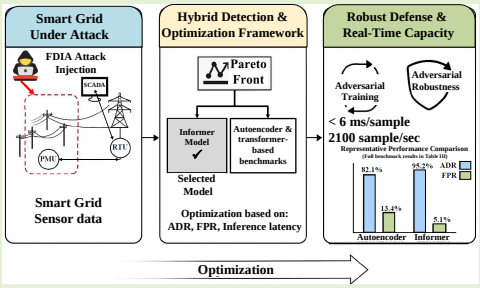
Enhancing Smart Grid Cybersecurity: Multi-Objective Optimization of Informer-Based FDIA Detection with Adversarial Defense Mechanisms

Lilia Tightiz, L. Minh Dang, Sanjeevikumar Padmanaban *Senior Member, IEEE*, and Hyosik Yang *Member, IEEE*

***Abstract*— In the era of smart grids, the use of digital technologies has introduced unprecedented cybersecurity challenges, particularly in safeguarding against false data injection attacks (FDIAs) that may potentially manipulate power system measurements and disrupt grid operation. This paper proposes an Informer-based FDIA detection framework with multi-objective optimization, a non-dominated sorting genetic algorithm II (NSGA-II), and adversarial defense techniques to maximize the attack detection rate (ADR), minimize the false positive rate (FPR), and reduce computational overhead. Leveraging ProbSparse self-attention, the Informer efficiently captures long-range temporal dependencies with low complexity, making it well-suited for real-time anomaly detection in power-system time series. Experimental evaluation on IEEE 5-bus, 14-bus, and 118-bus systems demonstrates the robustness and scalability of the proposed approach. Feature selection optimization, evaluated relative to the same Informer detector trained without feature selection, increased the attack detection rate (ADR) by 12.5 percentage points and reduced the FPR by 9 percentage points, confirming that redundant features should be omitted. Specifically, Bayesian hyperparameter optimization reduces Informer training time from 188.7 s to 145.2 s (22.0%), while subsequent NSGA-II-based multi-objective optimization further improves ADR from 87.3% to 95.1% under identical training conditions. The proposed defense mechanism is robust against fast gradient sign method (FGSM), basic iterative method (BIM), and momentum iterative method (MIM) attacks, and it outperforms a conventional Autoencoder-based detector. Moreover, the proposed detector consistently outperforms state-of-the-art Transformer time-series baselines (TimesNet, PatchTST, and FEDformer) under the same experimental protocol. Real-time feasibility experiments further show that the Informer achieves an inference latency of under 6 ms per sample and a maximum throughput of 2100 samples per second, supporting deployment in large-scale smart grid monitoring.**



***Index Terms*— False Data Injection Attacks, Smart Grid Cybersecurity, Informer Model, Multi-Objective Optimization, Adversarial Machine Learning, Bayesian Hyperparameter Tuning, Real-Time Anomaly Detection.**

This research was supported by the Technology Development Program to Solve Climate Changes through the National Research Foundation of Korea (NRF), funded by the Ministry of Science, ICT (NRF-2021M1A2A2065447). This work was also supported by the Institute of Information and Communications Technology Planning and Evaluation (IITP) grant funded by the Korean government (MSIT) (No. 2022-0-00106, development of explainable AI-based diagnosis and analysis frame work using energy demand big data in multiple domains).

L. Tightiz is with Department of Computer Science and Engineering, Sejong University, Seoul 05006, Republic of Korea, e-mail: liliatightiz@sejong.ac.kr

L. M. Dang is with Institute of Research and Development, Duy Tan University, Da Nang, 550000, Viet Nam, and Faculty of Information Technology, Duy Tan University, Da Nang, 550000, Viet Nam, danglienminh@duytan.edu.vn

S. Padmanaban is with Department of Electrical Engineering, IT and Cybernetic, University of South-Eastern, Norway–Campus Porsgrunn, 7430, Norway, email:sanjeev.padma@usn.no

Corresponding author: HS. Yang is with Department of Computer Engineering, Sejong University, 209, Neungdong-ro, Gwangjin-gu, Seoul 05006, Republic of Korea, email:hsyang@sejong.ac.kr

I. Introduction

A. Background on Smart Grid Cybersecurity

The smart grid is an updated power system incorporating state-of-the-art digital communication and information technology (IT) for efficiency, reliability, and sustainability in the energy distribution system [1]. Unlike in traditional power grids, where energy distribution occurs in one direction, smart grids have both energy and data flowing in both directions: from consumers to utilities and vice versa [4]. Energy sources integrated into the system may be distributed over different resources like solar panels, wind turbines, and electric vehicles [5]. The smart grid utilities have applied the Internet of things (IoT)-based devices, advanced metering infrastructure (AMI), and supervisory control and data acquisition (SCADA) systems for real-time monitoring and automated responses against system fluctuations [6]. Because the deployment of

these technologies creates additional network entry points that hackers can use to launch various cyberattacks against smart grids, such as FDIAs, denial of service attacks (DoS), distributed denial of service attacks (DDoS), and ransomware, the improvements in operational efficiency and resilience generated by these technologies are at risk. In order to trick state estimation algorithms into making poor operational decisions, a class of cyberattacks known as FDIAs manipulates sensor readings in the power grid. Attackers conduct unauthorized changes to PMU data and remote terminal unit (RTU) data at strategic sites, which avoids detection by standard anomaly detection systems [7].

Because FDIAs use sensor data to distribute energy in error, causing system overloads and shortages, they pose a risk to the stability and economic performance of the smart grid. In order to obtain unfair advantages, an attacker may also artificially raise or deflate electricity prices [2], [8]. Moreover, FDIAs promote instability in the grid and may result in blackouts by disturbing state estimation, hence triggering cascading failures; besides, they interact with demand-response mechanisms by falsely signaling adjustment in energy loads [9]. Due to their undetectable nature, FDIAs are generally hard to detect, as attackers can design these attacks to avoid detection using conventional state estimation methodologies and dynamically adjust strategies. On the other hand, anomaly detection suffers from high false positives due to the fact that threshold-based methods usually misclassify attacks with normal grid fluctuations, while machine learning models face class imbalance issues [10]. The problem is further exacerbated by computational complexity for real-time detection since deep learning models demand high processing power, and robust principal component analysis (RPCA)-based effective techniques insert latency into communication [1]. Another major challenge involves the lack of generalization among grid architectures, as

detection models are often trained and validated on small-scale benchmark systems, such as IEEE 5-bus networks, and may fail to generalize to larger and more complex grid topologies, including IEEE 14-bus and IEEE 118-bus systems [9]. The development of universal and reliable defense systems faces challenges because different regions and grid setups lead to different attack methods and operational patterns. Detection systems need to meet three requirements which include handling large scale operations, maintaining system integrity and following real-time performance standards.

The development of time-series learning has reached a new milestone through recent research, which shows that transformer-based systems can effectively model extended time periods while requiring less processing power. Models such as TimesNet and its variants have been successfully applied in real-time anomaly and spoofing detection scenarios, highlighting their effectiveness in capturing periodic and multi-scale temporal patterns under adversarial conditions [24]–[26]. However, most existing transformer-based approaches focus on forecasting or generic anomaly detection and do not explicitly address stealthy FDIA detection under strict real-time constraints or across heterogeneous grid architectures.

The challenges being addressed include developing AI-driven detection models and multi-objective optimization frameworks that help enhance FDIA detection and mitigation strategies [11].

B. Related Work

Several techniques have, therefore, been proposed to detect and mitigate cyber threats in smart grids; most of their focus has now shifted toward FDIAs. These can be broadly classified into traditional statistical detection methods, machine learning-based approaches, and optimization-driven frameworks. More recently, some adversarial defense mechanisms have worked to enhance the robustness of detection models.

Traditional statistical detection approaches have long been adopted as the basis for detecting anomalies within smart grids. Musleh et al. [12] investigated some residual-based state estimation approaches, namely weighted least squares (WLS) and Kalman filters, for identifying inconsistencies in the measurements of a power system. Though these methods are generally computationally efficient and interpretable, they cannot cope with sophisticated FDIAs bypassing residual thresholds. Whereas attackers develop even more sophisticated techniques, such as strategically injecting malicious data while keeping statistical consistency intact, traditional approaches, in isolation, are bound to be inadequate for the transition to more adaptive approaches using machine learning [13].

Machine learning-based techniques have gained momentum in identifying complex patterns associated with cyber attacks. Tightiz et al. [14] presented that both support vector machines (SVM) and decision trees are capable of identifying manipulated data from normal fluctuations within a grid. Mukherjee [15] further investigated a more powerful deep learning architecture, giving higher detection of the spatial and temporal dependencies present within power system data through improved convolutional neural networks (CNNs) and long-short-term memory (LSTM). On the other side, Zhang et al. [16] have shown the capabilities of Autoencoders and generative adversarial networks (GANs)-semi-supervised learning models that find new patterns of an attack with no large amount of labeled dataset required. But even these machine learning models are again too often haunted by issues with data scarcity, generalization of models, and potential vulnerability to adversarial attacks [17]. These challenges inspired further research into adversarial training techniques. Using this approach, for instance, in a very recent study, Ness [18] investigated artificial intelligence (AI)-driven adversarial defense mechanisms that improve robustness against perturbation-based attacks in smart grid systems. The approach entailed a structured adversarial retraining framework that significantly reduced attack success rates, especially against adaptive threat models. On the other side, Sampedro et al. [19] proposed a deep learning-based detection mechanism that utilized adversarially trained architectures for enhancing model resilience against evolving threats. However, such developments introduce computational overhead in most adversarial defenses and require very frequent retraining, hence becoming impractical in real-time grid environments.

TABLE I: Comparison of FDIA Detection Approaches and Closely Related Transformer-Based Time-Series Methods

Ref	Application Domain	Core Model	Learning Paradigm	Adversarial Consideration	Optimization Strategy	Scope / Limitation
[12]	Smart grid FDIA	WLS, Kalman filter	Model-based	No	No	Vulnerable to stealthy FDIAs
[14]	Smart grid FDIA	Decision Tree, SVM	Supervised ML	No	No	Limited generalization
[15]	Smart grid FDIA	CNN, LSTM	Supervised DL	No	No	High computational cost
[16]	Smart grid FDIA	Autoencoder, GAN	Semi-supervised DL	No	No	Sensitive to adversarial perturbations
[18]	Smart grid cybersecurity	DL with adversarial retraining	Supervised DL	Yes	No	Frequent retraining required
[19]	Cyber-physical systems	Adversarial DL	Supervised DL	Yes	No	Real-time constraints not addressed
[20]	Smart grid FDIA	Transformer-based FL	Federated DL	Partial	No	Communication overhead
[22]	Integrated energy systems	Informer	Supervised DL	No	No	Lacks adaptive defense
[27]	Cyber–physical power systems (FDIA attacker model)	Multi-objective stealthy FDIA construction (AC model)	Evolutionary optimization	Yes (attacker-side)	NSGA-II	Attack modeling only (no detection)
[21]	Smart grid FDIA	LSTM + GCN	Federated DL	No	No	No adversarial robustness; no optimization trade-offs
[28]	Smart grid (worst-case FDIA analysis)	Multi-objective strictly stealthy FDIA formulation	Evolutionary optimization	Yes (attacker-side)	SPEA2	Worst-case attack analysis only
[29]	Smart grid FDIA defense	DRL-based FDIA detection (MDP formulation)	DRL	No	Multi-objective RL	Detection parameter optimization
[30]	Power IoT security	AHP-based security risk assessment model	Expert-driven MCDM	No	AHP + PSO	Risk assessment only (no detection)
[23]	IEC 61850 substations	Informer + PPO	RL	Yes	RL-based	Not FDIA-specific
[24]	Cyber-physical spoofing	TimesNet	Supervised DL	Yes (spoofing)	No	Non-power-grid domain
[25]	Power systems	TimesNet + Crossformer	Supervised DL	No	No	Forecasting-only
[26]	Power systems	TimeXer	Supervised DL	No	No	Forecasting-only
This work	Smart grid FDIA	Informer + NSGA-II	Supervised DL + adversarial training	Yes	Multi-objective (NSGA-II)	Joint detection, adversarial robustness, scalability validation, and real-time feasibility analysis

To solve smart grid security's scalability and privacy issues, Li et al. [20] propose a federated deep learning-based approach where transformer-based architectures are used for secure, distributed FDIA detection. Unlike traditional centralized learning models, in federated learning, multiple grid nodes can be engaged in collaborative model training without sacrificing data privacy; thus, it is much closer to a practical implementation for a real-world smart grid. However, federated learning brings additional challenges in terms of convergence stability and communication overhead, and further optimization techniques must be developed in order to balance the trade-off between detection accuracy and computational feasibility.

Extending this line of research, Keçeci et al. [21] proposed a federated learning-based distributed framework for FDIA detection and localization in smart grids, combining graph convolutional networks (GCNs) and long short-term memory (LSTM) models to capture spatial–temporal dependencies across geographically distributed grid regions. Testing on IEEE benchmark systems, which comprise three distinct bus networks, 57-bus, 118-bus, and 300-bus, validated their methodology. The approach outperformed centralized systems in terms of scalability and privacy protection. The framework supports distributed representation learning and attack localization, but it needs to develop its adversarial defense system against adaptive and stealthy FDIAs and requires multi-objective optimization to achieve better attack detection accuracy and operational efficiency, and real-time constraint management.

Meanwhile, Sun et al. [22] introduced an Informer-based intrusion detection system to leverage long-sequence modeling in identifying cyber threats within integrated energy systems. The sparse self-attention mechanism of the Informer efficiently captures long-range time-series dependencies, making it well-suited for smart grid anomaly detection; however, existing Informer-based approaches generally lack adaptive mechanisms to counter dynamically evolving adversarial threats.

Closely related to this line of research, recent transformer-based time-series models have been actively explored in cyber-physical and power-system domains beyond FDIA detection. For example, an Informer-powered reinforcement learning framework for cybersecurity monitoring in IEC 61850-based substations was proposed in [23], demonstrating how long-sequence attention mechanisms can be combined with adaptive decision-making strategies to support real-time threat response in critical infrastructure environments.

In parallel, Wang *et al.* [24] employed a TimesNet-based architecture for real-time GPS spoofing detection in cyber-physical systems, showing that transformer-style temporal modeling can effectively capture periodic and multi-scale patterns under adversarial signal manipulation. Although this work is not focused on power grids, it provides evidence of the suitability of transformer-based time-series models for

detecting stealthy and structured attacks in real-time settings.

Within the power-system domain, transformer-based architectures such as TimesNet, Crossformer, and TimeXer have been investigated for short-term load forecasting and ramping demand prediction. Studies by He et al. [25] and Lu et al. [26] confirm that these models are effective in learning complex temporal dependencies from power-system measurements, further validating the applicability of transformer-based approaches to large-scale energy data. However, despite their demonstrated modeling capabilities, these works primarily target forecasting tasks or non-FDIA cyber threats and therefore do not explicitly address stealthy FDIA detection, adversarial robustness, or scalability across heterogeneous smart grid topologies.

A structured comparison of representative FDIA detection methods and closely related transformer-based time-series approaches is summarized in Table I. The table highlights that, although recent transformer-based models have significantly advanced time-series modeling capabilities in power and cyber-physical systems, existing works typically address only isolated aspects such as detection accuracy, privacy preservation, or optimization, without jointly considering adversarial robustness, multi-objective trade-offs, and scalability across different grid architectures.

In this respect, researchers have investigated multi-objective optimization frameworks to strike a balance between security and efficiency. A method has been designed to couple optimization with the main focus on precision in detection but also considering computational efficiency. Lu et al. [27] resorted to utilizing an NSGAII in balancing the trade-offs between detection accuracy and its efficiency. Also, evolutionary optimization techniques, like the strength Pareto evolutionary algorithm (SPEA2), have been used by Rahman et al. [28] to develop adaptive attack response strategies. In addition, Lin et al. [29] have presented multi-objective reinforcement learning-based optimization for dynamically adjusting the detection parameters, thereby reducing false positives significantly. The work of Sun et al. [30] has also demonstrated how the analytic hierarchy process with improved particle swarm optimization-based hierarchical optimization frameworks can be used to enhance the cybersecurity risk assessment of a power IoT environment through the optimal improvement of the attack detection strategy without resulting in an infeasible computation process, thereby enhancing resilience in smart grid applications.

Due to the increasing sophistication of FDIAs and their evasion abilities, advanced AI-driven solutions are needed for smart grid cybersecurity. The classical machine learning and deep learning approaches have been extensively examined; as mentioned, many of them suffer from high false alarm rates, computational inefficiency, or poor generalization across different grid architectures. Moreover, since new adversarial attacks are continuously being developed for active adaptation against detection mechanisms, it is essential to consider robust defense strategies [31].

Our proposed AI-based framework for detecting FDIA incorporates an Informer model for time series anomaly detection with multi-objective optimization, i.e., NSGA-II, increasing the detection capability in accuracy, efficiency, and scalability. Unlike traditional deep learning approaches that may be computationally expensive and suffer from misclassification, the sparse self-attention mechanism of the Informer model efficiently extracts long-range dependencies in smart grid data, making it suitable for real-time anomaly detection. Besides this, to defend against adaptive and stealthy cyber threats, adversarial defense mechanisms are integrated to strengthen the model against evolving attack strategies.

To validate the scalability and generalization of our proposed methodology, it has been applied in different grid topologies, such as the IEEE 5-bus, 14-bus, and 118-bus test networks. In general, by integrating an Informer model with multi-objective optimization and adversarial defenses, this work proposes an overall, efficient, and adaptive cybersecurity framework that remarkably enhances smart grid resilience against sophisticated FDIAs.

C. Contributions

In the proposed research, the Informer-based detection framework of FDIA for strengthening cybersecurity in smart grids will be advanced by including multi-objective optimization and adversarial defense mechanisms. The main contributions of the present work can be outlined below:

- Informer-Based Time-Series FDIA Detection: Design a new Informer model that captures the long-range dependency of smart grid data, enhancing real-time anomaly detection while reducing false alarms.
- Multi-Objective Optimization for Cybersecurity Trade-offs: Performing NSGA-II optimization to balance attack detection accuracy, FPR, and computational efficiency for practical real deployment in smart grids.
- Adversarial Defense Mechanism for Enhanced Robustness: Incorporates adversarial machine learning-based defenses to counter adaptive FDIAs, which will strengthen resilience against evolving cyber threats.
- Scalability across Various Smart Grids Architecture: This has been validated for IEEE 5-bus, 14-bus, and 118-bus networks and manifests the capability for generalization to different sizes and configurations of the grid.
- Benchmarking Against Strong Baselines: Extensive performance evaluation is conducted against an Autoencoder-based FDIA detector as well as state-of-the-art transformer time-series baselines, including TimesNet, PatchTST, and FEDformer, highlighting relative strengths and deployment trade-offs in terms of detection accuracy, false alarm rate, robustness, and inference latency.

The rest of the paper is organized as follows. Section II presents the proposed methodology, which consists of FDIA attack simulation, informer-based detection and benchmark model formulations, multi-objective optimization using NSGA-II with Bayesian hyperparameter refinement, and smart grid system modeling. The experimental setup and implementation details are described in Section III. This includes thorough benchmarking against reconstruction-based and transformer-based time-series baselines (Autoencoder,

TimesNet, PatchTST, and FEDformer), as well as assessments of system capacity, defense strength, and optimization performance across various grid sizes under operational conditions. While Section IV provides a summary of the key findings and conclusions of this work, Section V addresses the shortcomings of the suggested methodology and suggests avenues for further investigation.

II. Proposed Methodology

To tackle the detection and countermeasures of FDIAs in smart grids in an organized manner, our suggested approach, as shown in Fig. 1, is organized into different modules, each of which deals with an important function of the detection system. We elaborate on each of the modules in this section, describing their functions in facilitating adequate detection, optimization, and defense against malicious attacks.

A. Smart Grid System Model and FDIA Simulation

Smart grids integrate digital communication technologies and distributed energy resources (DERs) such as solar, wind, and electric vehicles, aiming to optimize reliability and efficiency. Their two-way data and energy flows, monitored via devices like PMUs, RTUs, and SCADA systems, increase operational visibility but also introduce significant cyber attack vectors. Among these, FDIAs are particularly insidious as they alter sensor data to deceive state estimation algorithms without violating physical system constraints [32].

To model these threats, we consider the AC power flow equation:

$$z = h(x) + e, \tag{1}$$

where x denotes system states, z the measurements, and $e \sim \mathcal{N}(0, R)$ is Gaussian noise. State estimation uses WLS optimization:

$$\hat{x} = \arg\min_{x} (z - h(x))^T R^{-1} (z - h(x)). \tag{2}$$

FDIAs inject an attack vector a such that $z^* = z + a$ while ensuring stealth:

$$\|z^* - h(\hat{x})\| \leq \tau, \tag{3}$$

ensuring that the injected false measurements remain stealthy with respect to residual-based bad data detection (BDD). When attackers have access to the Jacobian matrix H, they can construct $a = Hc$ to ensure residuals remain below the detection threshold.

In our simulation, IEEE 5-bus, 14-bus, and 118-bus systems are used to inject FDIAs that replicate realistic operational conditions. The injection process follows four practical steps:

1) Perform baseline WLS state estimation to obtain $\hat{x}$.
2) Generate adversarial perturbations using gradient-based techniques (FGSM, BIM, MIM).
3) Modify raw measurements to simulate corrupted input $z^* = z + a$.
4) Validate stealth by ensuring residuals $\|z^* - h(\hat{x})\|$ remain below threshold τ.

These steps allow us to evaluate our FDIA detection framework against sophisticated, stealth-aware adversarial threats.

FDIA Injection Protocol and Attack Configuration: Every FDIA is produced offline using simulated or recorded smart grid measurement sequences. This makes it possible to reproduce adversarial samples and have exact control over attack parameters. In Section III-F, the viability of real-time (streaming) deployment is assessed independently using latency and computational throughput analysis.

Measurement channels that are frequently accessible in PMU and SCADA systems, such as voltage magnitudes, voltage phase angles, and active and reactive power flow measurements, are attacked. In each attack instance, a sparse subset of sensors or buses is compromised, where k buses (or equivalently $p\%$ of measurement channels) are randomly selected to be attacked within a given time window, reflecting realistic attacker resource constraints.

Each FDIA is injected over a finite temporal duration. Specifically, attacks persist for a window of length W samples (corresponding to several seconds of operation), after which measurements revert to normal behavior. This temporal structure allows evaluation of both short-lived and sustained stealthy attacks.

Different adversarial generation methods produce distinct attack characteristics. FGSM generates a single-step perturbation constrained by an ℓ_∞ bound ϵ, resulting in fast but less adaptive attacks. BIM produces more precise perturbations by applying iterative updates with step size α within the same ϵ bound. By adding momentum to the iterative process, MIM creates attack samples that are more evasive and transferable, better simulating coordinated attacker behavior.

Every generated attack sample is verified against the residual constraint in Eq. (3) to guarantee stealthiness. If the residual exceeds the threshold τ, the stealth condition is satisfied by adaptively resampling or scaling down the perturbation magnitude. By doing this, all injected attacks are guaranteed to evade conventional residual-based detectors, producing challenging and realistic FDIA scenarios. For clarity and reproducibility, the complete FDIA generation and validation procedure is summarized in Algorithm 1.

Algorithm 1 FDIA Injection Protocol

Require: Measurement sequence z, state estimator $h(\cdot)$
Require: Attack method $\mathcal{A} \in \{\text{FGSM, BIM, MIM}\}$
Require: Stealth threshold τ, attack duration W, compromised buses k
Perform baseline WLS state estimation to obtain $\hat{x}$
Randomly select k buses (or $p\%$ of channels) to compromise
for each attack window of length W **do**
    Generate perturbation a using method $\mathcal{A}$
    Inject attack: $z^* = z + a$
    if $\|z^* - h(\hat{x})\| > \tau$ **then**
        Scale or resample a until stealth constraint is satisfied
    end if
end for
Ensure: Stealthy FDIA-injected measurement sequence z^*

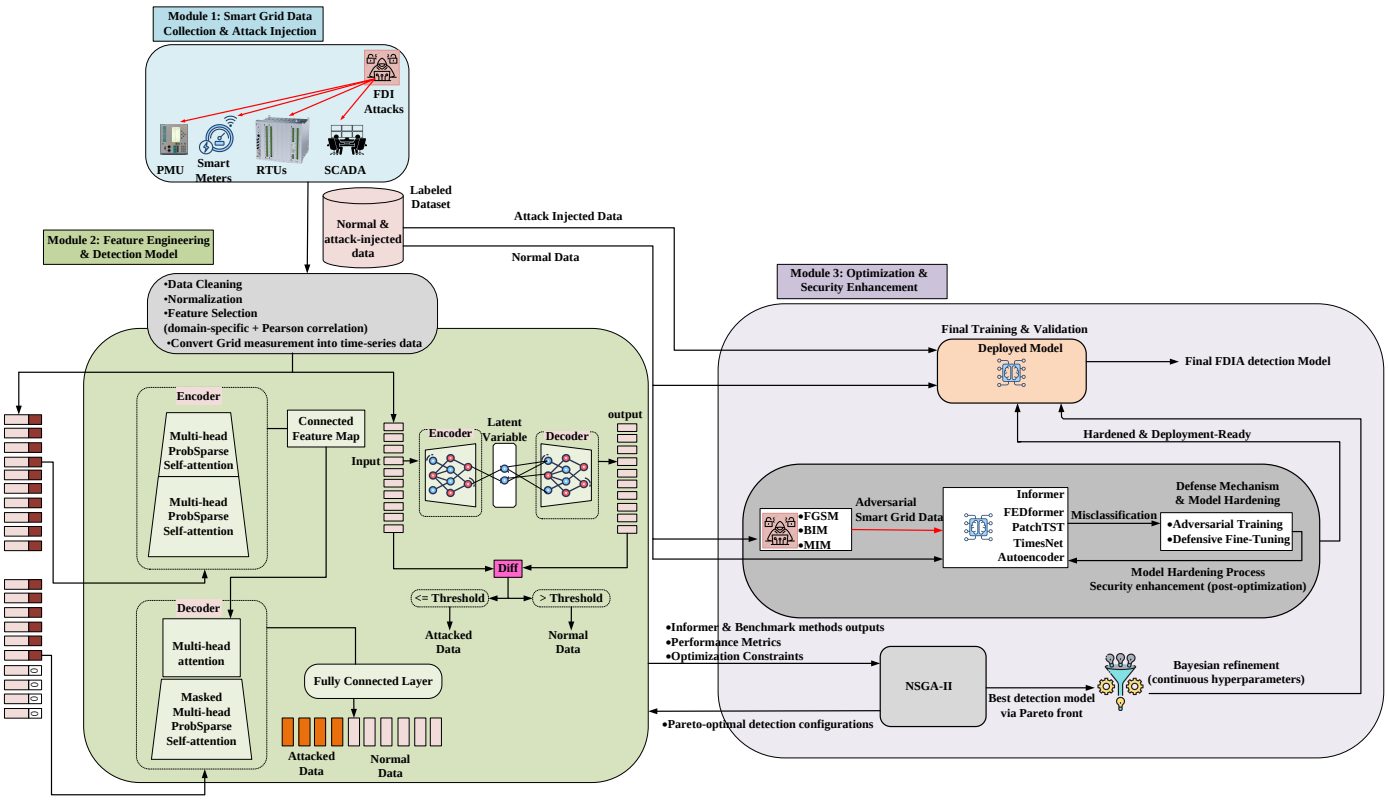


Fig. 1: An overview of the proposed AI-driven FDIA detection framework. Adversarial attacks (FGSM, BIM, MIM) are injected offline into smart grid measurement streams over finite temporal windows. NSGA-II performs global multi-objective optimization over feature subsets, detection thresholds, and model hyperparameters, while Bayesian optimization locally refines continuous hyperparameters.

B. Informer Model for FDIA Detection

The Informer model is a state-of-the-art deep learning model for long-term time-series forecasting and, therefore, of immense utility to FDIA detection in smart grids. Unlike typical recurrent neural networks (RNNs) or transformer-based models, Informer uses a ProbSparse self-attention mechanism, significantly lowering the computational overhead while being highly precise at anomaly detection.

1) Informer Model Architecture: The Informer model consists of the following key components:

- Encoder-Decoder Structure for processing time-series.
- ProbSparse Self-Attention for effectively modeling global dependencies.
- Conv1D & Max-Pooling Layers to identify high-level temporal features.

Given an input sequence of power system measurements over time:

$$\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_T] \in \mathbb{R}^{T \times d} \tag{4}$$

where T is the time horizon and d is the feature size. The Informer model provides a predicted future state:

$$\hat{\mathbf{X}} = Informer(\mathbf{X}) \tag{5}$$

When the discrepancy between the predicted state $\hat{\mathbf{X}}$ and the real observed state X^* is greater than a certain threshold δ, an FDIA is identified

$$\|\hat{\mathbf{X}} - \mathbf{X}^*\| \geq \delta \tag{6}$$

To achieve robustness, the Informer model is trained with a mix of supervised learning and adversarial data augmentation.

$$\mathcal{L} = \sum_{i=1}^{N} \|\mathbf{x}_i - \hat{\mathbf{x}}_i\|^2 + \lambda \sum_{i=1}^{N} \|\mathbf{x}_i^{adv} - \hat{\mathbf{x}}_i\|^2 \tag{7}$$

where $\mathcal{L}$ is the loss function (mean squared error + adversarial regularization), λ is the adversarial loss weighting factor, $\mathbf{x}_i^{adv}$ is adversarially perturbed data based on FGSM, BIM, or MIM attacks. With the Informer's capability to capture long-range dependencies, the detection model is able to detect stealthy FDIAs, which would otherwise go undetected.

C. Benchmark FDIA Detection Models

To ensure a rigorous and fair evaluation, the proposed FDIA detection framework is benchmarked against representative models from both reconstruction-based anomaly detection and state-of-the-art Transformer-based time-series learning paradigms. This subsection briefly introduces the comparative models and their core mathematical formulations to maintain methodological consistency with the proposed approach.

1) Autoencoder-Based FDIA Detection: Autoencoders are widely adopted for anomaly detection in cyber-physical systems due to their ability to learn compact representations of normal operating conditions in an unsupervised manner. An Autoencoder consists of an encoder-decoder pair defined as

$$f : \mathbb{R}^d \to \mathbb{R}^{d'}, \quad g : \mathbb{R}^{d'} \to \mathbb{R}^d, \tag{8}$$

where the encoder $f(\cdot)$ compresses the input measurement vector $\mathbf{x}$ into a latent representation, and the decoder $g(\cdot)$ reconstructs the input as

$$\hat{\mathbf{x}} = g(f(\mathbf{x})). \tag{9}$$

FDIA detection is performed by thresholding the reconstruction error:

$$\|\mathbf{x} - \hat{\mathbf{x}}\|_2 \geq \theta \Rightarrow \text{FDIA detected}, \tag{10}$$

where θ is a predefined detection threshold. The Autoencoder is trained by minimizing the mean squared reconstruction loss

$$\mathcal{L}_{\text{AE}} = \sum_{i=1}^{N} \|\mathbf{x}_i - g(f(\mathbf{x}_i))\|_2^2. \tag{11}$$

While effective for detecting gross anomalies, Autoencoders exhibit limited robustness against adaptive and stealthy FDIAs that preserve reconstruction statistics.

2) TimesNet-Based FDIA Detection: TimesNet is a recent state-of-the-art time-series backbone that models temporal variations by transforming one-dimensional sequences into structured two-dimensional representations based on multi-periodicity analysis. Given a multivariate time series $\mathbf{X} \in \mathbb{R}^{T \times C}$, dominant periods are identified via frequency-domain analysis:

$$\{p_k\} = \text{TopK}\big(\text{FFT}(\mathbf{X})\big), \tag{12}$$

where $\{p_k\}$ denotes the detected period lengths. The original sequence is reshaped into a set of 2D tensors $\mathbf{X}_{\text{2D}}^{(k)} \in \mathbb{R}^{p_k \times (T/p_k) \times C}$, enabling simultaneous modeling of intra-period and inter-period variations. A TimesBlock extracts temporal features via 2D convolutional kernels, and the output representation is obtained as

$$\mathbf{Z} = \sum_k \alpha_k \cdot \Phi(\mathbf{X}_{\text{2D}}^{(k)}), \tag{13}$$

where α_k are frequency-adaptive weights and $\Phi(\cdot)$ denotes the 2D feature extraction operator. FDIA detection is achieved by identifying deviations in the learned temporal variation patterns. Detailed architectural descriptions can be found in [33].

3) PatchTST-Based FDIA Detection: PatchTST improves Transformer-based time-series modeling by segmenting the input sequence into local temporal patches. For each channel c, the time series $\mathbf{x}^{(c)} \in \mathbb{R}^T$ is divided into N patches of length P:

$$\mathbf{x}^{(c)} \to \{\mathbf{x}_1^{(c)}, \ldots, \mathbf{x}_N^{(c)}\}, \quad \mathbf{x}_i^{(c)} \in \mathbb{R}^P. \tag{14}$$

Each patch is linearly projected into an embedding space and processed by a Transformer encoder:

$$\mathbf{Z} = \text{Transformer}\big(\mathbf{W}_p \mathbf{X}_{\text{patch}} + \mathbf{E}_{\text{pos}}\big). \tag{15}$$

FDIA detection is formulated by monitoring prediction or representation errors over reconstructed or forecasted patches. Patch-based tokenization significantly reduces attention complexity while preserving local temporal semantics. Further details are available in [34].

4) FEDformer-Based FDIA Detection: FEDformer enhances Transformer efficiency by operating in the frequency domain and incorporating seasonal–trend decomposition. Given a time series $\mathbf{X}$, Fourier transform is applied as

$$\mathbf{X}_f = \mathcal{F}(\mathbf{X}), \tag{16}$$

and a sparse subset of frequency components is selected to construct a compact representation. The model decomposes the signal into seasonal and trend components:

$$\mathbf{X} = \mathbf{S} + \mathbf{T}, \tag{17}$$

which are processed through frequency-enhanced Transformer blocks. FDIA detection is achieved by identifying inconsistencies in predicted temporal dynamics across frequency bands. FEDformer achieves linear computational complexity with respect to sequence length. Comprehensive architectural details are provided in [35].

D. Multi-Objective Optimization Framework

Model parameters (network weights) are learned using standard gradient-based optimization (Adam optimizer in PyTorch). The multi-objective optimization framework does not replace model training; instead, NSGA-II operates outside the weight-training loop to select optimal detection configurations, including feature subsets, decision thresholds, and selected hyperparameter settings that define candidate models.

We optimize the FDIA detection model using NSGA-II to balance detection accuracy, false alarm rate, and computational efficiency.

1) Optimization Objectives: The optimization framework seeks to jointly optimize the following objective functions:

1) Maximize the detection rate (Recall / ADR) ($\mathcal{F}_1$):

$$\mathcal{F}_1 = \frac{TP}{TP + FN} \tag{18}$$

2) Minimize FPR ($\mathcal{F}_2$):

$$\mathcal{F}_2 = \frac{FP}{FP + TN} \tag{19}$$

3) Minimize computational complexity ($\mathcal{F}_3$):

$$\mathcal{F}_3 = \frac{C_{comp}}{T_{eval}} \tag{20}$$

Where TP, FP, TN, and FN stand for true positives, false positives, true negatives, and false negatives, respectively, C_{comp} denotes the computational cost, and T_{eval} represents the model evaluation time.

The decision variable vector optimized by NSGA-II is explicitly defined as

$$\boldsymbol{\Theta} = [\boldsymbol{\theta}_{\text{model}}, \theta_\delta, \boldsymbol{\theta}_{\text{feat}}], \tag{21}$$

where $\boldsymbol{\theta}_{\text{model}}$ denotes the Informer architecture and training hyperparameters, in which discrete architectural choices (e.g.,

embedding dimension d_{model} and number of attention heads) are handled within the NSGA-II search, while continuous-valued training parameters (e.g., learning rate, dropout rate, and adversarial loss weight λ) are refined via Bayesian optimization; θ_δ represents the detection threshold used to convert prediction errors into binary attack decisions; and $\boldsymbol{\theta}_{\text{feat}}$ corresponds to the selected feature subset, such as a binary mask or top-K features.

Threshold Optimization within multi-objective optimization: The detection threshold θ_δ is treated as an explicit decision variable within the NSGA-II process. For each candidate configuration $\boldsymbol{\Theta}$, θ_δ is evaluated on a labeled validation set to compute $\mathcal{F}_1$ (ADR), $\mathcal{F}_2$ (FPR), and $\mathcal{F}_3$ (inference latency). Threshold selection is therefore jointly optimized with feature subsets and Bayesian optimization-refined hyperparameters, rather than being tuned independently.

Bayesian optimization is embedded as an auxiliary local-search step within the NSGA-II evaluation loop to refine continuous-valued hyperparameters for each candidate configuration. NSGA-II subsequently performs global multi-objective optimization over the combined decision space $\boldsymbol{\Theta}$, explicitly balancing ADR, FPR, and inference latency.

Remark on Optimization Objectives and Evaluation Metrics: Although accuracy is reported for completeness and comparability with prior work, it is not used as an optimization objective due to its sensitivity to class imbalance in FDIA detection. Accordingly, only ADR ($\mathcal{F}_1$), FPR ($\mathcal{F}_2$), and inference latency ($\mathcal{F}_3$) are optimized, while accuracy is computed post hoc on the held-out test set for evaluation purposes.

NSGA-II identifies a set of non-dominated solutions, and the optimal detection configuration is selected from the Pareto front as

$$\boldsymbol{\Theta}^* = \arg\max_{\boldsymbol{\Theta}} [\mathcal{F}_1, -\mathcal{F}_2, -\mathcal{F}_3]. \tag{22}$$

The resulting Pareto-optimal solution provides a principled balance between attack detectability, false alarms, and real-time computational feasibility.

E. Adversarial FDIA Attacks and Defense Mechanism

To assess resilience under evolving threats, we simulate adversarial FDIAs that strategically manipulate power grid measurements. These attacks target the input of the detection models by introducing small, structured perturbations to deceive both statistical estimators and AI-based detectors.

In practice, this corresponds to modifying voltage, power flow, and phase angle readings from PMUs and RTUs such that the attack vector a remains undetected by residual-based methods, yet leads to incorrect state estimation and possible grid misoperation. We implement the following attack strategies during simulation:

1) FGSM: This method perturbs input data x in a single step by maximizing the loss function gradient:

$$x^{\text{adv}} = x + \epsilon \cdot \text{sign}(\nabla_x L), \tag{23}$$

where ϵ controls the attack strength. FGSM introduces rapid, lightweight perturbations to voltage or power data, sufficient to mislead deep learning detectors without breaching grid thresholds.

2) BIM: An extension of FGSM, BIM applies multiple small perturbations iteratively:

$$x_{t+1}^{\text{adv}} = x_t^{\text{adv}} + \alpha \cdot \text{sign}(\nabla_x L), \tag{24}$$

increasing stealth by distributing the attack over time-series windows in PMU data.

3) MIM: This variant improves on BIM by incorporating momentum, leading to more persistent perturbations:

$$g_{t+1} = \mu \cdot g_t + \frac{\nabla_x L}{\|\nabla_x L\|_1}, \quad x_{t+1}^{\text{adv}} = x_t^{\text{adv}} + \alpha \cdot \text{sign}(g_{t+1}), \tag{25}$$

with μ as the decay factor. MIM produces highly evasive patterns, mimicking small, consistent fluctuations in grid parameters.

Adversarial Attack Parameterization: To ensure controlled and reproducible evaluation of attack strength, all adversarial perturbations are generated on normalized measurement sequences. For FGSM, the perturbation magnitude ϵ is selected from the set $\epsilon \in \{0.005, 0.01, 0.02, 0.05\}$, covering weak to strong attack intensities commonly used in power-system cyber attack studies. For BIM and MIM, iterative attacks are configured with a step size $\alpha = \epsilon/T$ and a fixed number of iterations $T = 10$, with projection onto the ℓ_∞-ball to enforce the same maximum perturbation budget as FGSM. For MIM, a momentum factor $\mu = 0.9$ is used to enhance attack transferability and evasiveness. All attacks are applied on a per-window basis to the selected measurement channels described in Section II-A, ensuring consistent temporal structure and comparable attack intensity across methods.

These attacks are applied to time-series data streams mimicking PMU signals. For each method, we generate adversarial examples and inject them into the testbed grids to emulate cyber-physical attack scenarios. Attack success is evaluated by the inability of classical BDD or untrained AI models to detect the perturbations while causing observable deviations in the power system state. This ensures a realistic assessment of both stealth and disruption potential.

To mitigate these attacks, we employ adversarial training, augmenting model training with adversarially perturbed data, and threshold fine-tuning, adjusting the detection boundary δ to balance sensitivity and false positives. To mitigate these attacks, we employ adversarial training by augmenting model learning with adversarially perturbed samples, together with threshold fine-tuning to balance sensitivity and false positives. The adversarially regularized loss function used during model hardening is identical to that defined in Eq. (7), which jointly minimizes prediction error on clean samples and robustness loss on adversarially perturbed inputs. The adversarial loss weight λ controls the trade-off between detection performance under normal operation and resilience against FGSM-, BIM-, and MIM-based attacks.

F. Feature Engineering and Preprocessing

Preprocessing and feature engineering are crucial stages in FDIA detection because they enhance data quality, remove noise, and choose appropriate features for the Informer and Autoencoder models, and other transformer-based benchmark

models. In order to improve detectability, we outline the feature extraction and selection procedures as well as the data pretreatment methods in this part.

1) Data Preprocessing: The raw smart grid data from PMUs, SCADAs, and RTUs is collected and converted into a more structured format so that the model is able to process it more efficiently. There are several steps of cleaning that are done to the dataset in order to add uniformity and decrease noise before running it through the models for detection. All preprocessing steps are applied independently within each dataset split and are performed after attack injection to preserve the statistical characteristics of FDIA-contaminated samples.

1) Data Cleaning
   - Dealing with Missing Values: To preserve the continuity of time-series data, missing sensor readings are imputed using a linear interpolation method:

$$x_t = x_{t-1} + \frac{(x_{t+1} - x_{t-1})}{2}. \tag{26}$$

   - Outlier detection and removal: We use the z-score method to suppress spurious measurement noise and simulation artifacts, while preserving FDIA-induced deviations that satisfy the stealth constraint. A feature x is labeled as an outlier if:

$$Z = \frac{x - \mu}{\sigma} > 3. \tag{27}$$

   Where μ and σ are the mean and standard deviation of the feature.

2) Data Normalization
   Min-Max feature scaling is performed to normalize all features for model convergence:

$$x' = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \tag{28}$$

   This prevents larger magnitude features from dominating the set and makes sure all the variables are in the range [0; 1].

2) Feature Extraction: The original smart grid data is a high-dimensional signal, including:

- Voltage Magnitude V_i
- Phase Angle θ
- Active Power P_i and Reactive Power Q_i

On the other hand, using these features directly might not be the best way. We apply Fourier transform (FFT) for frequency domain analysis and wavelet transform (WT) for transient detection. FFT encodes periodic trends in smart grid signals as follows:

$$X(f) = \sum_{n=0}^{N-1} x_n e^{-j2\pi fn/N}. \tag{29}$$

Where the Fourier-transformed signal at frequency f is denoted by $X(f)$. N is the number of data points. The complex sinusoidal basis functions are represented by the exponential term $e^{-j2\pi fn/N}$.

3) Feature Selection: Irrelevant features are eliminated using the following methods to enhance attack detectability (ADR) while reducing computational overhead (inference latency):

- Mutual information (MI): Calculates the information exchange between attack label Y and feature X.

$$I(X,Y) = \sum_{x\in X}\sum_{y\in Y} p(x,y)\log\frac{p(x,y)}{p(x)p(y)} \tag{30}$$

- Principal component analysis (PCA): preserves variance while reducing feature dimensionality:

$$\mathbf{Z} = \mathbf{X}W \tag{31}$$

  where W is the eigenvector transformation matrix, X represents the original data matrix, and Z is the transformed data.

G. Training and Validation Strategy

To ensure reliable attack detectability and robustness against adversarial attacks, the proposed detection framework is trained and evaluated using a systematic training and validation strategy consisting of the following modules.

1) Dataset Preparation: The dataset includes normal smart grid and FDIA-infected data (adversarial data from FGSM, BIM, and MIM).

1) Data Splitting: We divided the dataset into training, validation, and testing, including 70%, 15%, and 15% of the whole data, respectively.
2) Data Augmentation: To address class imbalance during model learning, oversampling techniques such as the synthetic minority oversampling technique (SMOTE) are applied exclusively to the training subset. No synthetic samples are introduced into the validation or test sets, which consist solely of physically simulated measurements, ensuring unbiased performance evaluation and preventing data leakage.

2) Model Training Strategy: All models are trained using the same training–validation splits described above to ensure a fair and consistent comparison. The proposed Informer-based detector and the benchmark models differ only in their learning objectives and robustness configurations, as detailed below.

1) Informer Model Training: Eq. (7) defines the adversarial regularized loss function that is used to train the Informer model. It simultaneously minimizes the reconstruction error on adversarial perturbed inputs produced by FGSM, BIM, and MIM attacks as well as the prediction error on clean samples. The trade-off between robustness under adversarial perturbations and clean-data performance is managed by the adversarial loss weight λ.
   Adversarial Training Configuration: During adversarial training, clean and adversarial samples are mixed within each training epoch using a fixed ratio of 70% clean samples and 30% adversarial samples. Adversarial examples are generated using a mixture of FGSM, BIM, and MIM attacks to improve robustness against both

single-step and iterative perturbations. The adversarial loss weight λ is treated as a tunable hyperparameter and is selected using Bayesian optimization on the validation set, with $\lambda \in [0.1, 1.0]$. This configuration balances ADR on clean data with robustness under adversarial perturbations.

2) Autoencoder Model Training: The Autoencoder baseline is trained using the reconstruction loss defined in Eq. (11), which minimizes the squared error between the input measurements and their reconstructions obtained from the latent representation under normal operating conditions.
3) Transformer-Based Benchmark Model Training (TimesNet, PatchTST, FEDformer): Following their standard formulation as state-of-the-art forecasting architectures, the Transformer-based benchmark models, such as TimesNet, PatchTST, and FEDformer, are trained as predictive time-series models under typical operating conditions. By minimizing a mean squared prediction error, each model learns to forecast future measurements from past input sequences. To represent their usual use in time-series analysis, these benchmark models are trained without adversarial augmentation. During inference, FDIA detection is performed by thresholding the prediction error, consistent with the anomaly detection protocol described in Section II-A.

3) Model Validation and Performance Metrics: Model validation and performance evaluation are conducted in strict accordance with the multi-objective optimization formulation defined in Section II-D. As detailed in Eqs. (18)–(20), the proposed framework jointly optimizes the attack ADR, FPR, and computational efficiency, and these same criteria are therefore adopted as the primary performance metrics during validation and testing.

For each candidate detection configuration $\mathbf{\Theta}$, FDIA detection is performed by thresholding the model prediction or reconstruction error using the optimized decision threshold θ_δ. The resulting binary decisions are used to compute the detection rate under adversarial conditions (ADR), corresponding to the optimization objective $\mathcal{F}_1$ in Eq. (18), and FPR, corresponding to $\mathcal{F}_2$ in Eq. (19). Computational efficiency is quantified by the end-to-end per-sample inference latency, which serves as the operational measure of $\mathcal{F}_3$ in Eq. (20).

During NSGA-II optimization, these three metrics are evaluated on the validation set to guide Pareto-front construction, while final performance is reported on a held-out test set using the same metrics for the selected non-dominated solution. This evaluation strategy ensures full consistency between the optimization objectives and the reported experimental results in Section III, avoiding any mismatch between training, validation, and deployment-oriented performance assessment.

III. Results and Discussion

A. Implementation and Experiment Setup

We evaluated our Informer-based FDIA detection algorithm in a multi-scale environment that emulated high-fidelity smart grid attack scenarios. Three IEEE-standard benchmark networks were used to test our approaches. The IEEE 5-bus system is used for baseline testing to measure basic performance. A medium-sized grid with numerous connections, the IEEE 14-bus system provides a more intricate testing environment. Finally, the large-scale network that closely resembles actual smart grids is the IEEE 118-bus system.

We integrated MATPOWER 7.1 into a Python-based machine learning pipeline for power flow analysis and state estimates. While NSGA-II optimization was carried out using the distributed evolutionary algorithms in Python (DEAP) framework, our models were constructed using PyTorch 2.0. The hardware configuration of our experiment environment consisted of a (24GB VRAM) NVIDIA RTX 3090 GPU, 128GB of RAM, an Intel Core i9-12900K CPU, and LTS for Ubuntu 22.04. We created a simulation-based dataset using conventional IEEE test networks. We used well-established attack injection techniques from adversarial attack research to faithfully replicate cyber threats. In particular, we created realistic FDIAs that closely mirrored actual security concerns using gradient-based adversarial approaches.

Based on the simulation and experimental environment described above, the datasets used for training, validation, and testing are summarized as follows.

Dataset Summary: For each IEEE benchmark network (5-bus, 14-bus, and 118-bus), we construct simulation-based multivariate time-series datasets using MATPOWER, where FDIA samples are generated offline following the attack protocol described above, physically feasible operating points are obtained through AC power flow analysis, and system states are estimated via WLS at each time step. For each grid, 200 independent time-series sequences are produced, each consisting of 500 consecutive samples, resulting in a total of 100,000 labeled samples per grid comprising both normal and FDIA-contaminated measurements. The measurement channels correspond to standard PMU/SCADA observations, including voltage magnitudes V, voltage phase angles θ, and active and reactive power flows $\{P, Q\}$. Measurement noise is modeled as additive Gaussian noise $e \sim \mathcal{N}(0, R)$ with $R = \sigma^2 I$ and $\sigma = 0.01$ p.u., and is applied to all channels prior to attack injection to emulate realistic sensing uncertainty. Stealthy FDIA labels are assigned to temporal windows in which adversarial perturbations are injected using FGSM, BIM, or MIM, while strictly enforcing the stealth constraint defined in Eq. 3. Unless otherwise stated, each attack compromises a sparse subset of $k = 3$ buses and persists for a duration of $W = 50$ samples. The dataset is partitioned into 70% training, 15% validation, and 15% testing subsets, and all preprocessing and attack generation procedures are conducted independently within each split to prevent information leakage. For leave-one-attack-out (LOAO) evaluation, experimental results are reported as mean$\pm$standard deviation over three independent random seeds.

Unless otherwise stated, all experiments assume a sparse attack setting with $k = 3$ compromised buses per attack instance and a fixed attack duration of $W = 50$ time samples.

Cross-Attack Generalization Protocol: In addition to the standard 70%/15%/15% train-validation-test split, we explic-

itly evaluate the generalization capability of the proposed detector under a *cross-attack* setting to mitigate overfitting to specific adversarial patterns. A LOAO protocol is adopted, in which the model is trained and validated on normal data together with two attack types (e.g., FGSM+BIM) and evaluated on an unseen attack type (e.g., MIM) that was never observed during training or validation; this process is repeated for all attack combinations. Early stopping is applied based on validation ADR and FPR to prevent overfitting, and all cross-attack results are reported as mean±standard deviation over three random seeds.

TABLE II: Feature Selection Performance Analysis

Feature	MI	PCA Contribution (%)
Voltage Magnitude (V_i)	0.89	28.4
Phase Angle (θ_i)	0.82	22.1
Active Power (P_i)	0.78	17.5
Reactive Power (Q_i)	0.72	15.8
FFT Peak Frequency	0.67	9.2
Statistical Features (Mean, Variance, Skewness, Kurtosis)	0.55	7.0

B. Feature Engineering and Feature Selection

To achieve better performance of FDIA detection, we employed a rigorous feature engineering and selection strategy based on the methods outlined in Section II.

Feature engineering was employed to transform raw measurements of power systems into informative structured data for ingestion by the detection models. Transformations employed included time-series feature extraction, frequency-domain processing, and statistical feature generation, as described in Section II. The combination of FFT and WT features was shown to be quite useful for identifying periodic and transient anomalies, respectively. Moreover, statistical features such as mean, variance, skewness, and kurtosis were shown to be very useful in separating normal system oscillation from malicious attacks. MI and PCA were employed to rank and assign higher priority to the most suitable features. Prior to NSGA-II optimization, Bayesian optimization was utilized to refine continuous model hyperparameters (e.g., learning rate, dropout, and adversarial loss weight λ) for each candidate configuration, while the feature subset itself was selected through the discrete MI/PCA ranking and the NSGA-II multi-objective search over $\{\theta_{\text{feat}}, \theta_{\delta}, \theta_{\text{model}}\}$.

In other words, MI and PCA provide an initial feature importance ordering, whereas NSGA-II determines the final deployment-ready subset size and composition under ADR–FPR–latency trade-offs.

The effectiveness of this feature selection process is depicted in Table II, which presents the MI values and PCA contribution percentages for each selected feature. As it can be observed, Voltage Magnitude (V_i) and Phase Angle (θ_i) contributed most significantly to FDIA detection with MI values of 0.89 and 0.82, respectively. Active Power (P_i) and Reactive Power (Q_i) ranked second, both having an MI value of 0.78 and 0.72, respectively. The incorporation of FFT Peak Frequency (MI = 0.67) and the statistical properties of Mean, Variance, Skewness, and Kurtosis (MI = 0.55) continued to enhance the model's robustness to distinguish normal variability and adversarial manipulations.

C. Performance Comparison with State-of-the-Art Time-Series Baselines

We outline a comparative evaluation of the Informer-based FDIA detection model and the other benchmark methods to contrast their respective strengths and weaknesses. The comparison is on the basis of their key performance metrics, i.e., FPR, ADR, and inference latency. Although accuracy is reported for completeness and comparability with prior work, it is not used as an optimization objective due to its sensitivity to class imbalance in FDIA detection.

In addition to the Autoencoder baseline, the Informer model is further benchmarked against multiple state-of-the-art Transformer-based time-series detection models, including TimesNet, PatchTST, and FEDformer, in order to provide a comprehensive and fair performance assessment.

All benchmark models were trained under identical data splits, hyperparameter tuning protocols, and attack scenarios, following the procedure in Sections III-A and III-B, to ensure a consistent and unbiased comparison. For benchmarking, the models were evaluated under normal operation, stealth FDIA attacks, and adversarially crafted attack environments (FGSM, BIM, MIM).

Table III presents a detailed quantitative comparison between the proposed Informer-based detector, the Autoencoder baseline, and state-of-the-art Transformer-based time-series models under the same experimental configuration. This table reports end-to-end per-sample detection latency on the IEEE 14-bus system, including model inference, anomaly scoring, and threshold-based decision. As seen in Table III, the Informer model achieved a considerably higher accuracy of 93.5% compared to the Autoencoder's 85.7%, indicating its superior ability to correctly distinguish between normal and compromised grid states. Compared with advanced Transformer baselines, Informer also demonstrates consistent gains in accuracy, outperforming TimesNet (91.6%), PatchTST (92.8%), and FEDformer (92.1%).

The Informer model further excelled in ADR, reaching 95.2% compared to 82.1% for the Autoencoder, indicating a substantially improved capability to detect both stealthy and adversarially crafted FDIA patterns. Notably, the ADR achieved by Informer also exceeds those of TimesNet (90.3%), PatchTST (93.1%), and FEDformer (92.4%), confirming its robustness in identifying complex FDIA dynamics in time-series grid measurements.

The observed ADR gains result from model design objectives which do not match FDIA characteristics. TimesNet develops its forecasting ability through its design, which uses frequency-domain analysis to achieve both periodicity and multi-scale temporal decomposition. The system can forecast structured variations, but it struggles to detect minor FDIA disturbances which maintain their main periodic patterns.

PatchTST uses localized temporal patch representations to manage attention complexity, but this approach prevents it from understanding long-range and cross-channel dependencies which arise during coordinated FDIA attacks. FEDformer operates largely in the frequency domain via seasonal–trend decomposition, and therefore may be less sensitive to stealthy manipulations that maintain spectral energy distributions while altering temporal consistency across measurement channels. In contrast, the Informer's ProbSparse self-attention prioritizes globally informative temporal dependencies with reduced complexity, supporting stronger detection under both stealthy and adversarial conditions, consistent with the FPR and latency trends reported below.

One of the significant requirements of FDIA detection is minimizing false positives to avoid unnecessary or destabilizing corrective actions. The FPR of the Informer (5.1%) was significantly lower than that of the Autoencoder (13.4%). Moreover, Informer achieves the lowest FPR among all evaluated models, outperforming TimesNet (7.8%), PatchTST (6.3%), and FEDformer (6.9%), which is critical for reliable deployment in operational power systems. All results in Table III report final performance on the held-out test set using the selected configuration obtained after NSGA-II convergence.

Inference latency is a key consideration for real-time FDIA detection under high-frequency PMU data streams. The Informer model processes samples faster (4.8 ms/sample) than the Autoencoder (7.9 ms/sample), as illustrated in Fig. 2. As further shown in Fig. 2, Informer also exhibits lower inference latency than TimesNet (9.4 ms), PatchTST (8.7 ms), and FEDformer (10.1 ms), indicating a more favorable balance between detection performance and computational efficiency.

Overall, the results in Table III and Fig. 2 demonstrate that the proposed Informer-based detector consistently achieves superior performance compared with both reconstruction-based and state-of-the-art Transformer-based baselines across accuracy, ADR, FPR, and inference latency. This advantage stems from Informer's ability to efficiently model long-range temporal dependencies while maintaining low computational overhead, yielding a more favorable accuracy–latency trade-off for real-time FDIA detection.

TABLE III: Performance Comparison with State-of-the-Art Time-Series Baselines. Inference latency reports end-to-end per-sample detection latency, including model forward pass, anomaly score computation, and threshold-based decision, evaluated on the IEEE 14-bus system with batch size = 1.

Model	Accuracy (%)	ADR (%)	FPR (%)	Inference Latency (ms/sample)
Autoencoder	85.7	82.1	13.4	7.9
TimesNet	91.6	90.3	7.8	9.4
PatchTST	92.8	93.1	6.3	8.7
FEDformer	92.1	92.4	6.9	10.1
Informer (Proposed)	**93.5**	**95.2**	**5.1**	**4.8**

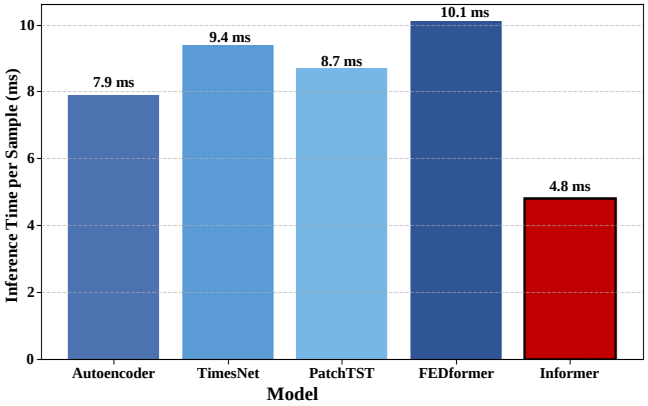


Fig. 2: Inference latency comparison across Autoencoder-based, state-of-the-art Transformer-based, and the proposed Informer-based FDIA detection models.

D. Impact of Multi-Objective Optimization

This section includes pre- and post-optimization results to evaluate the effect of NSGA-II on model performance. Optimization by means of NSGA-II was done for a population size of 100 for 50 generations, with a crossover probability of 0.8 and a mutation probability of 0.1. Model hyperparameters, feature subset selection, and decision thresholds were within the optimization search space to find an optimal trade-off between competing objectives. To assess the impact of optimization, we compared the performance of the pre-optimized model (default hyperparameters) and the NSGA-II optimized model. Table IV shows the gain in performance from utilizing multi-objective optimization. This table reports validation-stage metrics observed during NSGA-II search (for monitoring/comparability), whereas Table III reports final held-out test-set results for the selected converged configuration. The result indicates that multi-objective optimization significantly improved detection accuracy from 89.6% to 95.1% and reduced FPR from 10.8% to 4.9%. Besides, computational efficiency was also improved with the reduction of inference latency from 6.5 ms to 4.3 ms, and thus, the model became more suitable for real-time applications. This reduction in inference latency after NSGA-II optimization is achieved due to the better selection of subsets of features and hyperparameters. Optimization has avoided redundant computation, and the Informer model tackles only the key points of data without sacrificing precision. Specifically, NSGA-II optimized attention mechanism parameters and the dimensions of the features, yielding a fast-forward pass with minimized redundant calculations and enhancing real-time feasibility.

The multi-objective optimization process needs more validation for its effectiveness according to the results shown in Fig. 3, which uses NSGA-II to generate Pareto fronts that demonstrate ADR ($\mathcal{F}_1$), FPR ($\mathcal{F}_2$), and inference latency ($\mathcal{F}_3$) trade-offs. Each point represents a non-dominated detection configuration that exists as a unique detection configuration using specific feature subsets, detection thresholds, and model hyperparameters. The NSGA-II algorithm identified optimal trade-off solutions, which the plots show, but they do not

display any progression over time.

Fig. 3(a) illustrates the trade-off between ADR and inference latency. Higher accuracy configurations require more time for inference due to the need for additional feature details or the larger capacity of their models. The ADR–FPR relationship appears in Fig. 3(b), which shows that improving accuracy results in building tighter decision boundaries that lead to reduced false alarm rates. The trade-off between FPR and inference latency appears in Fig. 3(c), which shows that reduced false alarm rates require extra processing resources.

The final deployed model is selected from the Pareto-optimal set by jointly balancing ADR, FPR, and inference latency. As reported in Table IV, the NSGA-II–optimized configuration improves ADR from 91.2% to 97.5% while simultaneously reducing inference latency from 6.5 ms to 4.3 ms per sample. It should be noted that the accuracy values reported in Table IV are computed on the validation subset during the NSGA-II optimization process for performance monitoring and comparability purposes only; accuracy is not included as an optimization objective and does not represent final test-set performance. Together, Fig. 3(a–c) and Table IV confirm that NSGA-II enables principled selection of deployment-ready FDIA detection models under competing performance and real-time constraints.

TABLE IV: Impact of NSGA-II–based optimization on the Informer detector.

Metric	Pre-Optimized Model	NSGA-II Optimized Model
Accuracy (%)	89.6	95.1
ADR (%)	91.2	97.5
FPR (%)	10.8	4.9
Inference Latency per Sample (ms)	6.5	4.3

E. Attack Vulnerability and Adversarial Defense Analysis

Validation of FDIA detection model resiliency incorporates testing for both attack vulnerability and the effectiveness of adversarial defense mechanisms. This section ties together the susceptibility of the model to adversarial attacks, the detection performance on different FDIA scenarios, and the effectiveness of defensive mechanisms. Table V shows the impact of such added adversarial attacks on detection accuracy.

As shown in Table V, iterative adversarial attacks induce substantially larger degradation in attack detection performance than single-step perturbations across all evaluated models. In particular, the momentum-based MIM attack results in the highest ADR degradation, with drops of 21.3% for the Autoencoder and 12.7% for the Informer-based detector. This gap of more than 8.5 percentage points highlights the increased vulnerability of reconstruction-based approaches to accumulated gradient perturbations, while indicating that the Informer architecture is more effective at preserving temporal consistency under strong adversarial manipulation.

Compared with other Transformer-based baselines, the proposed Informer model consistently exhibits the smallest ADR degradation under all attack types. For example, under the strongest MIM attack, Informer limits the ADR drop to 12.7%, whereas TimesNet and PatchTST experience larger degradations of 15.8% and 14.2%, respectively. A similar trend is observed under stealthy FDIA scenarios, where Informer incurs only a 5.9% ADR reduction compared to 9.6% for TimesNet and 8.7% for PatchTST. These results indicate that Informer achieves a more robust balance between expressive temporal modeling and resistance to adversarial perturbations.

In contrast, the Autoencoder baseline shows pronounced sensitivity to adversarial attacks across all scenarios. Even under the relatively mild FGSM attack, the Autoencoder suffers a 15.2% ADR degradation, nearly twice that of the Informer model (8.4%). This sensitivity becomes more severe for iterative attacks, confirming that reconstruction-error-based detection is less effective at distinguishing adversarially manipulated signals when perturbations are temporally structured.

TABLE V: ADR Degradation Under Adversarial Attacks

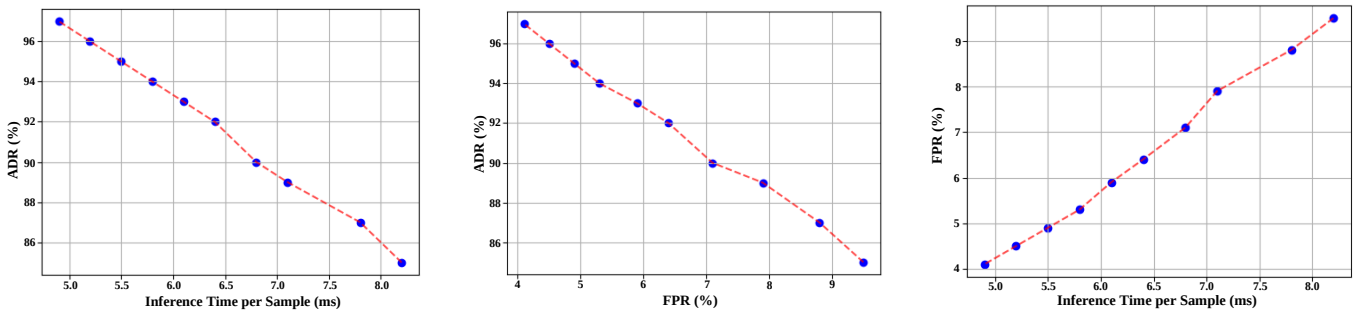
Attack Type	Autoencoder (% Drop)	TimesNet (% Drop)	PatchTST (% Drop)	Informer (% Drop)
FGSM Attack	15.2	10.6	9.8	8.4
BIM Attack	18.5	13.4	11.9	10.1
MIM Attack	21.3	15.8	14.2	12.7
Stealthy FDIA	14.7	9.6	8.7	5.9

TABLE VI: Cross-Attack Generalization Performance Under LOAO Evaluation (mean±std over three runs).

Train/Validation Attacks	Unseen Test Attack	Accuracy (%)	ADR (%)	FPR (%)
FGSM + BIM	MIM	91.8 ± 0.6	93.2 ± 0.5	6.4 ± 0.3
FGSM + MIM	BIM	92.4 ± 0.5	94.1 ± 0.4	5.9 ± 0.2
BIM + MIM	FGSM	93.0 ± 0.4	94.8 ± 0.3	5.6 ± 0.2

As shown in Table VI, the proposed Informer-based detector preserves strong detection capability even when evaluated on adversarial attack types that are completely unseen during training. Across all LOAO configurations, detection accuracy remains above 91.8% and ADR exceeds 93.2%, while the false positive rate stays below 6.4%. Compared with the random-split evaluation (Table III), the performance degradation is limited to less than 1.7% in accuracy and 2.0% in ADR, indicating that the model does not overfit to attack-specific patterns. These results confirm robust generalization to adaptive and previously unseen FDIA strategies.

To provide an intuitive understanding of how adversarial attacks manifest in power-system measurements, Fig. 4 illustrates a representative voltage magnitude time-series (p.u.) at a selected bus under normal and adversarial conditions. The signal is shown over a fixed temporal window of length $W = 50$ under the default sparse attack setting ($k = 3$). As observed, FGSM introduces relatively mild, single-step perturbations around the nominal trajectory, whereas BIM and MIM generate increasingly structured and stronger deviations due to their iterative update mechanisms. In particular, the



(a) ADR vs. inference Latency. (b) ADR vs. FPR. (c) FPR vs. inference latency.

Fig. 3: Pairwise Pareto fronts obtained using NSGA-II over the optimized objectives $\mathcal{F}_1$ (ADR), $\mathcal{F}_2$ (FPR), and $\mathcal{F}_3$ (inference latency). Each point represents a non-dominated detection configuration $\boldsymbol{\Theta}$.

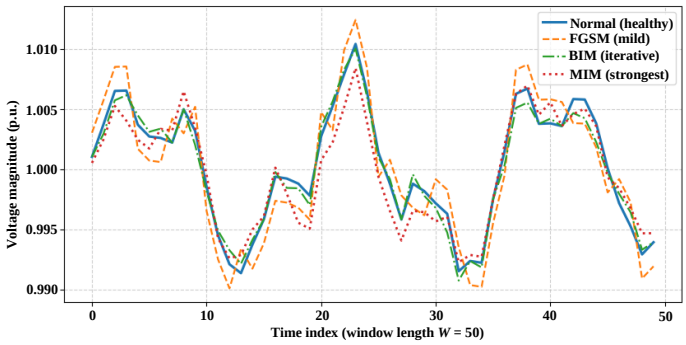


Fig. 4: Representative voltage magnitude time-series at a selected bus under normal and adversarial conditions.

MIM trace exhibits the largest divergence from the healthy signal, which is consistent with the higher detection performance degradation reported for MIM attacks in Table V and the ADR/FPR trends shown in Fig. 5.

To further analyze sensitivity to attack strength, we evaluated detection performance under increasing perturbation magnitudes $\epsilon \in \{0.005, 0.01, 0.02, 0.05\}$. Results indicate that the Informer model exhibits graceful degradation as attack intensity increases, maintaining high ADR and low FPR for $\epsilon \leq 0.02$, while performance degradation becomes more pronounced at $\epsilon = 0.05$. Across the same ϵ sweep, the baseline models (Autoencoder, TimesNet, PatchTST, and FEDformer) show stronger sensitivity to increasing perturbation magnitudes, with larger ADR reductions and/or higher FPR compared with the proposed Informer detector. These results confirm that the Informer architecture is less sensitive to moderate attack intensities and benefits more substantially from adversarial training and threshold fine-tuning.

To evaluate robustness against adversarially distorted FDIA samples, Fig. 5 presents the ADR and FPR under FGSM, BIM, MIM, and stealthy FDIA scenarios for all evaluated models. As shown in Fig. 5(a), the proposed Informer-based detector consistently achieves high detection performance, maintaining ADR values above 90% for FGSM, BIM, and MIM attacks. In contrast, the Autoencoder baseline exhibits substantially lower ADR, dropping below 80% under iterative attacks such as BIM and MIM, indicating higher vulnerability to adversarial perturbations.

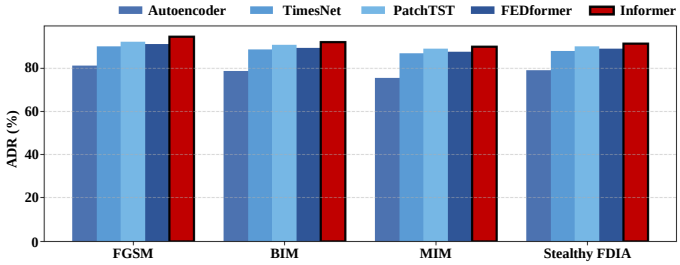
Compared with state-of-the-art Transformer-based baselines, Informer also demonstrates superior robustness. Under the strongest MIM attack, Informer maintains an ADR of approximately 90%, while TimesNet, PatchTST, and FEDformer achieve ADR values in the range of 87–89%. The same performance differences exist in stealthy FDIA tests, which show that Informer outperforms other solutions to detect adversarially manipulated time-related changes in electricity grid data.

The results in Fig. 5(b) demonstrate that Informer effectively achieves the lowest FPR across all tested adversarial conditions because its FPR values remain below 8% during MIM attacks. The Autoencoder model shows FPR increase which exceeds 14% during iterative attacks while Transformer-based systems demonstrate FPR values that range from approximately 6% to 10%. The results show that the Informer model delivers better protection against false alarms during adversarial attacks compared to existing systems.

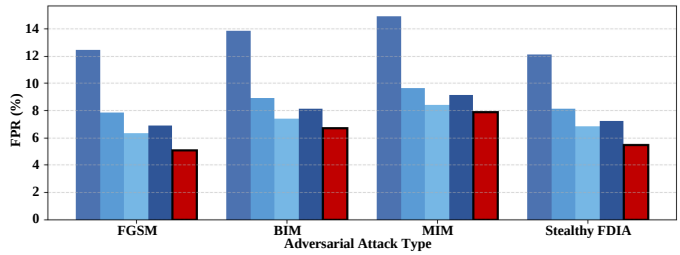
To further mitigate adversarial vulnerability, two complementary defense strategies were incorporated into the proposed framework:

- *Adversarial Training:* The detection model is retrained using a mixture of clean and adversarially perturbed samples to improve robustness against gradient-based attacks.
- *Threshold Fine-Tuning:* The system uses adaptive threshold adjustment to minimize false positive results while maintaining high levels of attack detection success during adversarial testing.

Table VII shows that adversarial training we examined produces significant improvements in ADR which demonstrates reduced failures to detect FDIA events during adversarial testing while achieving a 7.4% increase in classification accuracy. Threshold fine-tuning improves operational reliability by reducing false alarms and the FPR rate by 3.7% during regular grid operations. Together with the adversarial degradation analysis in Table V and the cross-attack generalization results in Table VI, the results show that the Informer-based detector performs better than other benchmark techniques across a range of adversarial testing scenarios. The combination of inherent architectural strength and dedicated security systems



(a) ADR under adversarial attacks.



(b) FPR under adversarial attacks.

Fig. 5: Performance under adversarial attacks for the Autoencoder baseline, state-of-the-art Transformer-based models (TimesNet, PatchTST, FEDformer), and the proposed Informer-based FDIA detector.

creates a protective barrier that protects against performance damage from adversarial threats, demonstrating that integrated adversarial security measures are essential for protecting FDIA detection systems used in smart grid operations.

TABLE VII: Impact of Defensive Strategies on Model Performance

Defense Strategy	Accuracy Improvement (%)	ADR Improvement (%)	FPR Reduction (%)
Adversarial Training	+7.4	+6.8	-4.2
Threshold Fine-Tuning	+1.9	+4.1	-3.7

F. Scalability and Real-Time Feasibility

It is important to clarify that the scalability and real-time feasibility analysis in this section focuses exclusively on the online inference and detection stage. As described in Section II-A, all FDIA samples are generated offline to ensure reproducibility and precise control of attack parameters. Here, we evaluate whether the trained detection models can operate under real-time streaming conditions by measuring inference latency, throughput, and memory usage during deployment. Ensuring the FDIA detection system is able to work well with varying grid sizes and real-time requirements is crucial to its actual deployment. In this section, we compare the scalability of our Informer-based proposed model with the Autoencoder-based baseline model in computational complexity, inference latency, and memory usage. To make a scalability comparison, we evaluated model performance on the following three IEEE benchmark networks:

- IEEE 5-bus system (Small-scale grid)
- IEEE 14-bus system (Medium-scale grid)
- IEEE 118-bus system (Large-scale grid)

The computational complexity of the evaluated models was measured in terms of training time, inference latency, and memory footprint across different grid topologies. Table VIII presents a comparative scalability analysis for the Autoencoder baseline, state-of-the-art Transformer-based models, and the proposed Informer-based detector. Table VIII presents a comparative scalability analysis for the Autoencoder baseline, state-of-the-art Transformer-based models, and the proposed Informer-based detector. In this scalability study, inference latency corresponds to the model forward-pass latency per sample measured with batch size $= 1$ under the same hardware configuration, rather than the end-to-end detection latency reported in Table III. As shown in Table VIII, the Informer model consistently demonstrates superior scalability characteristics as the grid size increases.

In terms of training efficiency, Informer requires substantially less training time across all grid sizes. For the IEEE 118-bus system, Informer completes training in 145.2 s, compared with 188.7 s for the Autoencoder and over 200 s for TimesNet, PatchTST, and FEDformer, indicating lower optimization and parameter-update overhead.

The analysis of inference latency demonstrates that Informer provides superior performance for real-time applications. The large-scale IEEE 118-bus system shows that Informer achieves a sample inference latency of 5.9 ms, which remains under the 6 ms threshold. In contrast, TimesNet, PatchTST, and FEDformer show higher latency periods, which range between 8.1 and 8.5 ms while the Autoencoder reaches 9.2 ms. This confirms that Informer maintains real-time feasibility under increasing grid complexity.

Memory footprint comparisons show that Informer is also more resource-efficient. For the IEEE 118-bus network, Informer consumes 980 MB of memory, compared with 1200 MB for the Autoencoder and up to 1340 MB for PatchTST, demonstrating a clear advantage for deployment in resource-constrained environments such as substations or edge devices.

This improved scalability is primarily attributed to Informer's ProbSparse self-attention mechanism, which selectively attends to the most informative long-range temporal dependencies rather than processing the full attention matrix exhaustively. By contrast, both reconstruction-based Autoencoders and full-attention Transformer architectures must process or reconstruct the entire input space, resulting in increased computational overhead, memory consumption, and inference latency as grid size grows.

For further real-time viability testing of the online detection stage, we subjected the trained models to different data stream processing rates, simulating real-life PMU streaming conditions. Fig. 6 illustrates the throughput (samples per second) that the models can sustain without encountering latency violations. As shown in Fig. 6, the proposed Informer-based detector achieves the highest throughput, processing up to 2100 samples per second, whereas the Autoencoder-based model sustains approximately 1450 samples per second before

TABLE VIII: Scalability analysis across IEEE 5-bus, 14-bus, and 118-bus systems. Inference latency corresponds to model forward-pass latency per sample, measured with batch size = 1 under the same hardware configuration.

Metric (Model)	IEEE 5-Bus	IEEE 14-Bus	IEEE 118-Bus
Training Time (Autoencoder)	18.5s	49.3s	188.7s
Training Time (TimesNet)	21.4s	56.8s	201.3s
Training Time (PatchTST)	24.9s	61.5s	214.6s
Training Time (FEDformer)	23.1s	58.7s	209.8s
Training Time (Informer)	**12.3s**	**35.7s**	**145.2s**
Inference Latency per Sample (Autoencoder)	5.6ms	6.9ms	9.2ms
Inference Latency per Sample (TimesNet)	4.9ms	6.1ms	8.5ms
Inference Latency per Sample (PatchTST)	4.6ms	5.8ms	8.1ms
Inference Latency per Sample (FEDformer)	4.8ms	6.0ms	8.3ms
Inference Latency per Sample (Informer)	**3.8ms**	**4.5ms**	**5.9ms**
Memory Usage (Autoencoder)	280MB	560MB	1200MB
Memory Usage (TimesNet)	310MB	590MB	1280MB
Memory Usage (PatchTST)	340MB	620MB	1340MB
Memory Usage (FEDformer)	325MB	605MB	1310MB
Memory Usage (Informer)	**230MB**	**450MB**	**980MB**

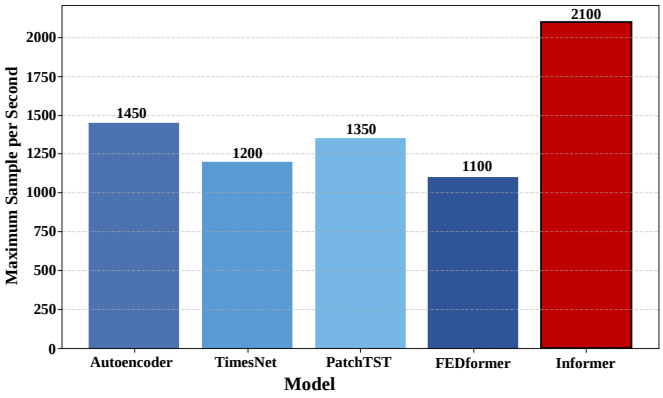


Fig. 6: Throughput analysis showing the maximum number of samples per second processed during online inference for Autoencoder-based, state-of-the-art Transformer-based, and the proposed Informer-based FDIA detection models.

experiencing delays. This indicates that the Informer model is more suitable for high-frequency smart grid monitoring while maintaining detection accuracy and real-time responsiveness.

Throughput measurements were obtained under streaming inference with mini-batched processing, whereas the latency results in Fig. 2 report per-sample inference latency, explaining the apparent difference between latency and sustained throughput values.

In comparison, state-of-the-art Transformer-based baselines exhibit intermediate throughput levels due to their increased architectural complexity. Specifically, TimesNet and PatchTST sustain throughput in the range of approximately 1200–1350 samples per second, while FEDformer achieves a lower throughput of around 1100 samples per second, primarily due to its frequency-domain processing overhead. These throughput trends are consistent with the inference latency results reported in Fig. 2, confirming that the ProbSparse attention mechanism enables Informer to achieve a more favorable balance between detection performance and real-time processing capability.

Beyond algorithmic performance, several system-level factors influence the practical deployment of FDIA detection models in operational power grids.

- *SCADA/PMU data ingestion and preprocessing:* In real-world environments, PMU measurements are streamed at rates of 30–60 frames per second. The reported inference latency and throughput results account only for the online detection stage; lightweight preprocessing operations (normalization, windowing) introduce negligible additional delay relative to model inference.
- *Model update and concept drift handling:* Grid operating conditions evolve over time due to load variation, topology changes, and seasonal effects. The proposed framework supports periodic offline retraining or incremental fine-tuning using newly collected data, while preserving low-latency inference during online operation.
- *Edge deployment pathway:* For edge-level deployment at substations or control centers, model compression techniques such as weight quantization, structured pruning, and knowledge distillation can be applied to further reduce memory footprint and inference latency. Batch-based inference can also be leveraged when processing synchronized PMU streams.
- *Cybersecurity operations and fail-safe integration:* The detection module is designed to operate as a monitoring layer integrated with existing EMS/SCADA infrastructures. Secure communication channels, integrity checks on measurement streams, and fail-safe fallback mechanisms ensure that detection alerts do not compromise grid stability under uncertain conditions.

G. Feature Selection and Hyperparameter Optimization Evaluation

Performance of detection against feature selection is compared in Table IX where models trained with all features available are compared to models trained on optimized features. According to this table, The optimized feature selection increased the ADR by 12.5% and decreased the FPR by 9%. The reductions validate the requirement for the elimination of redundant features to increase detection accuracy as well as to improve computational effectiveness.

The model's dramatic increase in ADR ($+12.5\%$) and reduction in FPR ($-$ 9.0%) after feature selection is best understood by attributing it to two main reasons. First, with the removal of redundant and noisy features, the detection model does not fit irrelevant variation, thus allowing it to generalize perfectly to new data. Secondly, using MI and PCA, it prioritized the most informative features, such as Voltage Magnitude (V_i) and Phase Angle (θ_i), most contributing to FDIA anomaly separation. The optimal feature set for this feature subset produces a robust decision boundary and, therefore, improves detection performance.

TABLE IX: Impact of Feature Selection on Detection Performance

Metric	Before Feature Selection	After Feature Selection
ADR (%)	82.7	95.2 (+12.5)
FPR (%)	14.1	5.1 (-9.0)
Inference Latency (ms/sample)	6.5	4.3 (-2.2)

Furthermore, Bayesian optimization was employed to efficiently tune continuous hyperparameters jointly with feature selection. Fig. 7 illustrates the convergence behavior of the Bayesian search by reporting the best validation accuracy achieved as a function of optimization trials. The optimization process achieves its best results through initial trials that match its target performance, followed by a stable performance period because the algorithm successfully explores all available hyperparameter settings. The system demonstrates its capability to complete training processes 22% faster, as observed in Table X, because it identifies optimal configuration through its rapid convergence method, which eliminates unnecessary testing. In addition, the stabilized best validation accuracy achieved through Bayesian optimization leads to improved detection performance and reduced false positives, as summarized in Table X.

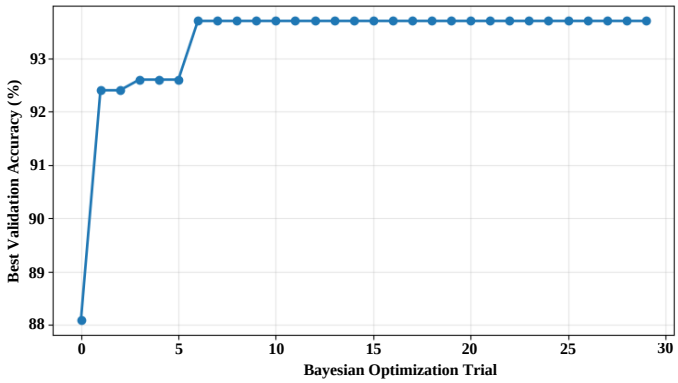


Fig. 7: Bayesian optimization convergence showing the best validation accuracy achieved as a function of optimization trials during hyperparameter tuning.

TABLE X: Hyperparameter Tuning Impact on Model Performance

Metric	Default Hyperparameters	Optimized Hyperparameters
Accuracy (%)	85.2	93.0
Training Convergence Time (epochs)	50	39

IV. Limitations and Future Direction

The research has shown performance improvements yet there remain multiple constraints that create opportunities for future studies. The combination of adversarial training with threshold fine-tuning provides a strong defense against FGSM, BIM, and MIM attacks, yet advanced adaptive adversaries who develop their skills through time can still overcome both detection systems. The upcoming research needs to focus on developing defense systems that use reinforcement learning and online adaptive methods for real-time tracking of new attack patterns.

While the Informer-based detector outperforms the Autoencoder and other Transformer-based benchmarks in large-scale inference efficiency, achieving per-sample latency below 6 ms on the IEEE 118-bus system, practical deployment under stringent real-time constraints with high-frequency streaming PMU data may remain challenging. In particular, large-scale grids with dense sensor deployments and strict end-to-end latency budgets could impose additional computational pressure. To address this, future research should explore lightweight Informer variants, model compression techniques, or edge-computing–assisted deployment strategies to further improve real-time feasibility.

Finally, although this work focuses on FDIA detection and robustness enhancement, effective cybersecurity operation in smart grids ultimately requires tighter integration with real-time energy management systems (EMS) and operational decision-making frameworks. The process of turning detection results into operational control measures and mitigation plans continues to present an ongoing challenge. The process of turning detection results into operational control measures and mitigation plans continues to present an ongoing challenge. The process of turning detection results into operational control measures and mitigation plans continues to present an ongoing challenge. Future research needs to study human-in-the-loop AI systems together with operator-based decision support systems because these technologies will help organizations deliver secure and quick cyber attack responses in actual grid operations [36].

V. Conclusion

The research developed a multi-objective optimization system that combines Informer-based FDIA detection with adversarial defense systems to improve security and operational capacity of contemporary smart grid systems. The framework solves fundamental problems of existing FDIA detection systems through its joint optimization of ADR, FPR, and inference latency. The elimination of redundant measurements through feature selection, which removed non-informative data, led to improved detection results with an ADR gain of 12.5 percentage points and a 9 percentage point decrease in FPR. The process of Bayesian hyperparameter optimization boosted model performance through a 7.8% increase in classification accuracy and a 22% decrease in training duration, which resulted in a better-performing and more reliable detection system. The Informer-based detector showed superior performance compared to Autoencoder, TimesNet, PatchTST, and FEDformer through all evaluated metrics, which included accuracy, ADR, FPR, and inference latency. The ProbSparse self-attention mechanism of the Informer enables efficient

long-distance time and cross-channel dependency modeling with minimal processing requirements, which results in better defense against stealthy attacks and adversarial FDIAs created through FGSM, BIM, and MIM methods. The proposed framework achieved operational efficiency through its processing capacity, which handled 2100 samples per seconds while its processing time per sample remained under 6 ms. This capability enables real-time operations on IEEE 5-bus, 14-bus, and 118-bus systems.

The existing detection frameworks face two major challenges because they need to handle larger grid environments that operate at higher frequencies, and they must defend against attackers who adopt new strategies. Research will explore lightweight Informer variants through their implementation in edge-computing systems and their use in online defense systems, which apply reinforcement learning to achieve better real-time results and enhanced system performance during prolonged periods. The system requires better integration between AI-based FDIA detection systems and EMS, together with operational decision-making processes, to convert detection results into precise grid response actions.

REFERENCES

[1] Abbas , K. Cybersecurity Challenges in Smart Grids: A Focus on Information Technology. *J. Electrical Systems* **2024**, *20(6s)*, 1394–1407.

[2] Andronikidis , G.; Eleftheriadis , C.; Batzos , Z.; Kyranou , K.; Maropoulos , N.; Sargsyan , G.; Grammatikis , PR.; Sarigiannidis , P. AI-Driven Anomaly and Intrusion Detection in Energy Systems: Current Trends and Future Direction. *IEEE International Conference on Cyber Security and Resilience (CSR)* **2024**, 777–782.

[3] Tooki , O.; Popoola , O. Intelligent-Based Techniques for Detection, Identification and Mitigation of Cyberattacks in Cyber-Physical Power Systems: A Review. *IEEE PES/IAS PowerAfrica* **2024**, 01–5.

[4] Tightiz , L.; Yang , H. A comprehensive review on IoT protocols' features in smart grid communication. *Energies* **2020**, *13(11)*, 2762.

[5] Ahmed , A.; Nadeem , MF.; Kiani , AT.; Ullah , N.; Khan , MA.; Mosavi , A. An improved hybrid approach for the simultaneous allocation of distributed generators and time varying loads in distribution systems. *Energy Reports* **2023**, *9*, 1549–60.

[6] Tightiz , L.; Dang , LM.; Padmanaban , S.; Hur K. Metaverse-driven smart grid architecture. *Energy Reports* **2024**, *12*, 2014–25.

[7] Inayat , U.; Zia MF.; Mahmood , S.; Berghout , T.; Benbouzid , M. Cybersecurity enhancement of smart grid: Attacks, methods, and prospects. *Electronics* **2022**, *11(23)*, 3854.

[8] Ghiasi , M.; Niknam , T.; Wang , Z.; Mehrandezh , M.; Dehghani M.; Ghadimi , N. A comprehensive review of cyber-attacks and defense mechanisms for improving security in smart grid energy systems: Past, present and future. *Electric Power Systems Research* **2023**, *215*, 108975.

[9] Alomari , MA.; Al-Andoli , MN.; Ghaleb , M.; Thabit , R.; Alkawsi , G.; Alsayaydeh , JA.; Gaid , AS. Security of Smart Grid: Cybersecurity Issues, Potential Cyberattacks, Major Incidents, and Future Directions. *Energies* **2025**, *18(1)*, 141.

[10] Khan , AA.; Chaudhari , O.; Chandra , R. A review of ensemble learning and data augmentation models for class imbalanced problems: combination, implementation and evaluation. *Expert Systems with Applications* **2024**, *244*, 122778.

[11] Abdullahi , M.; Alhussian , H.; Aziz , N.; Abdulkadir , SJ.; Alwadain , A.; Muazu , AA.; Bala , A. Comparison and investigation of AI-based approaches for cyberattack detection in cyber-physical systems. *IEEE Access* **2024**, *12* , 9038-9046.

[12] Musleh , AS.; Chen , G.; Dong , ZY. A survey on the detection algorithms for false data injection attacks in smart grids. *IEEE Transactions on Smart Grid* **2019**, *11(3)*, 2218–34.

[13] Irfan , M.; Sadighian , A.; Tanveer , A; Al-Naimi , SJ.; Oligeri , G. A survey on detection and localisation of false data injection attacks in smart grids. *IET Cyber-Physical Systems: Theory & Applications* **2024**, *9(4)*, 313–333.

[14] Tightiz , L.; Nasimov , R.; Nasab , MA. Implementing AI Solutions for Advanced Cyber-Attack Detection in Smart Grid. *International Journal of Energy Research* **2024**, *2024(1)*, 6969383.

[15] Mukherjee , D. Detection of data-driven blind cyber-attacks on smart grid: A deep learning approach. *Sustainable Cities and Society* **2023**, *92*, 104475.

[16] Zhang , Y.; Wang , J.; Chen , B. Detecting false data injection attacks in smart grids: A semi-supervised deep learning approach. *IEEE Transactions on Smart Grid* **2020**, *12(1)*, 623–34.

[17] Hao , J.; Tao , Y. Adversarial attacks on deep learning models in smart grids. *Energy Reports* **2022**, *8*, 123–9.

[18] Ness , S. Adversarial Attack Detection in Smart Grids Using Deep Learning Architectures *IEEE Access* **2025**, *13*, 16314–16323.

[19] Sampedro , GA.; Ojo , S.; Krichen , M.; Alamro , MA.; Mihoub , A.; Karovic , V. Defending AI Models Against Adversarial Attacks in Smart Grids Using Deep Learning. *IEEE Access* **2024**, *12*, 157408–157417.

[20] Li , Y.; Wei X.; Li Y.; Dong , Z.; Shahidehpour , M. Detection of false data injection attacks in smart grid: A secure federated deep learning approach. *IEEE Transactions on Smart Grid* **2022**, *13(6)*, 4862–72.

[21] Keçeci , C.; Davis , K. R.; Serpedin , E. Federated Learning-Based Distributed Localization of False Data Injection Attacks on Smart Grids, *IEEE Systems Journal* **2025**, *19(3)*,, 719–729.

[22] Sun , Y.; Hou , L.; Lv , Z.; Peng , D. Informer-based intrusion detection method for network attack of integrated energy system. *IEEE Journal of Radio Frequency Identification*, **2022**, *6*, 748–52.

[23] Tightiz , L.; Dang , L. Minh; Park , K. -W.; AIoT-Blockchain Security for Supply Chain Threats in IEC 61850 Substations Using Informer-Powered Reinforcement Learning. *IEEE Internet of Things Journal* **2026**, *13(2)*, 3138–3155.

[24] Wang , J.;Nie , L.; Gu , Z; Wang , J., Tan , R.; Kumari , S. Real-Time GPS Spoofing Detection in Consumer Drones Through Multi-Sensor Data Fusion and TimesNet, *IEEE Transactions on Consumer Electronics* **2025**, *71(2)*, 5569–5583.

[25] He , J. ; Yuan , K.; Zhong , Z.; Sun , Y. Enhancing Short-Term Power Load Forecasting With a TimesNet-Crossformer-LSTM Approach, *IEEE Access* **2024**,*12*, 56774–56788.

[26] Lu , K.; Gao , S.; Li , J.; Chen , K.; Yu ,C. Prediction of Power System Ramping Demand Using Meteorological Features, *IEEE Access* **2025**, *13*, 85775–85788.

[27] Lu , KD.; Wu , ZG. Multi-objective false data injection attacks of cyber–physical power systems. *IEEE Transactions on Circuits and Systems II: Express Briefs* **2022**, *69(9)*, 3924–8.

[28] Rahman , M.; Li Y.; Yan , J. Multi-objective evolutionary optimization for worst-case analysis of false data injection attacks in the smart grid. *IEEE Congress on Evolutionary Computation (CEC)* **2020**, 1–8.

[29] Lin , X.; An , D.; Cui , F.; Zhang , F. False data injection attack in smart grid: Attack model and reinforcement learning-based detection method. *Frontiers in Energy Research* **2023**, *10*, 1104989.

[30] Sun , Y.; Zhuang , L.; Jia , T.; Cheng , D.; Zhao , X.; Guo , J. A risk assessment method for power internet of things information security based on multi-objective hierarchical optimisation. *IET Smart Grid* **2025**.

[31] Jeon , S.; Seo , JT. A Synthetic Time-Series Generation Using a Variational Recurrent Autoencoder with an Attention Mechanism in an Industrial Control System. *Sensors* **2023**,*24(1)*, 128.
An , D.; Zhang , F., Yang , Q.; Zhang , C. Data integrity attack in dynamic state estimation of smart grid: Attack model and countermeasures. *IEEE Transactions on Automation Science and Engineering* **2022**, *19(3)*, 1631–44.

[32] Le , TD.; Anwar , A.; Loke , SW.; Beuran , R.; Tan , Y. Gridattacksim: A cyber attack simulation framework for smart grids. *Electronics* **2020**, *9(8)*, 1218.

[33] Wu , H.; Hu , T.; Liu , Y.; Zhou , H.; Wang , J.; Long , M. Timesnet: Temporal 2d-variation modeling for general time series analysis. *arXiv preprint arXiv:2210.02186* **2022**.

[34] Nie , Y. A Time Series is Worth 64Words: Long-term Forecasting with Transformers. *arXiv preprint arXiv:2211.14730* **2022**.

[35] Zhou , T.; Ma , Z.; Wen , Q.; Wang , X.; Sun , L.; Jin , R. Fedformer: Frequency enhanced decomposed transformer for long-term series forecasting. *InInternational conference on machine learning*, PMLR, **2022 Jun 28**, 27268-27286.

[36] Bayasgalan , A.; Park , Y.S.; Koh , S.B.; Son , S.Y. Comprehensive Review of Building Energy Management Models: Grid-Interactive Efficient Building Perspective. *Energies* **2024**, *17(19)*, 4794.