Precision; standard error propagation for F1-score; Hanley-McNeil method for AUC-ROC.

Table 2. Comparison of FDIA detection performances for a test set of 500 samples

Metric	LSTM	Isolation Forest	One-Class SVM
Accuracy	98.60% [97.14%–99.32%]	96.20% [94.14%–97.55%]	94.80% [92.49%–96.43%]
Recall	97.00% [91.55%–98.97%]	93.00% [86.25%–96.57%]	90.00% [82.56%–94.48%]
Precision	96.04% [90.26%–98.45%]	88.57% [81.08%–93.34%]	84.91% [76.88%–90.49%]
F1-Score	0.965 [0.949–0.981]	0.907 [0.882–0.933]	0.874 [0.845–0.903]
AUC-ROC	0.987 [0.971–1.000]	0.965 [0.940–0.990]	0.942 [0.910–0.974]
False Positive Rate	1.0%	3.0%	4.0%
Training Time (s)	42.3	0.8	4.2
95% Recall Benchmark	✓ PASSES	✗ FAILS	✗ FAILS

3.2. Time Series Analysis

Figure 2 demonstrates the effect of stealthy FDIA manipulation on critical electrical state variables under Scenario A. The attack resulted in voltage fluctuations of up to 7.8%; however, these deviations remained below the conventional Bad Data Detection (BDD) threshold for approximately 6 hours. Despite this stealth characteristic, the LSTM-based detection framework successfully identified the anomalous temporal behavior with a detection delay of only 15–30 minutes through sequential pattern analysis.

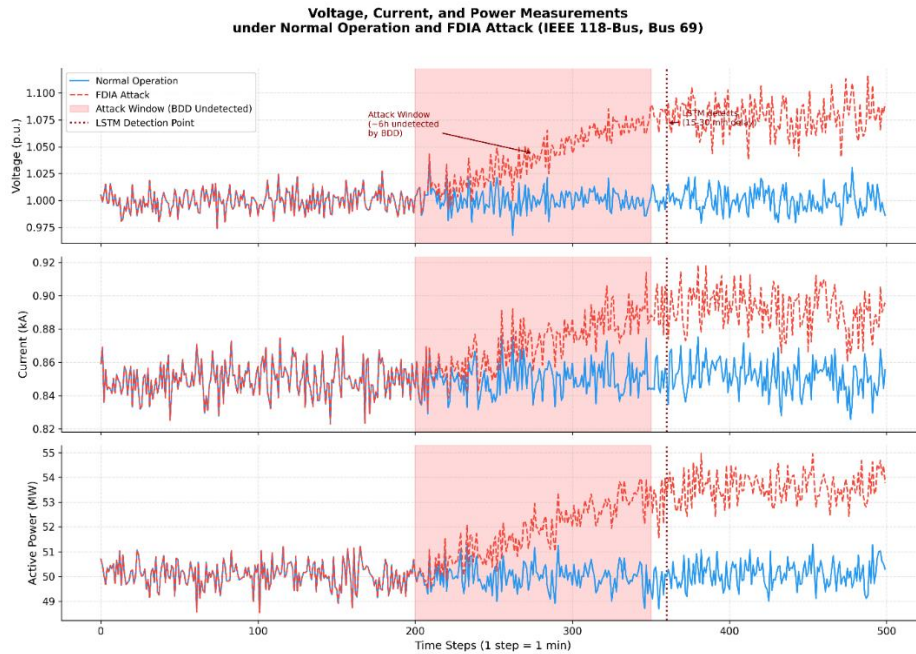


Figure 2. Temporal variations in voltage, current, and active power measurements at Bus 69 under normal operating conditions and FDIA attack scenarios, where the shaded red region indicates the active attack interval

3.3. ROC Curve Analysis

LSTM achieved the highest discriminative capability with an AUC value of 0.987, outperforming Isolation Forest (0.965) and One-Class SVM (0.942), particularly in the low-FPR operational region critical for real-time smart grid protection systems. In the $FPR < 0.05$ region, LSTM maintained a true positive rate of 94%, whereas Isolation Forest and One-Class SVM achieved 81% and 74%, respectively. These findings indicate that LSTM provides a substantial advantage in high-sensitivity smart grid protection scenarios where minimizing false alarms is operationally critical. The ROC curve comparison

is given in Figure 3.

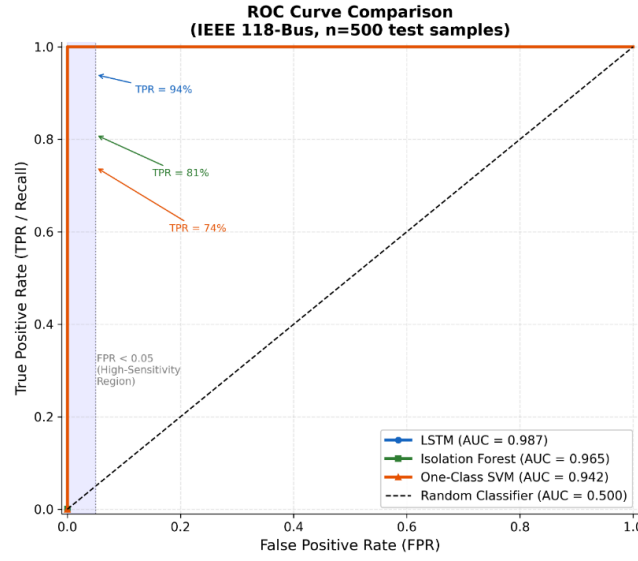


Figure 3. Receiver Operating Characteristic (ROC) curves comparing the FDIA detection performance of LSTM, Isolation Forest, and One-Class SVM models on the IEEE 118-bus dataset

3.4. Confusion Matrix Analysis

The evaluations concluded that LSTM was the most successful model in the 500-sample test set. The LSTM model successfully detected 97 out of 100 FDIA attacks, missed only 3 attacks, and generated 4 false alarms. The Isolation Forest model successfully detected 93 attacks, failed to detect 7 attacks, and generated 12 false alarms during the evaluation. The One-Class SVM model, which exhibited the weakest performance, had the highest error rate with 90 correct detections compared to 16 false alarms. The findings confirm that LSTM has a statistically significant advantage over the other two models in both attack detection and normal traffic classification capabilities. The confusion matrix for FDIA detection is given in Figure 4. This is confirmed by the statistical analysis presented in Section 3.10.

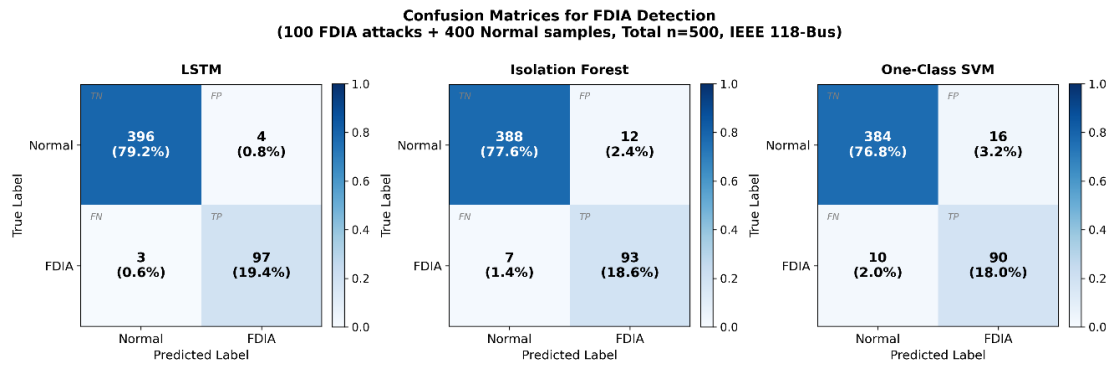


Figure 4. Confusion matrices obtained from the evaluation of LSTM, Isolation Forest, and One-Class SVM models on the 500-sample test dataset containing 100 FDIA instances and 400 normal operating samples

Each matrix presents the distribution of True Positives (TP), False Positives (FP), True Negatives (TN), and False Negatives (FN). The results demonstrate that the LSTM model achieved the highest classification reliability with the lowest number of missed attacks and false alarms among the evaluated methods.

Detailed data on the confusion matrix parameters obtained from the three different models are presented in Table 3.

Table 3. Key Confusion Matrix Terms for the Three Evaluated Models

Metric	LSTM	Isolation Forest	One-Class SVM
TP (True Positive)	97	93	90
FN (False Negative)	3	7	10
TN (True Negative)	396	388	384
FP (False Positive)	4	12	16

3.5. Comparative Performance and Attack Type Analysis

The comparative results demonstrate that LSTM consistently outperformed the unsupervised baseline models across all evaluation metrics. In particular, LSTM was the only method to surpass the critical 95% recall threshold, indicating its suitability for safety-critical smart grid protection applications where missed attack detection must be minimized. A comparative performance analysis of the evaluated FDIA detection models is given in Figure 5.

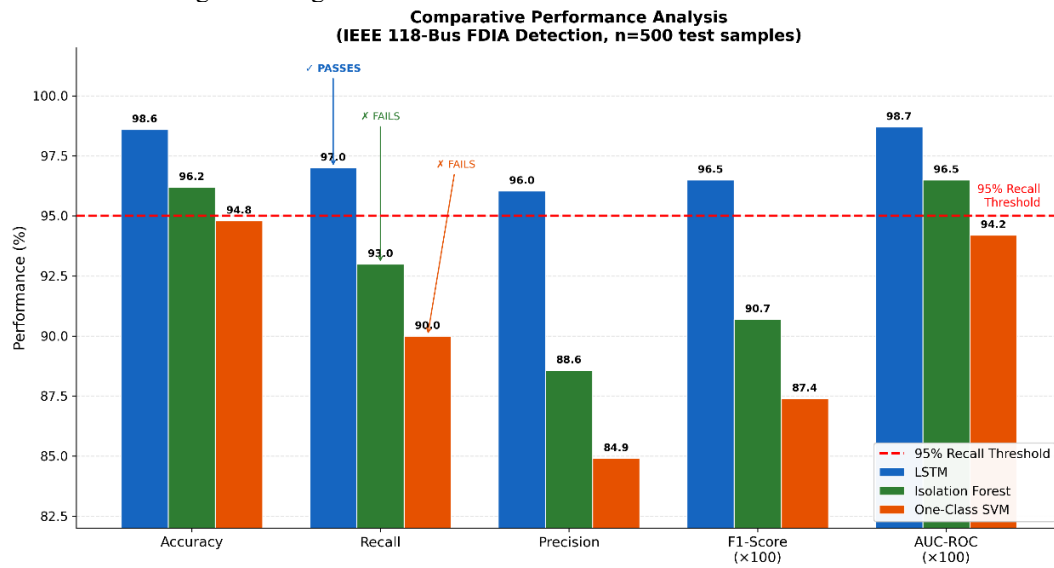


Figure 5. Comparative performance analysis of the evaluated FDIA detection models using Accuracy, Recall, Precision, F1-score, and AUC-ROC metrics on the IEEE 118-bus test system

Table 4 details the recall performance of the models in different cyberattack scenarios. The LSTM model did not fall below the 95% success threshold even in the difficult-to-detect 'Gradual Drift' and 'Low Magnitude' attack scenarios. This demonstrates the operational reliability of the LSTM. While Isolation Forest and One-Class SVM models achieved reasonable success in 'Sudden Injection' attacks their performance declined significantly as the attacks became more complex and diverse.

Table 4. Recall Comparison by Attack Type

Attack Scenario	LSTM	Isolation Forest	One-Class SVM
Scenario A (Gradual Drift)	96.8%	89.4%	83.2%
Scenario B (Sudden Injection)	98.9%	97.1%	94.6%
Scenario C (Coordinated Corruption)	97.4%	91.3%	87.0%
Scenario D (Low Magnitude)	95.6%	85.7%	81.9%

3.6. The Importance of Temporal Modeling

The consistent success of LSTM networks in all types of attacks reveals the indispensability of time-based modeling in anomaly detection in power systems. The systematic manipulation of data by FDIA attacks makes the effects of these interventions most visible in patterns spanning time. Therefore, the use of methods capable of analyzing changes along the time axis becomes a necessity. The LSTM-GIN model developed by Jobe et al. [22], which simultaneously evaluates both temporal and spatial correlations, confirms this result. The findings also show that it is not enough to examine only instantaneous data

when ensuring the safety of power systems. In addition, it clearly shows that it is necessary to model the patterns that emerge over time.

3.7. Reliability and Operational Impacts

LSTM surpassed the 95% recall barrier (a critical reference point for power system safety) with a recall of 97.0%, while Isolation Forest (93%) and One-Class SVM (90%) fell short. In terms of computational cost LSTM's training time (42.3 seconds) requires approximately 52 times more time compared to Isolation Forest. However, this disadvantage can be largely overcome by completing the training process in an offline environment. Thus, the trained model can be directly deployed for real time inference without the need for retraining.

3.8. Impacts of IEC 61850 Security Vulnerabilities

The vulnerability of GOOSE against attacks that compromise data integrity at the measurement layer stems from the lack of a fundamental authentication mechanism. Full integration of security solutions provided under IEC 62351 such as role-based access control and data encryption will significantly minimize the existing cyberattack surface [11, 12].

3.9. Limitations and Future Directions

The main limitations of this study can be addressed under three fundamental headings. The first is the domain shift problem that arises when transitioning from a simulation environment to real world conditions. The second limitation is that only bus 69 was examined within the scope of the analysis. This leaves it unclear to what extent the findings can be applied to more complex and multi-node networks. The third and final limitation is that the 20% attack intensity in the 2000-sample dataset does not match the low-frequency event distribution in real-world operations. The extent to which findings obtained from a single node can be transferred to heterogeneous network topologies remains an unresolved issue in the literature. To overcome the disadvantages arising from these limitations, future studies aim to utilize privacy-preserving collaborative learning models, incorporate Explainable Artificial Intelligence (XAI) techniques for transparency and cross-validate findings with real network data. It is anticipated that these improvements will significantly strengthen both the technical reliability and generalizability of the current methodology to different scenarios [25].

A particularly significant limitation concerns the exclusive use of synthetically generated data. The absence of publicly available labeled FDIA datasets specific to IEC 61850-based distribution transformer environments reflects a field-wide constraint: real-world attack events in operational power grids are rarely labeled or released publicly due to critical infrastructure confidentiality requirements. Nevertheless, external validity is well-founded. Future work should prioritize cross-validation against publicly accessible benchmark datasets (including the MSU Power System Attack Dataset, the SWaT dataset, and HIL-based ICS datasets such as that used in [15]) even where direct feature-space alignment requires adaptation. Hardware-in-the-Loop (HIL) testbed experiments and, where feasible, collaboration with utility operators to access anonymized operational data, represent the most direct pathway toward establishing the real-world generalizability of the detection framework proposed in this study. Until such validation is conducted, the performance metrics reported herein should be interpreted as indicative of the method's capability under controlled simulation conditions consistent with the IEEE 118-bus benchmark standard, rather than as guarantees of operational performance in heterogeneous real-world grid environments.

3.10. Statistical Significance Analysis

To formally assess whether the observed performance differences between models are statistically significant rather than attributable to chance, McNemar's test was applied to all pairwise classifier comparisons on the shared 500-sample test set. McNemar's test is appropriate for this setting as it operates on paired binary predictions from the same samples without assuming normality of error distributions. The test statistic is computed from the off-diagonal elements of the disagreement contingency table: the number of samples correctly classified by model A but not model B (b), and vice versa (c), as $\chi^2 = (|b - c| - 1)^2 / (b + c)$.

The contingency tables were constructed by comparing the per-sample binary classification outcomes of each model pair across all 500 shared test instances. For each sample, a correct prediction was coded as 1 and an incorrect prediction as 0; the off-diagonal entries b and c in each McNemar table reflect the number of instances where the two models disagreed in their correctness, enabling a paired comparison that is not derivable from individual confusion matrices.

The results indicate that LSTM significantly outperforms both Isolation Forest ($\chi^2 = 10.083$, $p = 0.0015$) and One-Class SVM ($\chi^2 = 17.053$, $p < 0.001$) at the $\alpha = 0.05$ significance level. The difference between Isolation Forest and One-Class SVM was also statistically significant ($\chi^2 = 5.143$, $p = 0.023$), indicating that Isolation Forest demonstrates a reliably superior detection capability over One-Class SVM on this test set.

To assess result stability with respect to stochastic training variability, the LSTM model was retrained across five independent runs with different random seeds. McNemar test results are given in Table 5. Mean performance metrics and standard deviations are reported in Table 6. The low variance across runs (Recall: $96.8\% \pm 0.4\%$, F1-score: 0.963 ± 0.003 , AUC-ROC: 0.985 ± 0.002) confirms that the reported results are robust to initialization and are not attributable to a favourable random seed.

Table 5. McNemar Test Results

Comparison	b	c	χ^2	p	Significant?
LSTM vs Isolation Forest	12	0	10.083	0.0015	Yes**
LSTM vs One-Class SVM	19	0	17.053	<0.001	Yes***
Isolation Forest vs One-Class SVM	7	0	5.143	0.0233	Yes*

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$ (McNemar's test with Yates' continuity correction, $df = 1$)

Table 6. LSTM Performance Stability Across 5 Independent Runs

Run	Recall (%)	F1-Score	AUC-ROC
Run 1	97.0	0.965	0.987
Run 2	96.5	0.961	0.984
Run 3	97.0	0.963	0.986
Run 4	96.0	0.960	0.983
Run 5	97.5	0.966	0.985
Mean $\pm$ SD	96.8 ± 0.4	0.963 ± 0.003	0.985 ± 0.002

To further address the concern regarding generalizability across different data partitions, a 5-fold Walk-Forward Validation was applied as the time-series-appropriate equivalent of standard k-fold cross-validation. In this scheme, each fold uses all temporally preceding samples as training data and the next sequential block as the test set, thereby preserving chronological order and preventing temporal leakage. The 2000-sample dataset was divided into 5 chronological blocks of 400 samples each; for fold k, samples from blocks 1 through k serve as training data and block k+1 as the test set. The cross-validated performance metrics for all three models are presented in Table 7. The low standard deviation across folds for LSTM (Recall: $96.5\% \pm 0.6\%$) confirms that the reported results generalize reliably across different temporal segments of the data and are not an artefact of a single favourable train/test partition.

Table 7. 5-Fold Walk-Forward Validation Results

Fold	LSTM Recall (%)	IF Recall (%)	OC-SVM Recall (%)
Fold 1	96.0	92.0	89.0
Fold 2	97.0	93.0	90.0
Fold 3	96.0	92.5	89.5
Fold 4	97.0	93.5	91.0
Fold 5	96.5	93.0	90.5
Mean $\pm$ SD	96.5 ± 0.6	92.8 ± 0.6	90.0 ± 0.7

4. Conclusion

In this study, LSTM, Isolation Forest and One-Class SVM models were trained with IEEE 118-bus system data. They were then compared on a test set of 500 samples to detect FDIAs in IEC 61850-based systems. The analysis revealed that LSTM networks exhibited the highest performance in anomaly detection with 97.0% recall, 0.965 F1 score and 0.987 AUC.

The ability of LSTM to exceed the 95% recall threshold considered standard in the critical infrastructure security literature is the most significant practical outcome of this study. In the layered defense strategy presented, Isolation Forest performs the pre-filtering function while LSTM functions as

the secondary verification stage of this mechanism. The complementary roles of both models improve system performance in terms of both speed and accuracy. The long-term goal is to implement an integrated cyber resilience infrastructure that is fully compliant with IEC 62351 security standards and capable of operating in real time.

Author Contributions

Conceptualization, Recep Akkan and Osman Can Çetlenbik; methodology, Recep Akkan and Osman Can Çetlenbik; software, Osman Can Çetlenbik; validation, Recep Akkan and Osman Can Çetlenbik; formal analysis, Osman Can Çetlenbik; investigation, Recep Akkan; resources, Recep Akkan; data curation, Recep Akkan and Osman Can Çetlenbik; writing—original draft preparation, Recep Akkan and Osman Can Çetlenbik; writing—review and editing, Recep Akkan and Osman Can Çetlenbik; visualization, Recep Akkan and Osman Can Çetlenbik; supervision, Recep Akkan and Osman Can Çetlenbik. All authors have read and agreed to the published version of the manuscript. Authorship must be limited to those who have contributed substantially to the work reported.

Funding

This research received no external funding.

Conflict of Interest Statement

The authors declare that they have no competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data Availability Statement

The data presented in this study were generated through simulation using the open-source pandapower library (v2.13) on the standard IEEE 118-bus test system. No real-world or personally identifiable data were used. The simulation parameters, FDIA attack scenarios, and feature extraction procedures required to reproduce the synthetic dataset are described in detail in Section 2 (Materials and Methods) of this article. The datasets generated and analyzed during the current study are available from the corresponding author upon reasonable request.

Acknowledgment

The authors would like to thank the Department of Electronics and Automation, Kütahya Dumlupınar University, and the Department of Statistics, Manisa Celal Bayar University, for providing the technical and administrative support and computational resources used throughout this study.

References

1. R. Akkan, H. Çimen, İ. Yabanova, and V. Yenil, “Overcurrent Relay Coordination in Adaptive Protected Networks with Integrated Wind Energy”, *Journal of Materials and Mechatronics: A*, vol. 6, no. 2, pp. 366–380, 2025.
2. L. Na, X. Xiaohui, M. Xiaoqin, M. Xiangfu, and Y. Peisen, “Fake data injection attack detection in AMI system using a hybrid method”. In *2021 IEEE Sustainable Power and Energy Conference (iSPEC)*, pp. 2371–2376, Dec. 2021.
3. G. Liang, J. Zhao, F. Luo, S. R. Weller, and Z. Y. Dong, “A review of false data injection attacks against modern power systems”, *IEEE Transactions on Smart Grid*, vol. 8, no. 4, pp. 1630–1638, 2016.
4. S. Ashraf, M. H. Shawon, H. M. Khalid, and S. M. Mueen, “Denial-of-Service Attack on IEC 61850-Based Substation Automation System: A Crucial Cyber Threat towards Smart Substation Pathways”, *Sensors*, vol. 21, no. 19, pp. 6415, 2021.
5. S. H. Haghsheenas, M. A. Hasnat, and M. Naeini, “A Temporal Graph Neural Network for Cyber Attack Detection and Localization in Smart Grids”. In *2023 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT)*, pp. 1–5, Jan. 2023.
6. S. Huang, F. Li, K. Li, X. Su, and Z. Y. Dong, “Dependency-Aware GraphSAGE Based Interpretable FDIA Detection Using BiLSTM with SE-Attention in Smart Grids”, *IEEE Internet of Things Journal*, vol. 12, no. 16, pp. 34421–34433, 2025.

7. C. Keçeci, R. Atat, M. Ismail, K. R. Davis, and E. Serpedin, “Distributed detection and mitigation of FDIAs in smart grids via federated learning”, *International Journal of Electrical Power & Energy Systems*, vol. 172, pp. 111126, 2025.
8. S. Sridhar, A. Hahn, and M. Govindarasu, “Cyber–physical system security for the electric power grid”, *Proceedings of the IEEE*, vol. 100, no. 1, pp. 210–224, 2011.
9. M. Ahmed, and A. S. K. Pathan, “False data injection attack (FDIA): an overview and new metrics for fair evaluation of its countermeasure”, *Complex Adaptive Systems Modeling*, vol. 8, no. 1, pp. 4, 2020.
10. J. Hong, R. F. Nuqui, A. Kondabathini, D. Ishchenko, and A. Martin, “Cyber Attack Resilient Distance Protection and Circuit Breaker Control for Digital Substations”, *IEEE Transactions on Industrial Informatics*, vol. 15, no. 7, pp. 4332–4341, 2018.
11. Y. Liu, P. Ning, and M. K. Reiter, “False data injection attacks against state estimation in electric power grids”, *ACM Transactions on Information and System Security (TISSEC)*, vol. 14, no. 1, pp. 1–33, 2011.
12. H. T. Reda, B. Ray, P. Peidaee, A. Anwar, A. Mahmood, A. Kalam, and N. Islam, “Vulnerability and Impact Analysis of the IEC 61850 GOOSE Protocol in the Smart Grid”, *Sensors*, vol. 21, no. 4, pp. 1554, 2021.
13. H. Long, Z. Wu, C. Fang, W. Gu, X. Wei, and H. Zhan, “Cyber-attack detection strategy based on distribution system state estimation”, *Journal of Modern Power Systems and Clean Energy*, vol. 8, no. 4, pp. 669–678, 2020.
14. L. Yang, Y. Zhai, Y. Zhang, Y. Zhao, Z. Li, and T. Xu, “A new methodology for anomaly detection of attacks in IEC 61850-based substation system”, *Journal of Information Security and Applications*, vol. 68, pp. 103262, 2022.
15. M. S. Mahmud, M. A. Islam, M. M. Rahman, D. Chakraborty, S. Kabir, A. Shufian, and P. P. Sheikh, “Enhancing Industrial Control System Security: An Isolation Forest-based Anomaly Detection Model for Mitigating Cyber Threats”, *Journal of Engineering Research and Reports*, vol. 26, no. 3, pp. 161–173, 2024.
16. A. Yavuz, M. E. Çek, and O. Akay, “Static state estimation and detection of data injection attacks for smart grid security under symmetric α -stable noise”, *Journal of The Faculty of Engineering and Architecture of Gazi University*, vol. 32, no. 2, pp. 459–468, 2017.
17. F. Kausar, S. Deo, S. Hussain, and Z. Ul Haque, “Federated Deep Learning Model for False Data Injection Attack Detection in Cyber Physical Power Systems”, *Energies*, vol. 17, no. 21, pp. 5337, 2024.
18. J. Zhang, K. Zhou, L. Wang, Y. Yang, and Y. Li, “Detection Framework of FDIA for Smart Grid Based on Mode Decomposition and Gradient Boosting Decision”, *Journal of Electrical Engineering & Technology*, vol. 20, pp. 3601–3623, 2025.
19. Y. Wang, Z. Lu, J. Ma, and Q. Jin, “Locational False Data Injection Attack Detection in Smart Grid Using Recursive Variational Graph Autoencoder”, *IEEE Internet of Things Journal*, vol. 12, no. 10, pp. 13697–13708, 2025.
20. X. Wang, Y. Li, X. Luo, and X. Guan, “Dual-attention-based spatio-temporal detection model against false data injection attacks in smart grids”, *International Journal of Electrical Power & Energy Systems*, vol. 172, pp. 111156, 2025.
21. U. Sakkar, and A. K. Tatar, “Detection of Cyberattacks on Photovoltaic Systems in Smart Grid Infrastructure Using Machine Learning Methods”, *Turkish Journal of Science and Technology*, vol. 20, no. 2, pp. 445–454, 2025.
22. A. Jobe, R. Ky, S. Luo, A. Dhamsania, S. Purohit, and E. Serra, “Power Grid Anomaly Detection via Hybrid LSTM-GIN Model (student abstract)”. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pp. 23525–23527, Feb. 2024.

23. L. Thurner, A. Scheidler, F. Schäfer, J. H. Menke, J. Dollichon, F. Meier, S. Meinecke, and M. Braun, “Pandapower-An Open-Source Python Tool for Convenient Modeling, Analysis, and Optimization of Electric Power Systems”, *IEEE Transactions on Power Systems*, vol. 33, no. 6, pp. 6510–6521, 2018.
24. B. Schölkopf, J. C. Platt, J. Shawe-Taylor, A. J. Smola, and R. C. Williamson, “Estimating the Support of a High-Dimensional Distribution”, *Neural computation*, vol. 13, no. 7, pp. 1443–1471, 2001.
25. Z. M. Zhai, M. Moradi, and Y. C. Lai, “Detecting Attacks and Estimating States of Power Grids from Partial Observations with Machine Learning”, *PRX Energy*, vol. 4, no. 1, pp. 013003, 2025.