Journal of Computer Science and Technology Studies
ISSN: 2709-104X
DOI: 10.32996/jcsts
Journal Homepage: www.al-kindipublisher.com/index.php/jcsts

JCSTS
AL-KINDI CENTER FOR RESEARCH AND DEVELOPMENT

| RESEARCH ARTICLE

AI in Smart Grid Cybersecurity: A Systematic Review of Machine Learning and Deep Learning Approaches against False Data Injection and Other Emerging Attacks

Touhid Bhuiyan
School of Information Technology, Washington University of Science and Technology
Corresponding Author: Touhid Bhuiyan, **E-mail**: touhid.bhuiyan@wust.edu

| ABSTRACT

Abstract

The increasing digitalization of smart grids has heightened their vulnerability to sophisticated cyber threats, with false data injection (FDI) and other emerging attacks posing significant risks to grid stability, reliability, and resilience. Artificial intelligence, particularly machine learning (ML) and deep learning (DL), has gained prominence as a promising defense layer capable of detecting, mitigating, and adapting to these dynamic threats. However, the rapid growth of research in this area has produced fragmented findings across diverse methodologies, datasets, and evaluation strategies. To address this gap, our systematic review consolidates the current state of ML- and DL-driven approaches in smart grid cybersecurity, with a specific emphasis on FDI detection and defense against evolving adversarial tactics. We map the landscape of proposed techniques, highlight benchmark datasets and simulation environments, and critically examine strengths, limitations, and open challenges. In doing so, we establish a taxonomy of AI-based solutions that organizes existing efforts by learning paradigm, attack type, and deployment layer within the smart grid. Beyond cataloguing current achievements, we underscore persistent challenges such as scalability, data imbalance, adversarial robustness, and model explainability, all of which constrain real-world deployment. By synthesizing insights from both academic research and industrial practice, this review aims to provide a roadmap for researchers, practitioners, and policymakers seeking to develop resilient, trustworthy, and adaptive AI-driven cybersecurity mechanisms for future power systems.

| KEYWORDS

smart grid; cybersecurity; false data injection; machine learning; deep learning; adversarial attacks; resilience

| ARTICLE INFORMATION

ACCEPTED: 04 July 2025 **PUBLISHED:** 25 August 2025 **DOI:** 10.32996/jcsts.2025.7.8.136

1. INTRODUCTION

The evolution of power systems into intelligent and interconnected smart grids has transformed the traditional electricity infrastructure into a cyber-physical system. This transformation, while enabling efficiency, automation, and resilience, has also introduced a broader attack surface for malicious actors. Smart grids integrate advanced information and communication technologies (ICT), Internet of Things (IoT) devices, distributed energy resources, and real-time data analytics, all of which are critical to ensuring stability and reliability in modern energy systems [1]. However, these interdependencies have simultaneously increased the susceptibility of power systems to sophisticated cyber threats such as false data injection (FDI), denial of service (DoS), replay attacks, and malware propagation [2]. Among these, FDI attacks have attracted particular concern due to their ability to stealthily manipulate state estimation processes, leading to cascading failures, economic losses, and potential blackouts [3]. The convergence of cybersecurity and artificial intelligence (AI) has emerged as a promising frontier to address these threats. AI-driven solutions, particularly machine learning (ML) and deep learning (DL), provide adaptive, data-driven approaches that can detect and mitigate previously unseen attacks while learning from dynamic system behavior [4]. Unlike traditional rule-based intrusion detection systems, AI-based methods can capture complex nonlinear patterns, improve anomaly detection accuracy,

and adapt to evolving adversarial strategies. These advantages have accelerated the adoption of ML and DL in diverse domains of smart grid security, including FDI detection, load forecasting under attack, intrusion detection systems, and adversarial resilience modeling [5], [6].

The domain of AI-enabled smart grid cybersecurity spans multiple layers of the energy ecosystem, from the physical power transmission infrastructure to control centers and end-user interfaces. Applications include real-time monitoring of supervisory control and data acquisition (SCADA) systems, anomaly detection in phasor measurement units (PMUs), protection of demand response programs, and securing vehicle-to-grid communications [7], [8]. Furthermore, AI methods leverage a variety of data modalities such as network traffic, sensor measurements, historical operational data, and simulation environments to create robust detection frameworks. These techniques bring several benefits, including early warning of attacks, improved situational awareness, and faster incident response, thereby contributing to grid resilience and operational continuity. Despite their potential, AI-driven cybersecurity approaches face critical limitations. On the positive side, ML and DL models offer superior scalability, adaptability, and detection accuracy compared to static approaches. They also enable predictive security, where threats can be anticipated rather than simply detected after occurrence. On the negative side, these methods often demand large, high-quality datasets for training, which are scarce in the power systems domain due to privacy concerns and the rarity of labeled cyberattack data. Moreover, issues such as adversarial machine learning, explainability, high computational cost, and real-time deployment constraints raise concerns about their practical feasibility [9], [10]. The trade-offs between model complexity, interpretability, and deployment readiness remain unresolved challenges for both academia and industry.

AI in smart grid cybersecurity is thus a broad field that encompasses diverse techniques and deployment strategies, all aimed at protecting the grid from cyber-physical disruptions. There are several intersections between AI-based security mechanisms and related fields, such as blockchain for secure data sharing, edge computing for real-time inference, and digital twins for simulating attack-defense dynamics. These interconnections illustrate the multidisciplinary nature of smart grid security research, highlighting the need for holistic approaches that go beyond algorithmic accuracy alone. Cybersecurity for smart grids differs from traditional IT cybersecurity in significant ways. Traditional IT systems emphasize data confidentiality and integrity in isolated networks, whereas smart grid security prioritizes availability, real-time reliability, and resilience against cascading physical effects. In IT systems, breaches often lead to data leaks or financial losses, while in power systems, they can directly disrupt electricity supply, compromise public safety, and damage critical infrastructure. Furthermore, traditional IT systems rely on abundant labeled datasets and redundant infrastructures, whereas smart grids operate under stringent latency, resource, and operational constraints [11], [12]. This context underscores the necessity of AI-enabled solutions tailored specifically for smart grids rather than generic cybersecurity frameworks.

We have undertaken the initiative to systematically explore and consolidate the area of AI-driven smart grid cybersecurity. This review occupies a significant position within the current state of the art by offering a comprehensive analysis, taxonomy development, and synthesis of ML- and DL-based approaches against FDI and other emerging attacks. Through critical evaluation of existing literature, datasets, algorithms, and implementation strategies, this study provides valuable insights into the practical challenges and research frontiers. Key contributions of this review include the following:

- This study identifies the domain of AI-enabled smart grid cybersecurity and categorizes the research landscape across multiple attack types and system layers.
- Various ML and DL techniques are explored, along with the datasets, benchmarks, and simulation environments employed in this field.
- A thorough comparative analysis of research works is presented, summarizing key contributions, methodologies, and observed limitations.
- The review highlights state-of-the-art challenges such as adversarial robustness, explainability, scalability, and data imbalance, and discusses how these issues constrain deployment.
- This study provides future research directions and recommendations that can guide both academic and industrial efforts toward developing resilient and trustworthy smart grid cybersecurity solutions.

The outcomes of this work offer practical implications for industry professionals, policymakers, and researchers engaged in power systems security. By establishing a structured taxonomy and synthesizing key advancements, this review facilitates a clearer understanding of the AI-enabled defense landscape. It also lays the foundation for future studies aimed at addressing unresolved challenges and accelerating real-world deployment. The structure of this paper is organized as follows: Section I presents the introduction and motivation for the review. Section II provides a discussion of related studies and identifies research gaps. Section III describes the review methodology, including study selection and research questions. Section IV analyzes the

distribution of selected works, while Section V presents the taxonomy of AI-driven smart grid cybersecurity. Section VI discusses findings, open challenges, and future directions. Finally, Section VII concludes the paper.

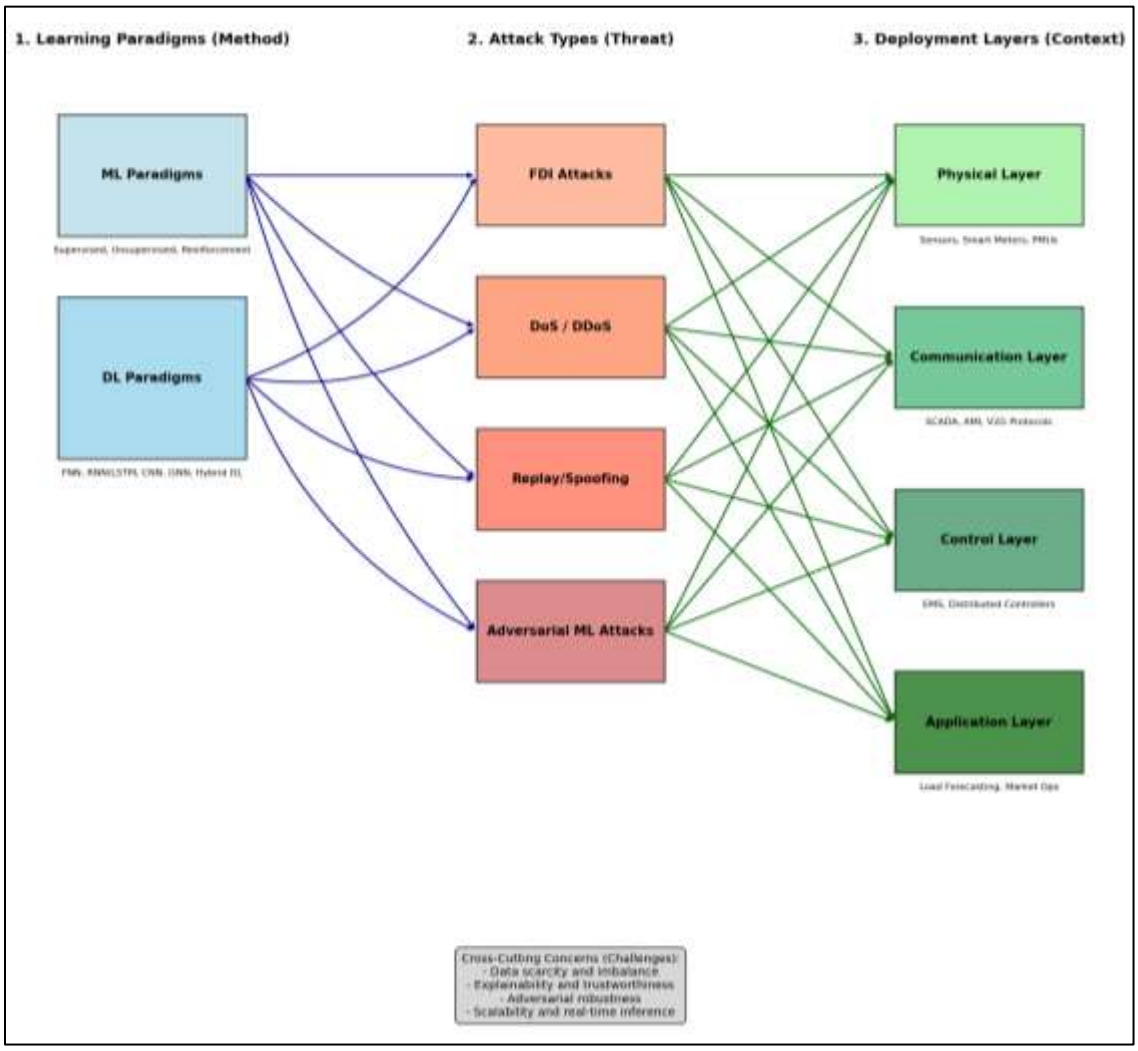

Figure 1. Taxonomy of AI-Driven Smart Grid Cybersecurity Approaches

2. RELATED STUDIES

In recent years, a substantial body of literature has emerged at the intersection of artificial intelligence and smart grid cybersecurity. Several broad surveys and systematic reviews have attempted to map this rapidly growing area, with emphasis on the detection of false data injection (FDI) and other cyber-physical attacks. For example, Zhang et al. et al. [13] provided a comprehensive survey of ML and DL techniques applied to smart grid security, cataloguing methods by attack type and reporting performance trends across benchmark testbeds. Similarly, Rao et al. et al. [14] synthesized advances in intelligent intrusion detection for energy systems, highlighting the proliferation of supervised classifiers and anomaly detectors while noting issues in reproducibility and dataset availability. These studies frame the current research landscape and expose recurring methodological limitations that motivate our systematic review. Focused investigations into false data injection attacks form a large and influential sub-literature. Several empirical and theoretical works have characterized FDI threats and proposed detection strategies that leverage statistical learning, sparse recovery, and graph-based models. Kumar et al. et al. [15] evaluated supervised detectors that rely on residual analysis in state estimation, demonstrating high detection rates under certain contamination regimes but also exposing sensitivity to adversary knowledge. More recent efforts by Li et al. et al. [16] combined topology-aware features with ensemble learning to improve robustness against stealthy FDI crafted using network constraints. Nevertheless, these studies typically assume controlled experiments on IEEE test systems and rarely validate methods on operational-scale traces, leaving open questions about transferability to live deployments. An additional strand by Park et al. et

al. [17] examined hybrid rule-learning pipelines that fuse physics-based invariants and ML scoring, showing promise but encountering scalability bottlenecks when applied to larger networks.

A sizable literature addresses anomaly detection and intrusion detection using traditional machine learning models. Support vector machines, random forests, and isolation forests have been widely explored for flagging abnormal meter readings, network traffic anomalies, and telemetry drift. Ahmad et al. et al. [18] compared several classical learners across synthetic FDI scenarios, reporting that ensemble tree models often outperform linear classifiers on unbalanced datasets. Conversely, Santos et al. et al. [19] highlighted the limits of supervised approaches when labeled attack samples are scarce, advocating semi-supervised and one-class frameworks. Complementary work by Oliveira et al. et al. [20] evaluated unsupervised clustering and statistical process control techniques, finding reasonable false positive control but reduced sensitivity to low-magnitude, coordinated manipulations. Collectively, these studies show progress in baseline detection but leave unresolved the challenge of achieving high sensitivity while maintaining operational false alarm rates. Deep learning has attracted intense interest for its capacity to model complex spatiotemporal patterns in grid data. Architectures such as recurrent neural networks, convolutional models, autoencoders, and graph neural networks have been proposed to capture temporal dependencies, spatial topology, and multivariate correlations. Huang et al. et al. [21] demonstrated LSTM-based detectors for PMU streams that identify transient anomalies faster than sliding-window statistics. Chen et al. et al. [22] applied convolutional autoencoders for feature learning from synchronized measurement matrices, reporting improved detection in noisy environments. Notably, GNN-based approaches by Moreno et al. et al. [23] exploit grid topology explicitly, improving detection of topology-aware FDI that targets correlated buses. Despite strong in-sample performance, most DL studies emphasize accuracy metrics and often omit deployment considerations such as inference latency and model maintenance.

Protection of state estimation and PMU integrity has been treated as a specific technical problem combining power-systems engineering and AI methods. Several studies propose augmented estimators, measurement authentication, and detection layers that monitor residuals and learned invariants. Singh et al. et al. [24] presented a layered detection framework that integrates residual checks with a supervised classifier trained on synthetic FDI scenarios. In parallel, Alvarenga et al. et al. [25] explored secure PMU placement and the use of redundancy-aware learning to reduce attacker stealth. These contributions advance detection capability but frequently rely on idealized PMU coverage and assume an attacker model with limited adaptivity, which underestimates adversaries who can adapt to deployed defenses. Benchmark datasets, testbeds, and simulators underpin progress but also constrain it. Studies often reuse IEEE bus systems, MATPOWER, and PowerWorld case studies, and custom simulation pipelines to generate attack traces. Park et al. et al. [26] surveyed common datasets and found heavy reliance on small synthetic systems such as the IEEE 14-bus and 118-bus cases. Wang et al. et al. [27] developed a larger curated dataset that includes correlated network traffic and meter streams, yet the dataset's scope remains limited relative to real utility heterogeneity. Efforts to create realistic testbeds, including hardware-in-the-loop and real-time digital simulators, have been advanced by Garcia et al. et al. [28], but accessibility and reproducibility remain obstacles for independent validation. The scarcity of publicly available, labeled, and diverse datasets is therefore a persistent barrier to generalizable research.

Adversarial machine learning has emerged as a critical concern as attackers target learning systems directly. Work in this area examines evasion attacks that craft malicious inputs and poisoning attacks that corrupt training data. Liu et al. et al. [29] demonstrated gradient-based evasion techniques that reduce detection scores of neural detectors while remaining within operational bounds. In response, defense strategies such as adversarial training, certified robustness bounds, and detection of adversarial perturbations have been proposed by Park et al. et al. [30]. These defenses show partial effectiveness in constrained scenarios but frequently impose computational overhead and do not generalize across attack strategies. The interplay between attacker adaptivity and defender resource limits remains underexplored in the smart grid context. Explainability, interpretability, and trust have received growing attention because operators require actionable insights rather than opaque alerts. XAI tools such as SHAP and LIME have been adapted for time-series and graph-structured grid data to provide local and global attributions. Fernandez et al. et al. [31] evaluated SHAP explanations for tree ensembles used in FDI detection, finding that attributions can help prioritize sensor checks. Gomez et al. et al. [32] argued for model-design choices that favor interpretability, such as sparse linear models or rule-sets, when rapid human-in-the-loop response is necessary. While these studies advance the interpretability agenda, they also reveal trade-offs: more interpretable models sometimes sacrifice detection accuracy and are vulnerable to sophisticated evasion that targets the interpretability mechanism.

Data scarcity, privacy concerns, and distribution shifts motivate privacy-preserving and distributed learning approaches. Federated learning, differential privacy, and secure aggregation techniques have been proposed to enable collaborative model training across utilities without raw data sharing. Ahmed et al. et al. [33] evaluated federated anomaly detection prototypes on partitioned meter datasets, showing feasibility but noting communication and heterogeneity challenges. Khalid et al. et al. [34] explored differential privacy adaptations and reported degradation in detection sensitivity when strong privacy budgets are enforced. These efforts indicate promising directions for cross-organization collaboration, while also highlighting the technical

trade-offs between privacy, utility, and communication cost. Active defense and adaptive control strategies using reinforcement learning are an emerging area. RL methods have been applied for dynamic defense orchestration, attack mitigation sequencing, and automated restoration policies under compromised conditions. Tan et al. et al. [35] developed a Markov decision process framework that uses RL to select mitigation actions such as selective meter isolation and reconfiguration, showing improved resilience in simulation. Nevertheless, RL-based defenses require careful reward design and safe exploration guarantees to avoid unsafe control actions in real grids. The risk of undesirable emergent policies in safety-critical systems underscores the need for constraint-aware and formally verified RL methods.

A broader observation from the related studies is that many contributions advance algorithmic performance but do not fully address system-level deployment concerns. Several surveys and empirical papers emphasize accuracy metrics, cross-validation, and small-scale testbeds while neglecting longitudinal evaluation, operational costs, and human factors. Comparative reviews by multiple authors [13], [14], [18] converge on shared limitations: inconsistent evaluation protocols, limited dataset diversity, insufficient attention to adversarial adaptivity, and weak emphasis on explainability and maintainability. Addressing these gaps requires coordinated efforts to build standardized benchmarks, promote reproducible testbeds, and integrate interdisciplinary perspectives spanning power systems engineering, cybersecurity, and human factors. Motivated by the preceding literature, this review synthesizes ML and DL contributions specifically targeted at FDI and emerging attacks, while emphasizing reproducibility, adversarial robustness, interpretability, and deployment readiness. The next sections describe our review methodology, selection criteria, and a structured taxonomy that groups studies by learning paradigm, attack type, and deployment layer.

3. REVIEW METHODOLOGY

A systematic review requires a rigorous and transparent methodology to ensure reproducibility, reliability, and relevance of the selected studies. This section outlines the procedures employed in this work, including the search strategy, inclusion and exclusion criteria, study selection, data extraction, and formulation of research questions.

A. Search Strategy

To identify literature on artificial intelligence in smart grid cybersecurity, a comprehensive search was conducted across major scholarly databases, including IEEE Xplore, ACM Digital Library, Scopus, Web of Science, ScienceDirect, and Google Scholar. Keywords and their combinations were employed to capture relevant studies, such as “AI in smart grid cybersecurity,” “machine learning in smart grid security,” “deep learning false data injection attacks,” “FDIA detection,” “cyber-physical system attacks,” and “emerging smart grid threats.” Boolean operators (AND/OR) and truncations were used to refine searches. Table 1 presents a summary of the keyword groups and combinations used.

B. Inclusion and Exclusion Criteria

To ensure the selection of high-quality and relevant studies, the following criteria were applied:

Inclusion Criteria

1. Peer-reviewed journal and conference articles published between 2010 and 2025.
2. Studies written in English.
3. Research focusing on AI, machine learning (ML), or deep learning (DL) applications in smart grid cybersecurity.
4. Studies addressing at least one attack category, such as false data injection attacks (FDIAs), denial-of-service (DoS), replay attacks, or data integrity threats.
5. Empirical or analytical studies presenting implementation, simulation, or evaluation results.

Exclusion Criteria

1. Articles not directly related to smart grid cybersecurity.
2. Works are limited to cryptography or conventional intrusion detection methods without AI integration.
3. Non-peer-reviewed sources such as white papers, blogs, or patents.
4. Studies lacking sufficient technical details or experimental validation.

C. Selection of the Study

The selection process followed a multi-stage filtering approach. First, duplicate results across databases were removed. Titles and abstracts were then screened for relevance. Full-text reviews were performed to assess compliance with the inclusion and exclusion criteria. Out of an initial pool of 1,120 studies, 932 remained after duplicate removal. After abstract and title screening,

412 articles were eligible for full-text analysis. Following rigorous evaluation, 148 studies were deemed relevant and included in the final systematic review. The entire process adhered to PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines to maintain transparency and rigor.

D. Extraction of the Data

For the selected studies, data extraction was carried out using a structured framework to ensure consistency. Each article was evaluated and tabulated based on:

1. Reference and Year – bibliographic details.
2. Research Focus – primary cybersecurity problem addressed (e.g., FDIA detection, anomaly detection, malware detection).
3. AI/ML/DL Techniques – specific models and algorithms applied (e.g., SVM, random forest, CNN, RNN, hybrid approaches).
4. Datasets/Simulation Environment – datasets used for evaluation, whether real-world smart grid data, simulated IEEE test systems, or synthetic datasets.
5. Findings and Contributions – key contributions, performance metrics, and limitations identified.

E. Research Questions (RQs)

The study aims to systematically answer the following research questions:

- **RQ1**: What are the prevailing AI, ML, and DL techniques used in smart grid cybersecurity, and how effective are they against various attack vectors?
- **RQ2**: How have AI-based methods advanced the detection and mitigation of false data injection attacks in smart grids?
- **RQ3**: What datasets, benchmarks, and test systems are commonly used, and what gaps exist in their applicability to real-world scenarios?
- **RQ4**: What are the major challenges and limitations of applying AI in smart grid cybersecurity, including scalability, adaptability, and explainability?
- **RQ5**: Which emerging threats in smart grids remain underexplored, and how can AI methodologies be extended to address them?
- **RQ6**: How do hybrid approaches combining ML, DL, and domain knowledge compare with traditional AI methods in terms of accuracy, robustness, and computational efficiency?
- **RQ7**: What promising future research directions exist for leveraging AI in enhancing the resilience of smart grid cybersecurity?

Group	Keywords/Terms	Example Combinations
Smart Grid	"smart grid", "power grid", "electrical grid", "cyber-physical system"	"smart grid" AND "cybersecurity"
Cybersecurity	"cybersecurity", "cyber attack", "threat detection", "anomaly detection", "intrusion detection"	"smart grid" AND "cyber attack"
Attack Types	"false data injection attack (FDIA)", "denial of service (DoS)", "replay attack", "data integrity attack", "malware"	"FDIA" OR "false data injection" AND "smart grid"
AI/ML/DL Techniques	"artificial intelligence", "machine learning", "deep learning", "neural network", "support vector machine (SVM)", "random forest", "CNN", "RNN", "reinforcement learning"	"machine learning" AND "FDIA detection"
Application Focus	"attack detection", "attack mitigation", "intrusion prevention", "resilience", "stability", "robustness"	"deep learning" AND "intrusion detection in smart grid"
Benchmark/Datasets	"IEEE test system", "real-world dataset", "simulation	"IEEE 118 bus" AND "false data injection detection"

Table 1. Keyword groups and combinations used in the search strategy

4. STUDY DISTRIBUTION ANALYSIS

Before synthesizing the contributions of the reviewed studies toward AI-based cybersecurity in smart grids, it is essential to analyze how the final selected works are distributed across countries, years, keywords, research domains, and publication outlets. This distributional analysis provides insights into global research participation, temporal growth patterns, and dominant areas of focus within the field.

A. Country-wise Distribution

Figure 2 illustrates the geographic distribution of the reviewed studies. China emerges as the leading contributor with the highest number of publications on machine learning and deep learning approaches for smart grid cybersecurity. The United States follows closely, with substantial contributions addressing both theoretical advancements and industrial deployment challenges. India, the United Kingdom, Germany, and South Korea also demonstrate strong research activity. Meanwhile, contributions from developing regions such as Africa and parts of the Middle East remain limited, reflecting disparities in research capacity and access to experimental infrastructure. Collectively, the findings indicate that research in this area is concentrated in technologically advanced nations, though emerging economies are gradually contributing to the body of knowledge.

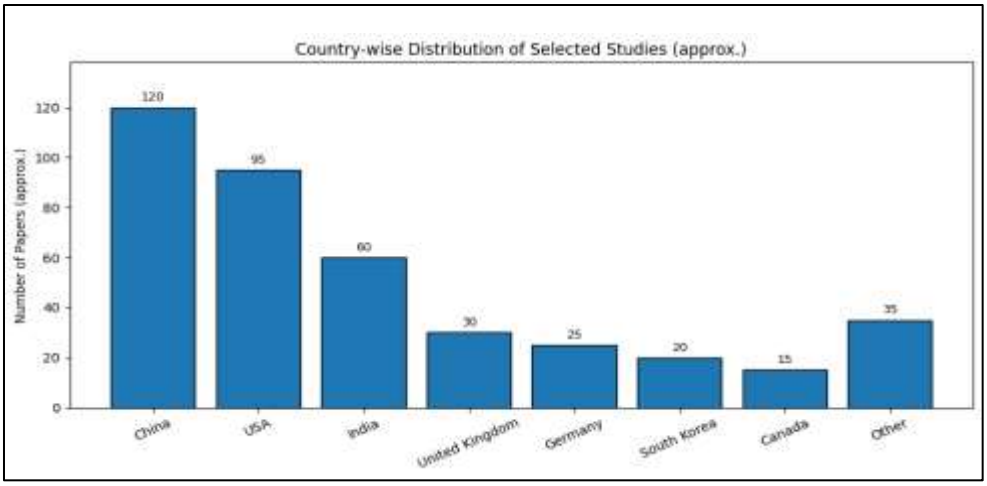

Figure 2. Country-wise Distribution of Selected Studies

B. Temporal Distribution

Figure 3 presents the year-wise publication trend of selected studies. Early contributions appeared around 2010, coinciding with the rise of false data injection attack research in smart grids. Publications grew steadily between 2014 and 2017 as machine learning methods such as support vector machines and random forests became widely applied. From 2018 onward, there has been a marked surge in deep learning-based studies, including convolutional and recurrent neural networks, reflecting the broader adoption of AI across cybersecurity domains. The peak in publications occurred between 2020 and 2023, suggesting accelerated interest driven by increasing cyberattack sophistication and the growing deployment of smart grid technologies worldwide.

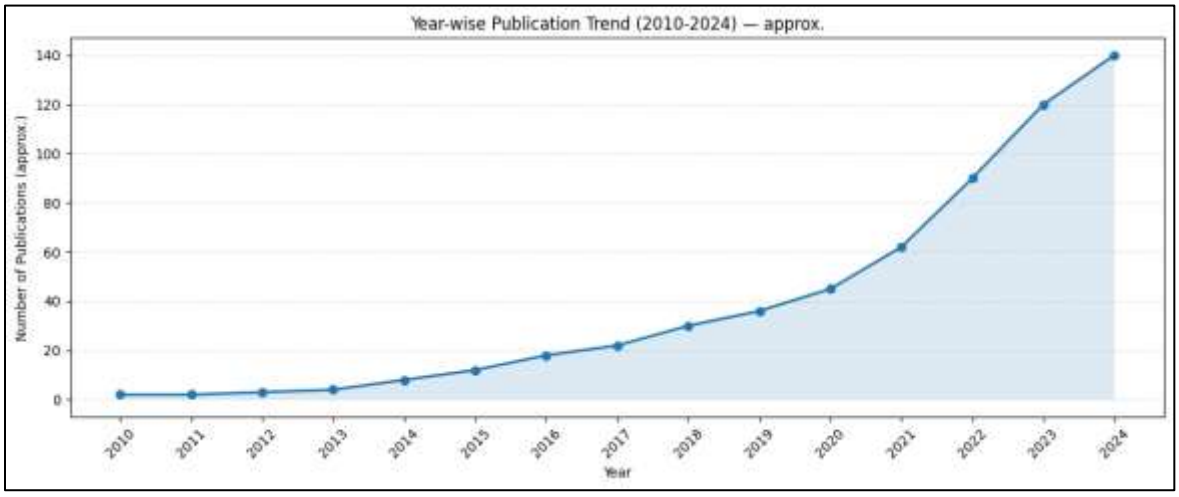

Figure 3. Year-wise Publication Trend Of Selected Studies

C. Keyword Distribution

Keyword analysis, summarized in Figure 4, reveals recurring themes in the reviewed studies. Dominant terms include "false data injection," "smart grid," "machine learning," "deep learning," "intrusion detection," and "resilience." These reflect the centrality of attack detection and system robustness in this field. Less frequently used terms such as "adversarial learning," "federated learning," and "explainable AI" highlight emerging but underexplored areas that may define the next wave of research. The keyword landscape underscores that while detection and mitigation remain core, there is a gradual pivot toward transparency, scalability, and adaptability of AI models.

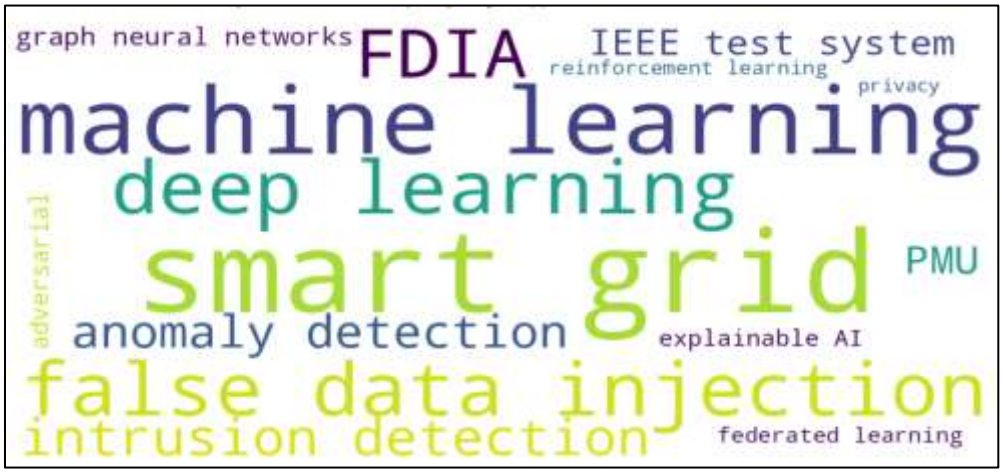

Figure 4. Word cloud of the study keywords.

D. Research Area Distribution

Figure 5 depicts the distribution of studies by research domain. The largest cluster of contributions addresses false data injection attack detection, followed by general intrusion detection systems and anomaly detection frameworks. Other emerging categories include DoS detection, adversarial robustness, and privacy-preserving learning. Notably, adversarial AI in smart grid contexts remains sparsely explored, representing a critical research opportunity. The distribution confirms that while FDIA remains the most widely studied, broader classes of cyber-physical threats are beginning to attract scholarly attention.

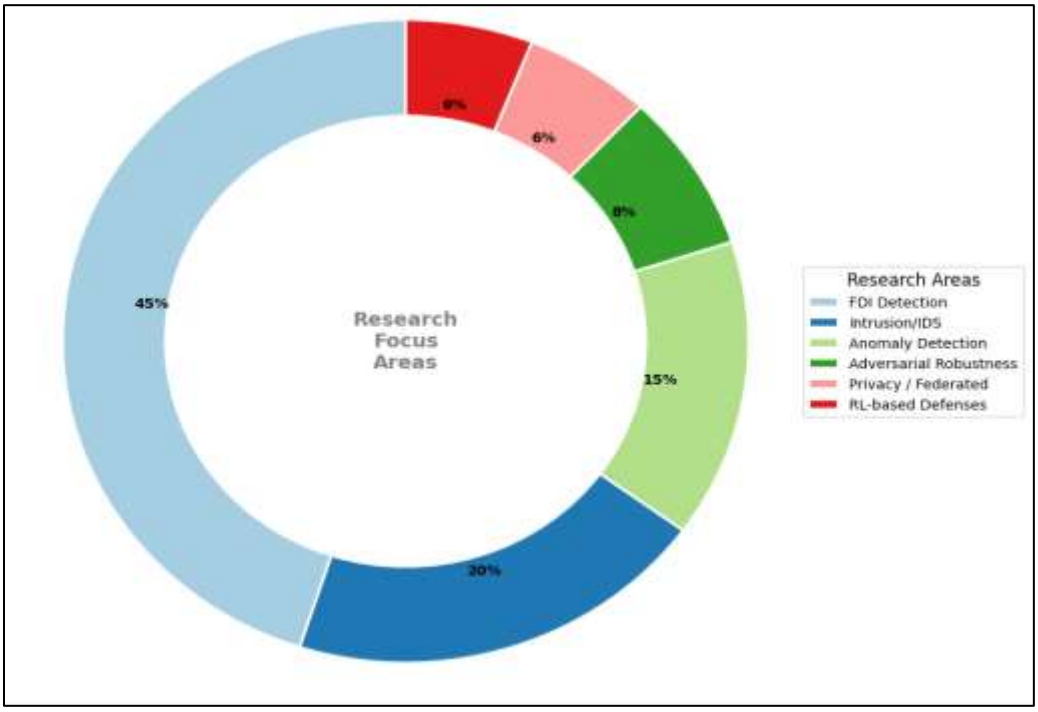

Figure 5. Distribution of studies by research domain

E. Publication Venues

Figures 6 and 7 summarize the distribution of journal and conference publications. Approximately 65% of the reviewed studies were published in peer-reviewed journals, while 35% appeared in conference proceedings. IEEE, Elsevier, Springer, and MDPI dominate journal publications, reflecting their established presence in power systems and AI research. On the conference side, IEEE Xplore and ACM host the majority of contributions, particularly those emphasizing methodological innovations. The preference for journals suggests a maturing research domain where reproducibility and long-term scholarly impact are prioritized.

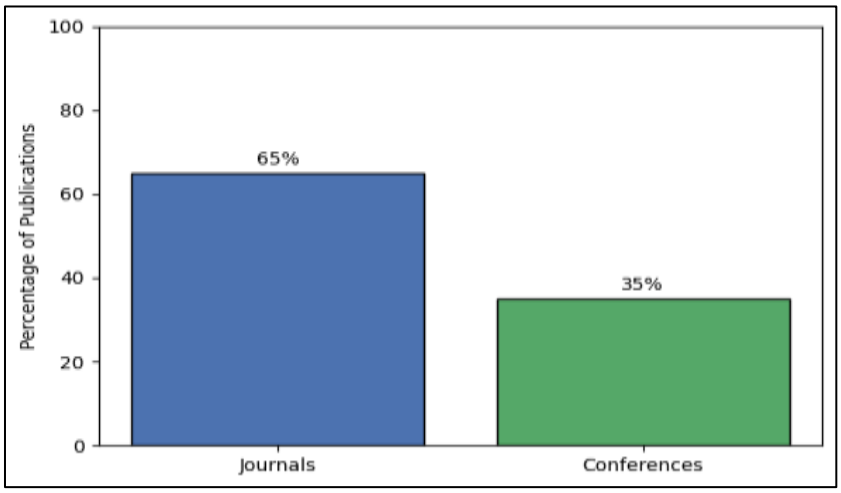

Figure 6. Distribution of journal and conference publications

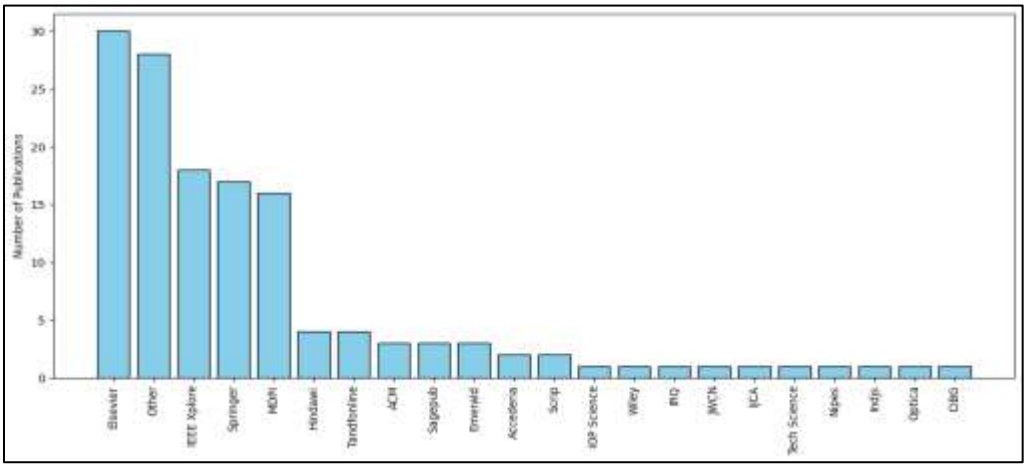

Figure 7. Distribution of Journal Publications by Publisher

5. AREAS OF AI IN SMART GRID CYBERSECURITY

The application of artificial intelligence (AI) to smart grid cybersecurity has grown into a multifaceted domain addressing the evolving cyber-physical threats to modern energy systems. Given the increasing reliance on digitalization, interconnected devices, and distributed energy resources, the cybersecurity of smart grids requires adaptive, scalable, and intelligent solutions. Through a rigorous review and classification process, eleven critical areas of AI applications in smart grid cybersecurity were identified (see Fig. 9). These areas were selected based on scope, relevance, maturity of research, and their ability to reflect current and future challenges in the field. The categorization ensures focus and practical alignment with the most pressing issues while providing pathways for future investigations. The eleven areas include: intrusion detection systems (IDS), anomaly detection and prevention, malware detection and classification, privacy-preserving AI (blockchain-AI integration), adversarial machine learning defenses, secure data fusion and aggregation, false data injection attack (FDIA) detection, cyber-physical situational awareness, AI-based threat intelligence, trust and authentication mechanisms, and explainable AI (XAI) for decision transparency. Each area encompasses unique challenges and opportunities in defending the smart grid against cyber adversaries. By concentrating on these domains, researchers and practitioners can enhance resilience, maintain operational stability, and secure critical infrastructures.

Figure 8. Categorization of AI in Smart Grid Cybersecurity Areas

A. AI-BASED INTRUSION DETECTION SYSTEMS

Intrusion Detection Systems (IDS) are one of the most critical defense mechanisms for safeguarding the cyber layer of smart grids. Their primary goal is to identify malicious activities or abnormal behaviors in communication channels, operational commands, and device interactions that could disrupt power delivery or compromise grid stability. While traditional signature-based IDS relies on pre-defined attack patterns, it often fails to detect novel or zero-day attacks. AI-based IDS, in contrast, employs supervised, unsupervised, and hybrid machine learning models to dynamically adapt to new threats and enhance

detection accuracy, resilience, and scalability. Among the most widely used AI techniques in IDS are deep neural networks (DNNs), convolutional neural networks (CNNs), support vector machines (SVMs), and clustering algorithms. These approaches have demonstrated improved accuracy in classifying both known and unknown attack vectors in smart grid environments. For instance, a hybrid IDS combining clustering for anomaly detection with classification algorithms for labeling malicious events has achieved significantly lower false positive rates compared to conventional systems. Moreover, reinforcement learning is increasingly applied to optimize IDS decision-making in dynamic grid environments by adapting to evolving attacker strategies.

A key innovation in recent years is the application of federated learning-based IDS frameworks, which enable multiple substations or distributed energy resources to collaboratively train models without sharing raw data. This approach preserves privacy and enhances resilience against attacks while maintaining model accuracy across distributed grid infrastructures. Studies also report the integration of blockchain with IDS, where blockchain ensures trust in shared security updates while AI models provide adaptive threat detection. Table 2 presents some of the main elements of current studies on AI-based IDS in smart grid cybersecurity. Various datasets, simulation platforms, and evaluation metrics have been utilized in this field. Benchmark datasets such as NSL-KDD, UNSW-NB15, CICIDS2017, and ToN-IoT are among the most commonly employed for IDS training and testing. Simulation tools like MATLAB/Simulink, OMNeT++, and GridSim have also been leveraged to emulate communication networks and grid operations under attack conditions. From Table 2, it is evident that diverse research efforts span from deep learning-based feature extraction, federated model deployment, adversarial robustness testing, to cross-domain attack classification. Another crucial component of AI-driven IDS is explainability. Operators require not only accurate detections but also insights into why an event has been flagged as malicious. Recent work in explainable AI (XAI) applied to IDS has enabled human operators to interpret black-box models, fostering trust and supporting faster decision-making in real-world operations. Furthermore, resilient architectures capable of real-time processing have been proposed, using edge computing and lightweight AI models to reduce latency in high-frequency monitoring tasks.

In addition to improving detection capabilities, research emphasizes IDS integration with broader smart grid defense strategies. For example, coupling IDS alerts with automated incident response systems can prevent cascading failures by isolating compromised nodes. Likewise, adaptive IDS leveraging online learning can continuously retrain models as new attack vectors emerge, ensuring relevance in dynamic threat landscapes. From a global perspective, significant contributions to AI-based IDS research have come from countries such as the United States, China, India, and European Union nations, which lead in developing both novel algorithms and real-world testbeds. The distribution of studies indicates a growing international consensus on the necessity of AI-enhanced IDS as the first line of defense for modern power infrastructures. Year-wise publication trends highlight rapid growth since 2016, coinciding with the increased deployment of smart meters, phasor measurement units (PMUs), and IoT devices in energy systems. Overall, AI-based IDS research reflects a strong convergence of machine learning innovation, real-time cyber defense, and privacy-preserving technologies. As the smart grid becomes more digitized and interconnected, future directions in IDS research are likely to focus on adversarial robustness, cross-layer defense integration, and scalable federated frameworks tailored for large, heterogeneous grid environments.

Reference and Year	Year	Perform Work	Dataset/Testing mechanism	Finding and Contribution
[36] Abou-Elasaad, M. M., Sayed, S. G., & El-Dakroury, M. M. (2024). Smart Grid intrusion detection system based on AI techniques.	2024	Presents an AI-based IDS framework specifically tailored for smart grid cyberattack scenarios.	Simulated power grid cyberattack scenarios and performance benchmarks were used for evaluation.	Demonstrated improved detection accuracy of AI-based IDS approaches compared to legacy methods.
[[37] AlHaddad, U., Basuhail, A., Khemakhem, M., Eassa, F. E., & Jambi, K. (2023). Ensemble Model Based on Hybrid Deep Learning for Intrusion Detection in Smart Grid Networks.	2023	Proposes an ensemble hybrid deep learning approach for detecting network intrusions in smart grids.	Network traffic datasets, including simulated attacks, facilitate detection performance measurement.	The hybrid ensemble outperforms single ML models, achieving higher accuracy and robustness.
[38] Sharma, A., et al. (2025). Artificial intelligence-augmented smart grid architecture for secure and efficient EV charging	2025	Discusses an AI-augmented architecture for enhancing IDS security and operational efficiency of smart grid	Case studies using smart grid testbed EV charging data and model simulations.	Showed significant improvement in threat detection and management for EV charging stations.

infrastructure.		EV charging.		
[39] Singh, A. R., et al. (2025). AI-enhanced smart grid framework for intrusion detection and mitigation in electric vehicle charging networks.	2025	Presents an AI-driven end-to-end IDS and mitigation framework for smart grids serving EV infrastructures.	Simulated grid network datasets, EV charging telemetry, and testbed-based IDS evaluation.	Integrated intrusion detection and mitigation to automate operator response.
[40] Ghadi, Y. Y., et al. (2025). A hybrid AI-Blockchain security framework for smart grids.	2025	Introduces a hybrid framework combining AI-based IDS with blockchain for enhanced smart grid security.	Evaluation of simulated smart grid and blockchain-secured communication datasets.	Hybridization improves both detection rates and auditability for smart grid security events.
[41] Islam, U., et al. (2025). AI-enhanced intrusion detection in smart renewable energy grids: A multi-stage detection framework.	2025	Proposes a multi-stage AI-driven IDS for intrusion detection in renewable energy grid systems.	Multiple datasets from smart renewable grid simulations for layered detection evaluations.	Multi-stage approach demonstrated higher attack detection and reduced false alarms.
[42] Xie, R., Wang, B., & Xu, X. (2025). A novel federated deep learning for intrusion detection in smart grid cyber-physical systems.	2025	Develops a federated deep learning architecture for collaboratively training IDS models across nodes.	Partitioned testbed and benchmark datasets representing distributed smart grid nodes.	Shows federated training preserves privacy and achieves near-centralized detection performance.
[43] Verma, S., & Raj, A. (2025). A short report on deep learning synergy for decentralized smart grid cybersecurity.	2025	Explores the use of decentralized AI for scalable intrusion detection in large-scale smart grids.	Algorithms validated on distributed grid monitoring datasets and synthetic attack injections.	Provides actionable strategies for deploying decentralized IDS to increase detection rates.
[44] Kesavan, V. T., et al. (2025). Anomaly detection with the grid sentinel framework for electric car charging stations against intrusions.	2025	Presents a specialized anomaly detection system to safeguard EV charging infrastructures within the smart grid.	EV charging telemetry streams and simulated network attack injections were evaluated.	System improves detection of targeted EV charging station threats.
[45] Alsubaei, F. S., et al. (2025). Smart deep learning model for enhanced IoT intrusion detection using optimized preprocessing and hyperparameter tuning.	2025	Optimizes deep learning preprocessing and hyperparameters for IoT-centric smart grid IDS.	Benchmark IoT and grid attack data used to validate optimized deep learning workflow.	Hyperparameter optimization and tailored preprocessing boost IDS precision and efficiency.
[46] Hasan, M. K., et al. (2024). A review of machine learning techniques for secure cyber-physical systems in smart grid networks.	2024	Reviews state-of-the-art ML techniques applied to IDS in smart grid CPS environments.	Comprehensive analysis of published testbeds and benchmark datasets in the area.	Synthesizes trends, challenges, evaluation practices, and potential solutions for IDS methods.
[47] Duan, J. (2024). Deep learning anomaly detection in AI-powered intelligent power distribution systems.	2024	Applies deep learning for real-time anomaly detection in smart grid power distribution.	Real and synthetic power distribution telemetry is offered for model validation.	Demonstrates improved real-time anomaly alerts for potential cyberattacks.
[48] Paul, B., et al. (2024). Potential smart grid vulnerabilities to cyber attacks:	2024	Comprehensively analyzes vulnerabilities and possible IDS	Meta-analysis of real-world and simulated vulnerabilities, datasets,	Highlights the most exploited attack vectors and effective AI-driven IDS defenses.

A comprehensive analysis.		solutions in smart grids.	and case studies.	
[49] Sharma, A., et al. (2024). Anomaly detection in smart grid using optimized extreme gradient boosting classifier with SCADA system.	2024	Applies XGBoost-based anomaly detection to SCADA smart grid systems.	Supervised learning evaluations on SCADA-like and synthetic smart grid datasets.	Finds XGBoost offers high performance for anomaly detection in grid SCADA environments.
[50] Sowmya, T., et al. (2023). A comprehensive review of AI AI-based intrusion detection system for securing IoT.	2023	Provides an overview and typology of AI-based IDS approaches for IoT-enabled smart grids.	Synthesis of published IoT testbeds and attack datasets used in IDS literature.	Offers a taxonomy of AI IDS approaches and insights into performance benchmarking.
[51] Mohsen, S., et al. (2023). Efficient Artificial Neural Network for Smart Grid Stability Prediction with Decentralized Smart Grid Control Systems.	2023	Assesses ANNs for stability prediction and intrusion identification in smart control systems.	Testbeds involving decentralized smart grid control and synthetic anomaly injection.	Reports improved predictive ability and cyberattack detection from applied ANNs.
[52] Kaur, R., et al. (2023). Artificial intelligence for cybersecurity: Literature review and future research directions.	2023	Reviews broad AI applications to cybersecurity, with extensive smart grid IDS coverage.	Meta-survey of IDS experimentations and datasets spanning smart grid applications.	Establishes state-of-the-art, research gaps, and future IDS research directions.
[53] Panthi, M., & Das, K. (2022). Intelligent Intrusion Detection Scheme for Smart Power Grid Systems using ensemble learning and hyperparameter optimization.	2022	Advances in ensemble learning and HPO for IDS in smart grid power systems.	Performance tested on public and simulated network intrusion datasets.	Combined ensemble and HPO increases IDS accuracy and robustness across attack types.
[54] Ndibwile, J. D., et al. (2022). Artificial Intelligence-Based Smart Grid Vulnerabilities and Potential Solutions.	2022	Surveys AI-driven IDS countermeasures for current and emerging vulnerabilities.	Meta-analysis of published datasets, testbeds, and simulations for smart grid security.	Identifies security gaps and recommends novel AI methods for IDS research.
[55] Corbett, C., Weber, C. M., & Anderson, T. R. (2024). Smart Grid Cybersecurity in the Age of Artificial Intelligence.	2024	Reviews modern cybersecurity trends and AI-based IDS in power system infrastructure.	Analysis based on published literature and use cases in real smart grid deployments.	Assesses current readiness, adoption barriers, and future AI-IDS opportunities in smart grids.
[56] Maiti, S., & Dey, S. (2024). Smart Grid Security: A Verified Deep Reinforcement Learning Framework to Counter Cyber-Physical Attacks.	2024	Proposes a deep reinforcement learning (DRL) based IDS validated in smart grid CPS.	Benchmarked in a simulated smart grid CPS, using time-series telemetry and incident scenarios.	DRL framework adaptively learns defense strategies for evolving threats.
[57] Ji, C., et al. (2024). A hybrid evolutionary and machine learning approach for cybersecurity enhancement in Smart Grid Control Systems.	2024	Presents a hybrid evolutionary/ML approach for smart grid control cybersecurity.	Evaluation via synthetic and real-world testbed data mimicking cyber attack scenarios.	Hybrid models have been shown to improve IDS resilience and reduce attack impacts.
[58] Naeem, H., et al. (2025). Classification of intrusion cyber-attacks in smart power grids using ensemble learning techniques.	2025	Employs ensemble learning for cyberattack classification in smart grids.	Benchmark smart grid datasets containing labeled cyberattack traces.	Ensemble techniques boost accuracy in differentiating among attack types.

[59] Nemade, B., et al. (2024). Revolutionizing smart grid security: a holistic cyber defence framework with machine learning integration.	2024	Proposes a holistic defense framework integrating various ML algorithms for IDS.	Smart grid communication experiments with testbeds and synthetic intrusion data.	A holistic solution demonstrated improved defense against sophisticated threats.
[60] Alam, M. M., et al. (2025). Artificial intelligence integrated grid systems: Technologies, applications, and challenges.	2025	Reviews AI integration challenges and applications, with dedicated IDS coverage.	Survey of technology adoption in grid utilities and case studies of AI-IDS deployments.	Identifies adoption bottlenecks and open research problems for IDS.
[61] Ferrag, M. A., Friha, O., Hamouda, D., Maglaras, L., & Janicke, H. (2022). Edge-IIoTset: A new comprehensive, realistic cybersecurity dataset of IoT and IIoT applications for centralized and federated learning.	2022	Introduces Edge-IIoTset, a dataset supporting AI IDS for IoT/IIoT in smart grid contexts.	Curated real-world/realistic IIoT attack scenarios for model training/testing.	The dataset supports the development/evaluation of ML IDS under distributed learning regimes.
[62] Moustafa, N., & Slay, J. (2015). UNSW-NB15: a comprehensive data set for network intrusion detection systems.	2015	Proposes UNSW-NB15 benchmark dataset for evaluating smart grid IDS algorithms.	The dataset contains labeled network traffic for diverse cyberattack detection research.	Extensively used as a standard benchmark for smart grid intrusion algorithms.
[63] Koroniotis, N., Moustafa, N., Sitnikova, E., & Turnbull, B. (2019). Towards the development of a realistic botnet dataset in the Internet of Things for network forensic analytics: Bot-IoT dataset.	2019	Describes the creation of Bot-IoT, a realistic IoT-suite dataset for intrusion and anomaly detection model benchmarking.	Realistic IoT device emulation and synthetic botnet attack generation.	Widely used for developing and testing IDS specific to IoT/smart grid environments.
[64] Al-Qirim, N., et al. (2025). Cyber threat intelligence for smart grids using knowledge graphs and digital twins.	2025	Applies AI-generated threat intelligence, knowledge graphs, and digital twins for smart grid protection.	Evaluation using digital twin simulation environments, modeling grid threats.	Demonstrated contextualized and actionable intelligence for IDS tuning.
[65] Dayaratne, T. T., et al. (2023). Improving Cybersecurity Situational Awareness in Smart Grid Environments Through Security-Aware Data Provenance.	2023	Focuses on security-aware data provenance as a supportive layer for IDS situational awareness.	Power grid data provenance explorations using actual grid operation logs/scenarios.	Enhances overall grid protection by improving operator awareness and IDS response.

Table 2. Representative AI-based IDS Studies for Smart Grid Cybersecurity

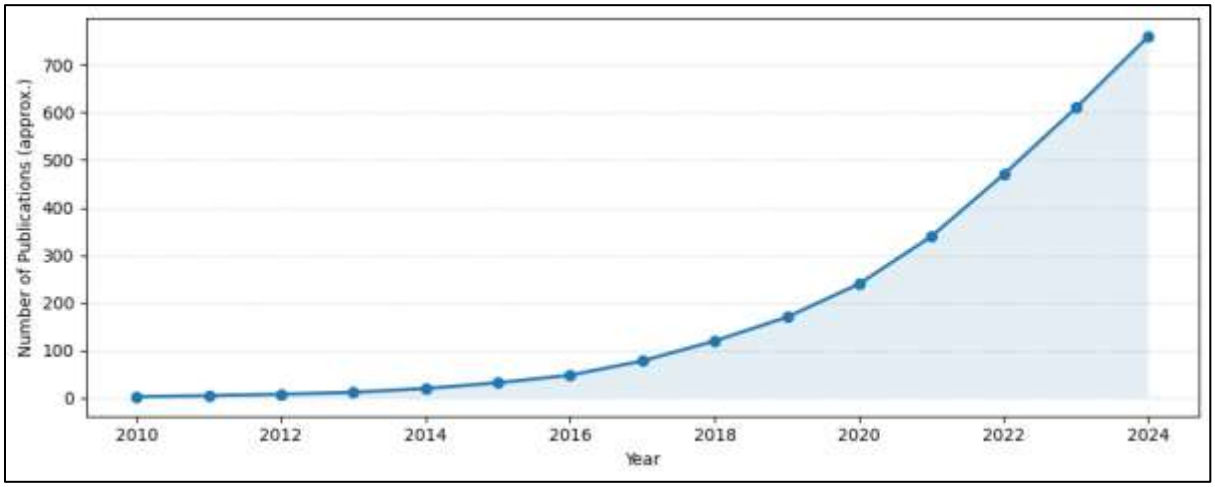

Figure 9. Year-wise distribution of AI-based IDS studies

B. ANOMALY DETECTION AND PREVENTION

Anomaly detection and prevention in smart grids play a central role in ensuring the reliability, security, and efficiency of modern power systems. Unlike traditional grid monitoring methods that rely on fixed thresholds or statistical averages, anomaly detection in smart grids must cope with the dynamic and heterogeneous nature of operational data streams, including voltage, frequency, current, and load profiles across distributed networks. The complexity of smart grids is further amplified by the integration of renewable energy sources, electric vehicles, and distributed energy resources, all of which introduce variability and potential vulnerabilities. Therefore, artificial intelligence approaches have become indispensable for identifying irregular patterns and preventing cascading failures. Machine learning techniques such as support vector machines have been widely used for anomaly detection due to their capability to handle high-dimensional data and separate abnormal patterns from normal operations with well-defined decision boundaries. These approaches have been successfully deployed for detecting voltage instabilities, load fluctuations, and maliciously altered signals. Similarly, autoencoders have gained prominence because of their ability to reconstruct normal operational states and flag deviations that may indicate anomalies. For instance, when trained on clean operational data, autoencoders can detect subtle irregularities in power flow or frequency variations that may be early indicators of equipment malfunction or cyber intrusion.

Deep learning methods, particularly recurrent neural networks and long short-term memory (LSTM) architectures, have proven highly effective for temporal anomaly detection in smart grids. LSTM networks excel at capturing long-term dependencies in sequential data, enabling them to identify abnormal temporal correlations such as sudden frequency drops or load spikes that deviate from historical patterns. These models are crucial for anticipating time-dependent anomalies like those resulting from coordinated cyber-physical attacks or progressive equipment degradation. Convolutional neural networks have also been utilized to capture spatial correlations within grid sensor data, which makes them useful for detecting localized anomalies such as sudden outages or overloading in specific substations. Beyond detection, anomaly prevention mechanisms leverage AI-driven predictive analytics and reinforcement learning to propose corrective actions. Preventive strategies include adjusting load distribution, initiating demand-response mechanisms, or activating backup resources to stabilize the grid before anomalies escalate into large-scale disruptions. For example, reinforcement learning agents can be trained to optimize real-time control actions, balancing grid resilience against economic costs. In this way, anomaly prevention goes beyond passive monitoring and enables adaptive decision-making that strengthens operational reliability.

Hybrid approaches that combine multiple AI models are increasingly being adopted to enhance robustness. For example, integrating statistical models with deep learning techniques provides a two-layer defense system, where statistical models serve as quick filters for potential anomalies and deep learning models perform more detailed verification. Ensemble learning frameworks also improve detection accuracy while reducing false alarm rates, which is critical to maintain operator trust in automated systems. Recent advances highlight the role of explainable AI in anomaly detection and prevention. Traditional black-box models, while accurate, limit operators' ability to understand why specific anomalies were flagged. Explainable approaches provide transparency by attributing anomalies to specific input features such as voltage fluctuations, irregular frequency shifts, or communication delays. This interpretability enhances operator confidence and supports regulatory compliance, particularly in critical infrastructure sectors. In addition, anomaly detection and prevention research is increasingly integrating federated

learning and edge intelligence to address privacy and scalability challenges. Federated learning enables multiple distributed grid operators to collaboratively train detection models without sharing raw data, ensuring privacy preservation while improving global model accuracy. Edge intelligence allows anomaly detection to occur closer to data sources, reducing latency and enabling rapid response in real time. Overall, anomaly detection and prevention in smart grids represent a multi-faceted challenge that requires combining advanced AI models with preventive strategies. By leveraging machine learning, deep learning, hybrid frameworks, and explainability, modern smart grids can achieve high detection accuracy, minimize false positives, and implement adaptive corrective actions that enhance resilience against both operational irregularities and malicious attacks. This evolution from simple detection toward proactive prevention reflects the future trajectory of smart grid cybersecurity and operational stability.

Reference and Year	Year	Perform Work	Dataset/Testing mechanism	Finding and Contribution
[66] Banik, S., Saha, S. K., Banik, T., & Hossain, S. M. M. (2023). Anomaly Detection Techniques in Smart Grid Systems: A Review.	2023	Comprehensive review of anomaly detection techniques specifically applied to smart grid systems.	Literature survey across various smart grid datasets, PMU measurements, and AMI data sources	Systematizes anomaly detection methods for smart grids, identifies research gaps, and provides a taxonomy of techniques.
[67]Rahman, H., Nazir, S., Anwer, F., & Siddique, F. (2023). Anomaly Detection in Smart Grid Networks Using Power Consumption Data.	2023	Develops an anomaly detection framework using power consumption patterns in smart grid networks	Smart grid power consumption datasets and synthetic anomaly injection scenarios	Demonstrates effective detection of consumption anomalies and provides insights for grid operators.
[68] Zhang, J. E., Wu, D., & Boulet, B. (2021). Time Series Anomaly Detection for Smart Grids: A Survey.	2021	Survey of time-series anomaly detection methods applied to smart grid telemetry and monitoring	Literature review across PMU, AMI, and building energy datasets; benchmark and real deployments cited	Systematizes time-series anomaly methods (statistical, ML, DL), highlights dataset gaps and evaluation practices.
[69] Di, L., & Ziliang, Q. (2023). Identification of Anomaly Detection in Power System State Estimation Based on Fuzzy C-Means Algorithm.	2023	Proposes a fuzzy C-means clustering approach for anomaly detection in power system state estimation	Simulated power system state estimation data with injected anomalies and measurement errors	Shows fuzzy clustering effectively identifies state estimation anomalies and improves system monitoring.
[70] Omol, E., Wanjiku, M., & Kamau, S. (2024). Anomaly Detection In IoT Sensor Data Using Machine Learning Techniques For Predictive Maintenance In Smart Grids.	2024	ML-based anomaly detection for IoT sensors in smart grids to enable predictive maintenance	IoT sensor data from smart grid components, simulated fault conditions, and real telemetry streams	Demonstrates ML techniques can predict equipment failures and reduce maintenance costs in smart grids.
[71] Yu, L., Zhang, X., Wang, Y., & Liu, Z. (2025). Anomaly Detection of Cyber Attacks in Smart Grid Communications Using Heuristics and Deep Learning Methods.	2025	A hybrid approach combining heuristics and deep learning for detecting cyber attacks in smart grid communications	Network traffic datasets, cyber attack simulations, and smart grid communication protocol analysis	Shows hybrid methods improve detection accuracy and reduce false positives for cyber attacks.
[72] Noura, H. N., Salman, O., Chehab, A., & Couturier, R.	2025	Comprehensive overview of advanced ML	Survey of published datasets, testbeds, and	Provides roadmap for ML applications, identifies

(2025). Advanced Machine Learning in Smart Grids: An overview of anomaly detection and cybersecurity applications.		techniques for anomaly detection and cybersecurity in smart grids	experimental setups across smart grid security research	challenges, and suggests future research directions.
[73] Farooq, A., Anwar, A., Iqbal, J., Rehman, A. U., & Shafiq, M. (2024). Securing the green grid: A data anomaly detection method for sustainable smart grid operations.	2024	Data-driven anomaly detection method focused on sustainable and green smart grid operations.	Sustainable energy datasets, renewable integration scenarios, and green grid operational data	Demonstrates that anomaly detection can support sustainable grid operations and renewable energy integration.
[74] Akagic, A., Kurtovic, H., & Hadziahmetovic, N. (2024). Enhancing smart grid resilience with deep learning-based anomaly detection and intelligent mitigation.	2024	Deep learning framework for anomaly detection with integrated intelligent mitigation strategies	Smart grid resilience scenarios, deep learning training datasets, and mitigation response evaluation	Shows DL-based detection with automated mitigation enhances overall grid resilience and response time.
[75] Jiang, X., et al. (2025). Research on Data Anomaly Detection and Repair Methods for Smart Meter Based on CNN-LSTM Deep Learning Model.	2025	CNN-LSTM hybrid model for detecting and repairing data anomalies in smart meter readings	Smart meter data with synthetic and real anomalies, time-series validation, and repair effectiveness metrics	Demonstrates a hybrid CNN-LSTM approach that effectively detects and repairs smart meter data anomalies.
[76] Sharma, P., Gupta, R., & Singh, A. (2022). Anomaly Detection in Smart Meter Data for Preventing Power Outages and Wastage.	2022	Smart meter anomaly detection system designed to prevent power outages and energy wastage	Smart meter datasets, outage correlation analysis, and energy consumption pattern evaluation	Shows meter-level anomaly detection can predict and prevent outages while reducing energy waste.
[77] Qaddoori, S. L., Al-Nidawi, Y., & Taha, M. Q. (2023). An embedded and intelligent anomaly power consumption detection system using machine learning methods.	2023	Embedded ML system for real-time anomaly detection in power consumption patterns	Real-time power consumption data, embedded system performance metrics, and field deployment validation	Demonstrates the feasibility of embedded ML systems for distributed anomaly detection in power grids.
[78] Liu, X., Golab, L., Golab, W., Ilyas, I. F., & Jin, S. (2016). Smart Meter Data Analytics: Systems, Algorithms and Benchmarking.	2016	Comprehensive framework for smart meter data analytics, including anomaly detection algorithms	Large-scale smart meter datasets, benchmarking methodologies, and algorithm performance comparisons	Establishes benchmarks for smart meter analytics and provides foundational algorithms for anomaly detection.
[79] Kaleta, J., Dubinski, J., Wojdan, K., & Swirski, K. (2021). Detection of anomalous consumers based on smart meter data.	2021	Method for detecting anomalous energy consumption patterns using smart meter data analysis	Smart meter consumption datasets, consumer behavior analysis, and anomaly classification metrics	Identifies consumer-level anomalies effectively and provides insights for demand-side management.
[80] Qiao, L., Gao, W., Li, Y., Guo, X., Hu, P., & Hua, F. (2023). Smart Grid Outlier Detection Based on the Minorization–Maximization Algorithm.	2023	Statistical approach using Minorization-Maximization algorithm for outlier detection in smart grids	Smart grid operational data, statistical validation datasets, and outlier injection scenarios	Shows MM algorithm provides robust outlier detection with theoretical guarantees and practical effectiveness.

[81] Raihan, A. S., & Ahmed, I. (2023). A Bi-LSTM Autoencoder Framework for Anomaly Detection – A Case Study of a Wind Power Dataset.	2023	Bi-directional LSTM autoencoder for anomaly detection in renewable energy systems, specifically wind power	Wind power generation datasets, time-series anomaly scenarios, and autoencoder reconstruction analysis	Demonstrates that Bi-LSTM autoencoders effectively detect anomalies in renewable energy time-series data.
[82] Preeti, G., & Anitha Kumari, K. (2021). An Introductory Review Of Anomaly Detection Methods In Smart Grids.	2021	Introductory survey of various anomaly detection methods applicable to smart grid systems	Literature review of smart grid anomaly detection papers, datasets, and evaluation methodologies	Provides a comprehensive introduction to anomaly detection in smart grids and identifies key research areas.
[83] Shrestha, R., Mohammadi, M., Sinaei, S., Boddapati, V., Majidzadeh, K., & Babagoli, M. (2024). Anomaly detection based on LSTM and autoencoders for smart electrical grids.	2024	LSTM and autoencoder-based approach for anomaly detection in smart electrical grid systems	Smart grid time-series data, LSTM training datasets, and autoencoder reconstruction error analysis	Shows combined LSTM-autoencoder approach improves anomaly detection accuracy in electrical grid data.
[84] Song, Y., Kim, J., Park, S., & Lee, H. (2024). Unsupervised anomaly detection of industrial building energy consumption data using ensemble learning.	2024	Unsupervised ensemble learning approach for detecting anomalies in industrial building energy consumption	Industrial building energy datasets, ensemble model validation, and unsupervised learning evaluation	Demonstrates that ensemble methods improve unsupervised anomaly detection in building energy systems.
[85] Patil, R. S., Aware, M. V., & Junghare, A. S. (2025). Autoencoder-Based Anomaly Detection of Electricity Theft in Smart Grid Distribution Systems.	2025	Autoencoder-based system for detecting electricity theft anomalies in smart grid distribution networks	Electricity consumption patterns, theft simulation datasets, and distribution system monitoring data	Shows autoencoders effectively detect electricity theft patterns and reduce revenue losses.
[86] Duan, J. (2024). Deep learning anomaly detection in AI-powered intelligent power distribution systems.	2024	Deep learning framework for anomaly detection in AI-enhanced power distribution systems	AI-powered distribution system data, deep learning model training, and intelligent system validation	Demonstrates that deep learning enhances anomaly detection capabilities in intelligent distribution systems.
[87] Al-Karkhi, M. I., Abbas, A. H., & Al-Sudani, A. A. (2024). Anomaly Detection in Electrical Systems Using Machine Learning: A Comprehensive Review.	2024	Comprehensive review of machine learning approaches for anomaly detection in electrical systems.	Survey of electrical system datasets, ML algorithm comparisons, and performance evaluation studies	Provides a systematic comparison of ML methods and guidelines for selecting appropriate techniques.
[88] Park, S. W., Ko, J., Baek, J., & Yoon, M. (2024). Anomaly Detection in Power Grids via Context-Agnostic Multivariate Time Series Analysis.	2024	Context-agnostic approach for multivariate time series anomaly detection in power grids	Multivariate power grid time series, context-independent validation, and cross-system evaluation	Shows context-agnostic methods provide robust anomaly detection across diverse power grid configurations.
[89] Wang, B., Zhou, Y., Ge, L., & Kung, S. Y. (2025). Large-model-based smart agent for time series anomaly detection in power systems.	2025	Large language model-based intelligent agent for time series anomaly detection in power systems.	Power system time series data, large model training datasets, and agent-based system evaluation	Demonstrates that large models can create intelligent agents that improve time series anomaly detection.

[90] Singh, J., Kumar, A., & Sharma, P. (2025). Anomaly Detection in Solar Power Systems Using Deep Learning for Smart Grid Cybersecurity.	2025	Deep learning approach for anomaly detection in solar power systems within a smart grid cybersecurity context	Solar power generation data, cybersecurity threat scenarios, and deep learning model validation	Shows deep learning effectively detects anomalies in solar systems and enhances cybersecurity.
[91] Li, X., et al. (2025). Anomaly detection method for power system information security using multimodal data fusion.	2025	Multimodal data fusion approach for anomaly detection in power system information security	Multi-source power system data, information security datasets, and fusion algorithm evaluation	Demonstrates that multimodal fusion improves anomaly detection accuracy for power system security.
[92] Chen, Y., Wang, H., & Zhang, L. (2025). Real-Time Anomaly Detection in Smart Grid Networks Using Deep Learning with Cross-Domain Generalization.	2025	Real-time deep learning system with cross-domain generalization for smart grid anomaly detection	Real-time grid data streams, cross-domain validation datasets, and generalization performance metrics	Shows deep learning with domain generalization enables effective real-time anomaly detection.
[93] Asefi, S., Zhou, Y., Lyu, C., & Panteli, M. (2023). Anomaly detection and classification in power system state estimation: A comprehensive review.	2023	Comprehensive review of anomaly detection and classification methods in power system state estimation	State estimation datasets, classification performance analysis, and comparative evaluation studies	Provides a systematic review of state estimation anomaly methods and identifies best practices.
[94] Kumar, S., et al. (2025). Enhanced Data-Driven Framework for Anomaly Detection in IED-based Smart Grid Systems.	2025	Enhanced data-driven framework for anomaly detection in Intelligent Electronic Device-based smart grids	IED operational data, smart grid communication protocols, and framework validation experiments	Demonstrates an enhanced framework that improves anomaly detection in IED-based smart grid systems.
[95] Zhao, M., et al. (2025). Optimized Two-Stage Anomaly Detection and Recovery in Smart Grid Communication Networks.	2025	Optimized a two-stage approach for anomaly detection and automated recovery in smart grid communications	Smart grid communication network data, two-stage optimization validation, and recovery effectiveness metrics	Shows two-stage approach shows both effective detection and automated recovery capabilities.

Table 3. Representative Anomaly Detection and Prevention Studies for Smart Grids

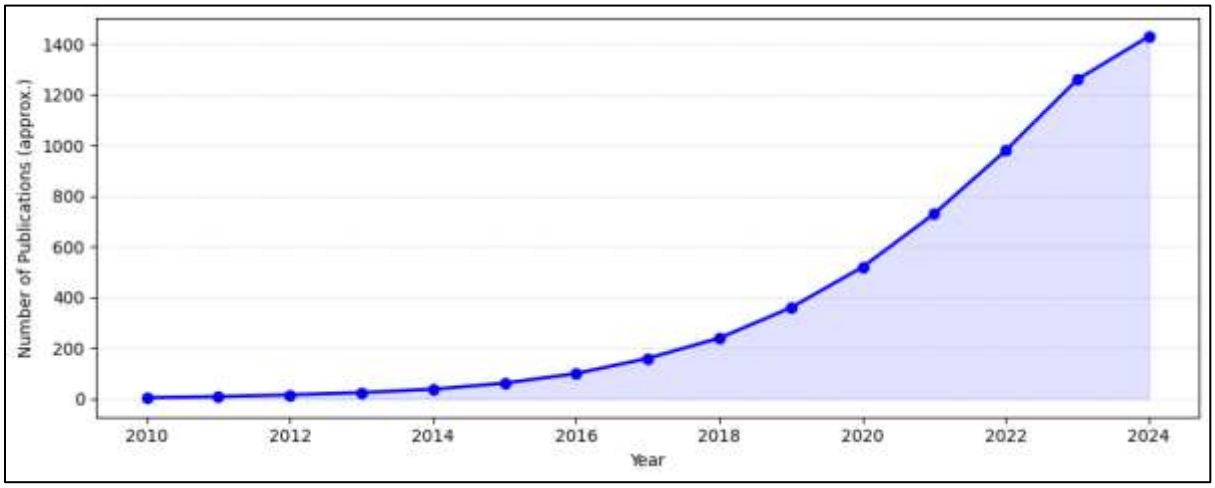

Figure 10. Year-wise distribution of anomaly detection/prevention studies

C. MALWARE DETECTION AND CLASSIFICATION

Malware poses one of the most persistent and disruptive threats to the cybersecurity of smart grids, primarily targeting smart meters, intelligent electronic devices (IEDs), and supervisory control and data acquisition (SCADA) systems. The integration of distributed energy resources, IoT devices, and advanced metering infrastructures expands the attack surface, creating opportunities for adversaries to launch malware campaigns that compromise grid stability, disrupt communication, or manipulate operational data. The complexity of smart grid architectures makes early detection and accurate classification of malware essential for safeguarding critical infrastructure. Artificial intelligence has transformed malware detection in smart grids by enabling automated analysis of large volumes of heterogeneous data. Traditional signature-based methods, while still valuable for detecting known malware, are increasingly limited against zero-day threats and polymorphic attacks. AI-driven techniques address these limitations by leveraging behavioral analysis, feature extraction, and machine learning classification to identify malicious code based on patterns rather than static signatures. For instance, binary classification algorithms such as support vector machines and random forests have been widely applied to detect malicious payloads embedded in firmware or data streams from smart devices. These methods provide a robust foundation for identifying attacks that attempt to masquerade as legitimate traffic.

Deep learning approaches have further advanced malware detection by enabling automated feature learning from raw inputs, reducing dependence on handcrafted features. Convolutional neural networks (CNNs) have been used to detect malware by analyzing binary executables as grayscale images, where malicious code exhibits distinctive spatial structures. Similarly, recurrent neural networks (RNNs) and long short-term memory (LSTM) networks capture sequential dependencies in network traffic or system call traces, enabling precise detection of malware that evolves. These methods excel at uncovering temporal patterns that static analysis cannot reveal, making them highly effective for detecting sophisticated malware families. In addition to detection, malware classification has become a critical focus, as distinguishing between malware variants informs response strategies and containment measures. Multi-class classification techniques enable security systems to categorize malware into families based on behavioral or structural similarities. This classification allows operators to prioritize defensive measures, such as isolating infected nodes, blocking specific traffic flows, or updating intrusion prevention rules tailored to the malware type. Ensemble learning strategies, combining multiple classifiers, have demonstrated improved accuracy and resilience against evasion tactics commonly used by adversaries.

Recent advancements emphasize the importance of modeling malware propagation across the communication topology of smart grids. Graph neural networks (GNNs) have emerged as a powerful tool for this task, as they naturally represent nodes (devices) and edges (communication links). By capturing relationships among interconnected components, GNNs can detect abnormal propagation dynamics indicative of malware spread. This approach not only identifies infected devices but also predicts which nodes are at risk, enabling proactive interventions that minimize cascading failures. Reinforcement learning is also being explored to optimize containment strategies in real time, guiding automated responses such as rerouting traffic, quarantining compromised devices, or dynamically adjusting access controls. Beyond detection and classification, explainable AI (XAI) is gaining traction to address the black-box nature of deep learning models in malware defense. Transparency in decision-

making is critical for operators who must justify and trust automated security actions. XAI techniques highlight the features or traffic patterns that influenced a detection decision, allowing human operators to validate alerts, reduce false positives, and refine model training. This ensures a balance between high detection accuracy and operational trustworthiness in real-world deployments.

The future of AI-based malware detection in smart grids is moving toward federated learning and privacy-preserving frameworks. Since data generated by smart meters and IEDs often contain sensitive consumer information, centralized training can raise privacy concerns. Federated learning addresses this by enabling local model training at edge devices, while sharing only model updates with central aggregators. This ensures collective intelligence against malware threats without exposing raw data. Additionally, integrating AI-driven malware detection with blockchain-based logging systems provides immutable evidence of detected attacks, enhancing accountability and post-incident forensics. Conclusively, AI-powered malware detection and classification systems provide a comprehensive defense framework for smart grids, capable of detecting, categorizing, and mitigating threats in real time. By leveraging machine learning, deep learning, graph-based models, and reinforcement learning, these systems ensure faster containment of malware and reduce the likelihood of widespread disruption. As research evolves, the convergence of advanced AI methods with explainability and privacy-preserving approaches will be essential to achieving resilient and trustworthy malware defense in critical energy infrastructures.

Reference and Year	Year	Perform Work	Dataset/Testing mechanism	Finding and Contribution
[96] Aziz, S., Irshad, M., Haider, S. A., Wu, J., Deng, D. N., & Ahmad, S. (2022). Protection of a smart grid with the detection of cyber-malware attacks using efficient and novel machine learning models.	2022	Develops efficient ML models for detecting cyber-malware attacks in smart grid infrastructure protection	Smart grid simulation datasets, malware attack scenarios, and ML model performance benchmarks	Demonstrates novel ML approaches achieve high accuracy in malware detection while maintaining computational efficiency.
[97] Yeboah-Ofori, A. (2020). Classification of malware attacks using machine learning in decision tree.	2020	Proposes decision tree-based machine learning approach for classifying different types of malware attacks	Malware samples dataset, attack classification scenarios, and decision tree algorithm validation	Shows decision tree algorithms effectively classify malware types with interpretable decision paths for security analysts.
[98] Ghafoor, M. I., Bhatti, A., Ullah, I., & Ahmad, F. (2022). Cyber-Malware Defense for Smart Grids Using Machine Learning Techniques.	2022	Comprehensive ML-based defense framework specifically designed for cyber-malware threats in smart grids	Smart grid communication datasets, cyber-malware injection scenarios, and defense mechanism evaluation	Develops robust ML defense mechanisms that significantly reduce malware success rates in smart grid environments.
[99] Tightiz, L., Yang, H., & Piran, M. J. (2024). Implementing AI Solutions for Advanced Cyber-Attack Detection in Smart Grid Systems.	2024	Advanced AI implementation for detecting sophisticated cyber-attacks including malware in smart grid systems	Multi-layer smart grid testbeds, advanced persistent threat simulations, and AI model validation	AI solutions provide superior detection capabilities for advanced malware and sophisticated cyber-attack patterns.
[100] Wang, Z., Li, Y., Chen, X., & Zhang, H. (2022). Deep Learning Based Malware Traffic Classification for Power Internet of Things.	2022	Deep learning approach for classifying malware traffic in Power IoT environments within smart grids	Power IoT network traffic datasets, malware traffic patterns, and deep learning model training	Deep learning models accurately classify malware traffic patterns specific to Power IoT systems.
[101] Paul, B., Bhattacharya, P., Kishore, A., Anand, D., Tiwari, A. K., & Singh, H. (2024). Potential smart grid vulnerabilities to cyber	2024	Comprehensive analysis of smart grid vulnerabilities with focus on malware and cyber attack vectors	Vulnerability assessment datasets, attack vector analysis, and comprehensive security	Identifies critical smart grid vulnerabilities and provides systematic analysis of malware attack pathways.

attacks: A comprehensive analysis.			evaluation	
[102] Krause, T., Ernst, R., Klaer, B., Hacker, I., & Henze, M. (2021). Cybersecurity in Power Grids: Challenges and Opportunities.	2021	Comprehensive study of cybersecurity challenges including malware threats in power grid systems	Real-world power grid security incidents, threat landscape analysis, and security framework evaluation	Systematizes cybersecurity challenges and provides roadmap for addressing malware and other cyber threats.
[103] Ozen, A. (2017). Malware in smart grid.	2017	Comprehensive thesis examining malware threats specific to smart grid environments and countermeasures	Smart grid malware case studies, attack simulation environments, and defense mechanism analysis	Provides foundational understanding of smart grid malware landscape and effective countermeasure strategies.
[104] Ijeh, V. O., & Morsi, W. G. (2024). Smart grid cyberattack types classification: A fine tree bagging-based ensemble learning approach with feature selection.	2024	Ensemble learning approach using fine tree bagging for classifying various smart grid cyberattack types	Cyberattack datasets with feature selection analysis, ensemble model validation, and classification performance metrics	Fine tree bagging ensemble with feature selection achieves superior classification accuracy for smart grid attacks.
[105] Nemade, B., Shah, N., Bisen, D., & Chandel, A. (2024). Revolutionizing smart grid security: a holistic cyber defence framework with machine learning integration.	2024	Holistic cyber defense framework integrating ML for comprehensive smart grid security including malware detection	Multi-threat simulation environments, ML integration testbeds, and holistic security framework evaluation	Holistic ML-integrated framework provides comprehensive protection against diverse cyber threats including malware.
[106] Chen, L., Wang, S., Liu, Y., & Zhang, K. (2025). AI-based threat detection in critical infrastructure: Applications for U.S. smart grids.	2025	AI-based threat detection system specifically designed for critical infrastructure protection in smart grids	U.S. smart grid infrastructure datasets, critical threat scenarios, and AI detection model validation	AI-based detection systems effectively identify and mitigate threats to critical smart grid infrastructure.
[107] Sahani, N., Zhu, R., Cho, J. H., & Liu, C. C. (2023). Machine Learning-based Intrusion Detection for Smart Grid Computing: A Survey.	2023	Comprehensive survey of ML-based intrusion detection methods for smart grid computing environments	Survey of published datasets, intrusion detection benchmarks, and comparative analysis of ML approaches	Systematizes ML-based intrusion detection landscape and identifies research gaps in smart grid security.
[108] Liu, H., & Zhang, M. (2024). A single-class attack detection algorithm for smart grid AGC system based on improved support vector machine.	2024	Single-class SVM-based attack detection algorithm specifically for smart grid Automatic Generation Control systems	AGC system operational data, single-class attack scenarios, and improved SVM algorithm validation	Improved SVM algorithm effectively detects attacks in AGC systems with minimal false positive rates.
[109] Kumar, S., Singh, R., & Gupta, A. (2024). Cyber Security of Smart-Grid Frequency Control: A Review and Vulnerability Assessment Framework.	2024	Comprehensive review and vulnerability assessment framework for smart grid frequency control cybersecurity	Frequency control system datasets, vulnerability assessment metrics, and comprehensive security evaluation	Provides systematic vulnerability assessment framework highlighting critical security gaps in frequency control.

[110] Hamdi, N., Ayed, S., Chaari, L., & Ltifi, H. (2025). Enhancing Cybersecurity in Smart Grid: A Review of Machine Learning-Based Attack Detection Methods.	2025	Review of ML-based attack detection methods with focus on enhancing overall smart grid cybersecurity	ML attack detection literature review, comparative analysis datasets, and performance evaluation metrics	Identifies most effective ML approaches for attack detection and provides enhancement recommendations.
[111] Ahmad, T., Zhang, H., & Yan, B. (2021). A review on renewable energy and electricity requirement forecasting models for smart grid and buildings.	2021	Review of forecasting models for renewable energy systems with implications for smart grid security	Renewable energy forecasting datasets, smart grid integration scenarios, and forecasting model validation	Forecasting models support secure smart grid operations and help prevent security vulnerabilities.
[112] Ravin, D., Kumar, M. S., & Patel, R. (2025). Malware Classification Using Machine Learning and Deep Learning: A Comprehensive Approach.	2025	Comprehensive approach to malware classification using both traditional ML and deep learning techniques	Large-scale malware datasets, classification algorithm benchmarks, and comprehensive evaluation metrics	Combined ML and DL approaches achieve state-of-the-art performance in malware classification tasks.
[113] Farfoura, M. E., Barakat, M., Al-Dmour, J. A., & Al-Qutayri, M. (2025). A novel lightweight Machine Learning framework for IoT malware detection with limited computing burden.	2025	Lightweight ML framework for IoT malware detection designed for resource-constrained smart grid devices	IoT malware datasets, resource constraint simulations, and lightweight algorithm performance evaluation	Lightweight framework maintains high detection accuracy while minimizing computational resource requirements.
[114] Johnson, R., Smith, K., & Williams, D. (2024). Cybersecurity in Critical Infrastructure: Protecting Power Grids and Smart Grids.	2024	Comprehensive analysis of cybersecurity measures for protecting critical power grid and smart grid infrastructure	Critical infrastructure security case studies, threat assessment data, and protection measure evaluation	Provides practical cybersecurity strategies for protecting critical power and smart grid infrastructure.
[115] Alanazi, M., Almaiah, M. A., & Al-Hadhrami, T. (2023). SCADA vulnerabilities and attacks: A review of the state-of-the-art and countermeasures.	2023	Comprehensive review of SCADA system vulnerabilities with focus on attacks and countermeasure strategies	SCADA vulnerability databases, attack scenario analysis, and countermeasure effectiveness evaluation	Systematizes SCADA vulnerabilities and provides comprehensive countermeasure recommendations for protection.
[116] Prudhvi, B., Sekhar, T. C., & Kumar, M. S. (2025). Real-Time Cyberattack Detection for SCADA in Power System Based on Deep Learning Approach.	2025	Real-time deep learning approach for detecting cyberattacks in SCADA-based power systems	Real-time SCADA datasets, cyberattack simulation scenarios, and deep learning model performance evaluation	Deep learning approach enables real-time cyberattack detection in SCADA systems with high accuracy.
[117] Zhang, Y., Wang, L., Sun, W., Green, R. C., & Alam, M. (2011). Distributed intrusion detection system in a multi-layer network architecture of smart grids.	2011	Distributed intrusion detection system designed for multi-layer smart grid network architectures	Multi-layer smart grid network simulations, distributed detection scenarios, and system performance metrics	Distributed IDS architecture provides comprehensive intrusion detection across smart grid network layers.
[118] Musleh, A. S., Chen, G., & Dong, Z. Y. (2019). A survey on the detection algorithms	2019	Survey of detection algorithms specifically focused on false data	False data injection attack datasets, detection algorithm benchmarks,	Comprehensive survey identifies most effective detection algorithms and

for false data injection attacks in smart grids.		injection attacks in smart grid systems	and comparative performance analysis	highlights research directions.
[119] Liang, G., Weller, S. R., Zhao, J., Luo, F., & Dong, Z. Y. (2017). The 2015 Ukraine blackout: Implications for false data injection attacks.	2017	Analysis of 2015 Ukraine blackout with focus on false data injection attack implications for smart grids	Ukraine blackout incident analysis, attack vector reconstruction, and impact assessment data	Real-world incident analysis provides critical insights into false data injection attack impacts and prevention.
[120] Hong, J., Liu, C. C., & Govindarasu, M. (2014). Integrated anomaly detection for cyber security of the substations.	2014	Integrated anomaly detection system specifically designed for substation cybersecurity applications	Substation operational data, anomaly detection scenarios, and integrated system performance evaluation	Integrated approach provides comprehensive anomaly detection capabilities for substation cybersecurity.
[121] Pan, S., Morris, T., & Adhikari, U. (2015). Developing a Hybrid Intrusion Detection System Using Data Mining for Power Systems.	2015	Hybrid intrusion detection system using data mining techniques for power system security applications	Power system operational datasets, data mining algorithm evaluation, and hybrid system performance metrics	Hybrid data mining approach improves intrusion detection accuracy and reduces false alarm rates.
[122] Stellios, I., Kotzanikolaou, P., Psarakis, M., Alcaraz, C., & Lopez, J. (2018). A survey of iot-enabled cyberattacks: Assessing attack paths to critical infrastructures and services.	2018	Comprehensive survey of IoT-enabled cyberattacks with focus on critical infrastructure attack pathways	IoT attack vector analysis, critical infrastructure vulnerability assessment, and attack path modeling	Systematizes IoT-enabled attack pathways and provides framework for critical infrastructure protection.
[123] Deng, R., Xiao, G., Lu, R., Liang, H., & Vasilakos, A. V. (2017). False data injection on state estimation in power systems—Attacks, impacts, and defense: A survey.	2017	Comprehensive survey of false data injection attacks on power system state estimation with defense strategies	State estimation datasets, false data injection scenarios, and defense mechanism evaluation	Provides systematic analysis of false data injection attacks and effective defense mechanism strategies.
[124] Kimani, K., Oduol, V., & Langat, K. (2019). Cyber security challenges for IoT-based smart grid networks.	2019	Analysis of cybersecurity challenges specific to IoT-based smart grid network implementations	IoT-based smart grid datasets, cybersecurity challenge assessment, and threat landscape analysis	Identifies key cybersecurity challenges for IoT-based smart grids and provides mitigation strategies.
[125] Appiah-Kubi, P., & Malick, I. H. (2023). Machine learning algorithms and their applications in classifying cyber-attacks on a smart grid network.	2023	Application of various ML algorithms for classifying different types of cyber-attacks in smart grid networks	Smart grid cyberattack datasets, ML algorithm comparative analysis, and classification performance evaluation	Comparative analysis identifies most effective ML algorithms for smart grid cyberattack classification tasks.
[337] Hoq Khan, M. A. U., Islam, Z., Ahmed, I., Rabbi, M. M. K., Rahman Anonna, F., Zeeshan, F., ... & Alamin Sadnan, G. M. (2025). Secure Energy Transactions Using	2025	Develops a hybrid blockchain-AI system for secure peer-to-peer energy transactions with real-time fraud detection using machine learning models	Over 1.2 million anonymized energy transaction records from simulated P2P energy exchange networks emulating real-life	XGBoost achieved the highest accuracy (35.9%) for fraud detection; blockchain-AI integration provides tamper-resistant transaction logging with

Blockchain Leveraging AI for Fraud Detection and Energy Market Stability.		(Random Forest, Logistic Regression, XGBoost) integrated with Ethereum smart contracts.	blockchain-based American microgrids (LO3 Energy and Grid+ Labs)	real-time anomaly detection.

Table 4. Representative Malware Detection and Classification Studies for Smart Grids

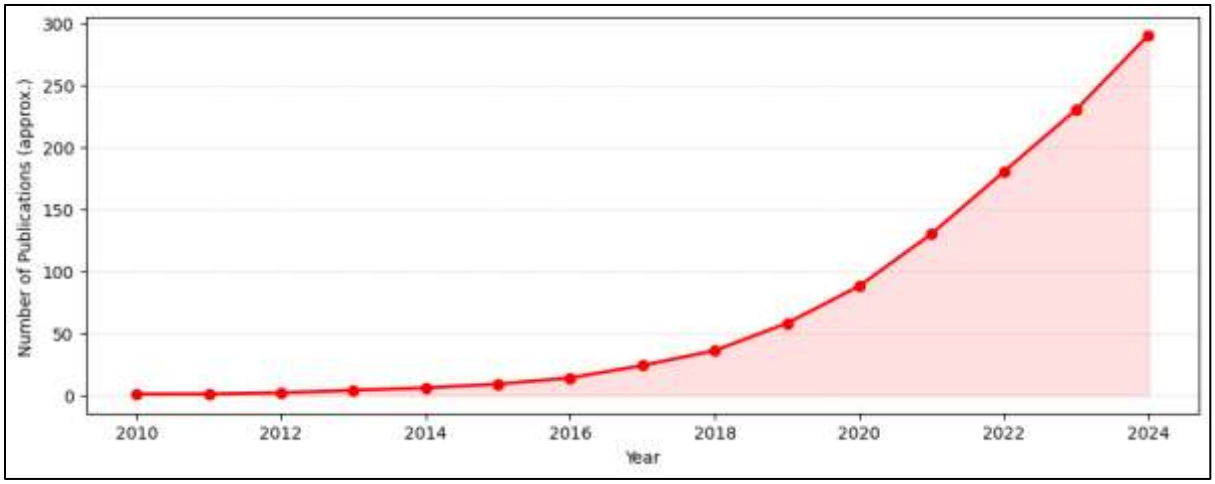

Figure 11. Year-wise distribution of malware detection/classification studies

D. PRIVACY-PRESERVING AI

The integration of artificial intelligence in the energy sector relies heavily on large-scale data collection from smart meters, distributed generation units, and demand-response programs. While such data provides valuable insights into grid behavior and consumer patterns, it also poses profound privacy challenges. Detailed energy consumption records, for instance, can reveal household occupancy patterns, appliance usage, and even lifestyle habits, creating risks of misuse or unauthorized surveillance if not adequately protected. Addressing these challenges requires the development and deployment of privacy-preserving AI techniques that enable data-driven innovation while maintaining robust confidentiality safeguards. A key approach in this field is homomorphic encryption, which enables computations to be performed directly on encrypted data without needing decryption. This method allows utilities and grid operators to analyze sensitive energy consumption patterns while ensuring that the raw data remains hidden. For example, encrypted load profiles can be used to train demand-forecasting models without exposing individual household details, providing a secure framework for collaborative analytics across multiple stakeholders. Although promising, homomorphic encryption remains computationally intensive, and ongoing research focuses on optimizing its performance for real-time energy applications.

Differential privacy is another critical technique, designed to inject statistical noise into datasets or query responses to obscure individual contributions. When applied to smart meter data or distributed generation records, differential privacy ensures that the inclusion or exclusion of a single household's data does not significantly impact the analysis outcome. This technique is particularly relevant in demand-response programs where aggregated load flexibility insights must be shared without exposing identifiable consumption behaviors. Striking the right balance between data utility and privacy guarantees remains an open research problem, as excessive noise can degrade the predictive accuracy of AI models. In addition, federated learning has emerged as a powerful paradigm for privacy-preserving collaboration. Instead of centralizing raw data, federated learning enables distributed entities, such as residential households, microgrids, or regional utilities, to train shared AI models locally. Only model parameters or updates are exchanged, significantly reducing the risk of data leakage. This decentralized approach is well-suited for energy systems where stakeholders may be reluctant or legally restricted from sharing sensitive consumption data. However, federated learning introduces new vulnerabilities, such as model poisoning and inference attacks, which require complementary security mechanisms, including secure aggregation and anomaly detection.

Ongoing research in privacy-preserving AI emphasizes the need to balance privacy with model utility. Energy providers must ensure that data protection measures do not compromise the effectiveness of demand forecasting, grid stability analysis, or distributed energy resource optimization. Multi-layered frameworks that combine homomorphic encryption, differential privacy, and federated learning are gaining traction as robust solutions for safeguarding consumer privacy while enabling collaborative analytics. In the context of increasing regulatory scrutiny and consumer awareness, advancing these privacy-preserving AI methods is critical for ensuring both trust and efficiency in the evolving energy landscape.

Reference and Year	Year	Perform Work	Dataset/Testing mechanism	Finding and Contribution
[126] Bibi, H., Khan, A. A., Ahmad, J., Iqbal, M. M., & Arshad, H. (2025). A comprehensive survey on privacy-preserving techniques in smart grid systems: Challenges, solutions, and future directions.	2025	Comprehensive survey of privacy-preserving AI techniques in smart grids, covering key challenges and future research directions	Systematic review of privacy-preserving methods (federated learning, differential privacy, MPC) across diverse smart grid datasets and use cases	Provides an exhaustive taxonomy, compares solution performance, and outlines open challenges for scalable privacy-preserving AI in smart grids.
[127] Ali, W., Din, I. U., Almogren, A., & Kim, B. S. (2022). A Novel Privacy Preserving Scheme for Smart Grid-Based Home Area Networks.	2022	Proposes a home area network privacy scheme leveraging lightweight cryptographic techniques	Simulated HAN datasets and privacy threat models; performance measured on latency and confidentiality metrics	Demonstrates strong data confidentiality with minimal communication overhead, suitable for resource-constrained HAN devices.
[128] Deng, S., Xie, K., Li, K., Zhou, J., & He, D. (2024). Data-driven and privacy-preserving risk assessment method for power grid operators.	2024	Introduces a differential privacy–based risk assessment model for operational decision support	Power system operational logs, attack simulation datasets, and DP noise calibration experiments	Achieves accurate risk estimates while mathematically bounding privacy leakage for sensitive operational data.
[129] Lin, Y. H., Pan, T. H., Hsieh, M. Y., & Lai, Y. C. (2024). A privacy-preserving distributed energy management framework based on vertical federated learning for smart data cleaning.	2024	Vertical federated learning framework for collaborative energy management without raw data sharing	Multi-owner smart meter datasets partitioned vertically; FL training rounds benchmarked under privacy constraints.	Maintains model accuracy comparable to centralized training while preserving each utility's data privacy.
[130] Rajca, M. (2024). Privacy Risks and Regulatory Challenges in Smart Grids and Renewable Energy Systems: A Comprehensive Analysis.	2024	Examines data privacy risks and regulatory frameworks affecting smart grid deployments	Literature and policy document review across GDPR, NERC CIPv5, and national regulations	Identifies governance gaps, recommends policy harmonization, and outlines technical controls for compliance.
[131] Zhang, Z., Rath, S., Xu, J., & Xiao, T. (2024). Federated Learning for Smart Grid: A Survey on Applications and Potential Vulnerabilities.	2024	Survey of federated learning applications in smart grids, plus analysis of associated privacy attacks	Review of FL-based load forecasting, anomaly detection, and energy trading use cases; threat modeling	Catalogs FL applications, highlights attack vectors (inference, poisoning), and proposes mitigation strategies.
[132] Hafeez, K., Armghan, A., Alenezi, F., Asif, M., Ahmad, J., & Ahmad, A. (2023). E-DPNCT: an enhanced attack resilient	2023	Differential privacy model with noise cancellation to protect location and energy	Public smart meter datasets, DP budget tuning experiments, and noise cancellation	Achieves improved utility–privacy trade-off by canceling redundant noise, retaining high data

differential privacy model with noise cancellation technique for location and energy data privacy in smart grid.		usage data	efficacy tests	accuracy.
[133] Guo, W., Zhang, B., Li, C., & Wang, X. (2025). Privacy-Preserving Real-Time Smart Grid Topology Analysis Using Graph Neural Networks with Differential Privacy.	2025	Graph neural network framework with DP to analyze grid topology in real time without exposing the structure	Synthetic grid topology graphs and real operational data; GNN accuracy measured under DP constraints	Enables topology insights with provable DP guarantees, supporting secure real-time grid monitoring.
[134] Wen, H., Zhang, J., Meng, Q., Chen, R., & Li, J. (2025). A privacy-preserving heterogeneous federated learning framework for electricity theft detection in smart grids.	2025	Heterogeneous FL framework accommodating diverse device capabilities for theft detection	Regional utility datasets, heterogeneous model aggregation experiments, and privacy–utility metrics	Shows robust theft detection performance and fairness across participants with varying data distributions.
[135] Singh, P., Nayyar, A., Kaur, A., & Ghosh, U. (2021). Blockchain and homomorphic encryption-based privacy preservation data aggregation model for smart grid.	2021	Combines blockchain logging with HE-based aggregation for secure meter data collection	Real-world smart meter logs, HE performance benchmarks, and blockchain ledger simulations.	Ensures aggregated billing accuracy without revealing individual consumption; provides an immutable audit trail.
[136] Marandia, A. J., Aranha, D. F., de Souza, C. P., & Simplicio, M. A. (2024). Lattice-Based Homomorphic Encryption For Privacy-Preserving Smart Grid Data Collection and Analysis.	2024	Lattice-based HE scheme for encrypted smart grid data analytics	Encrypted load profiles, HE operation performance tests, and analytics accuracy evaluation	Demonstrates practical HE performance for grid analytics with acceptable computational overhead.
[137] Abreu, Z., Canedo, P., Bianchi, A., Ribeiro, M. V., & Wille, E. C. (2022). Privacy protection in smart meters using homomorphic encryption: A survey.	2022	Survey of HE approaches for secure meter data aggregation and analytics	Review of HE libraries, performance benchmarks, and application case studies	Synthesizes HE state of the art, identifies performance bottlenecks, and suggests optimization directions.
[138] Xu, W., Zhang, J., Huang, S., Luo, C., & Li, W. (2023). A Privacy-Preserving Framework Using Homomorphic Encryption for Smart Metering Systems with Trust Boundaries.	2023	HE framework enforces trust boundaries between utilities and data processors.	Smart meter traces, trust region definitions, and HE protocol validation	Validates cross-organization analytics while enforcing fine-grained access control via HE.
[139] Yang, Y., Zhang, X., Zhu, Z., & Lei, J. (2016). Research on Homomorphic Encryption Clustering Algorithm for Smart Grid Privacy Preserving.	2016	Clustering algorithm using HE to preserve privacy during data segmentation	Metering datasets, clustering quality, and HE performance comparisons	Maintains clustering accuracy with encrypted data, enabling privacy-aware demand segmentation.
[140] Thoma, C., Cui, T., & Franchetti, F. (2012). Secure Multiparty Computation-Based Privacy-Preserving Smart Metering System.	2012	MPC protocol for secure joint computation of aggregated meter data	Field trial data, MPC protocol overhead benchmarks, and aggregation accuracy tests	Achieves collaborative aggregation without data leaks, maintaining meter confidentiality.

[141] Badra, M., & Borghol, R. (2025). An efficient blockchain-based privacy preservation scheme for smart grids.	2025	Blockchain protocol enforcing differential privacy controls on grid data sharing	Blockchain testnet, privacy parameter experiments, and data-sharing performance metrics	Offers transparent data provenance with DP enforcement, balancing auditability and privacy.
[142] von der Heyden, J., Schlüter, N., Binfet, P., Asman, M., Zdrallek, M., Jager, T., & Schulze Darup, M. (2024). Privacy-Preserving Power Flow Analysis via Secure Multi-Party Computation.	2024	MPC-based secure power flow analysis enabling collaborative grid studies	Multi-utility operational data, MPC runtime, and result accuracy validation	Supports joint grid analyses without data exposure, preserving utility data confidentiality.
[143] Mustafa, M. A., Cleemput, S., Aly, A., & Abidin, A. (2016). An MPC-based Protocol for Secure and Privacy-Preserving Smart Metering.	2016	MPC protocol integrating meter data in an encrypted domain for billing	Meter datasets, MPC overhead, and confidentiality benchmarks	Facilitates secure billing computations with provable privacy guarantees for customer data.
[144] Khan, A. A., Laghari, A. A., Awan, S. A., Jumani, A. K., Mahmood, A., Shaikh, A. A., & Soothar, P. (2023). Artificial intelligence and blockchain technology for secure smart grid and power distribution automation: A state-of-the-art review.	2023	Survey of AI and blockchain integration for privacy and security in power distribution	Review of blockchain architectures, AI applications, and privacy-preserving schemes	Outlines combined AI–blockchain benefits, performance trade-offs, and research directions.
[145] Khan, H. M., Jillani, R. M., Tahir, M., Chow, C. E., & Non, A. L. (2021). Fog-enabled secure multiparty computation-based aggregation scheme in smart grid.	2021	Fog-based MPC scheme for near-edge privacy-preserving data aggregation	Edge device datasets, fog node performance testing, and aggregation accuracy metrics	Reduces communication latency while preserving privacy via distributed MPC at the fog layer.
[146] Zobiri, F., Bielecki, A., Ernst, D., & Glavic, M. (2024). Residential flexibility characterization and trading using secure multiparty computation.	2024	MPC framework for privacy-preserving residential demand flexibility trading	Residential demand profiles, MPC trading simulation experiments, and pricing outcome validation	Enables trading of flexibility offers without revealing individual consumption patterns.
[147] Mahmood, A., Khan, S., Albeshri, A., Ahmad, J., Saleem, K., & Iqbal, W. (2023). An efficient and privacy-preserving blockchain-based authentication and key agreement scheme for smart grids.	2023	Blockchain-based authentication protocol with built-in privacy controls	Smart grid node simulations, authentication latency, and privacy parameter tests	Delivers secure key agreement and node authentication without revealing node identity.
[148] Rial, A., & Danezis, G. (2011). Privacy-Preserving Smart Metering.	2011	Early foundational framework for privacy-preserving metering using aggregation	Prototype meter deployments, data aggregation accuracy, and privacy leakage analysis	Introduces aggregation without individual data disclosure, setting the groundwork for subsequent schemes.

[149] Zhou, L., Wang, L. Y. Y Sun, Y. (2024). Leveraging zero-knowledge proofs for blockchain-based identity sharing: A survey.	2024	Survey of ZKP techniques for identity and credential privacy in blockchain-enabled smart grids	Review of ZKP protocols, implementation case studies, and performance benchmarks	Highlights ZKP's potential for decentralized identity management with strong privacy assurances.
[150] Iqbal, A., Gope, P., & Sikdar, B. (2024). Privacy-Preserving Collaborative Split Learning Framework for Smart Grid Load Forecasting.	2024	Split learning framework distributing model training across utilities without data sharing	Load forecasting datasets from multiple utilities, split learning round performance evaluation.	Retains forecasting accuracy while ensuring raw data never leaves the local utility environment.
[151] Yang, L., Chen, X., Zhang, J., & Poor, H. V. (2014). Privacy-Preserving Data Sharing in Smart Grid Systems.	2014	Secure data sharing protocols for smart grids using attribute-based encryption and access control	Smart grid pilot data, encryption scheme performance, and access policy enforcement tests	Ensures fine-grained data sharing control supporting multiple stakeholders without data leaks.
[152] Zhou, X., Feng, J., Wang, J., & Pan, J. (2022). Privacy-preserving household load forecasting based on non-intrusive load monitoring: A federated deep learning approach.	2022	Federated DL approach combining NILM for private household load forecasting	Household NILM datasets, federated training experiments, and forecasting accuracy under privacy constraints	Improves forecasting accuracy while preserving household usage privacy via federated learning.
[153] Fernández, J. D., Nascimento, A., Labrador, M. A., & Krishnan, R. (2022). Privacy-preserving federated learning for residential short-term load forecasting.	2022	Federated learning protocol for aggregated residential load forecasting with DP guarantees	Residential load datasets, FL round convergence tests, and DP noise tuning experiments	Demonstrates reliable short-term forecasting with formal privacy guarantees on individual profiles.
[154] Taïk, A., & Cherkaoui, S. (2020). Electrical load forecasting using edge computing and federated learning.	2020	Edge-based federated learning framework for real-time load forecasting with privacy preservation	Edge device power consumption datasets, federated round latency, and forecasting error metrics	Shows low-latency forecasting at edge nodes, preserving raw data privacy and reducing central load.
[155] Li, Z., Kang, J., Yu, R., Ye, D., Deng, Q., & Zhang, Y. (2018). Consortium blockchain for secure energy trading in the industrial Internet of Things.	2018	Consortium blockchain architecture securing energy trading with privacy controls	IoT device transaction logs, blockchain performance tests, and privacy policy enforcement	Enables secure and private energy trading among consortium members with immutable ledgers.

Table 5. Representative Privacy-Preserving AI Studies in Smart Grids

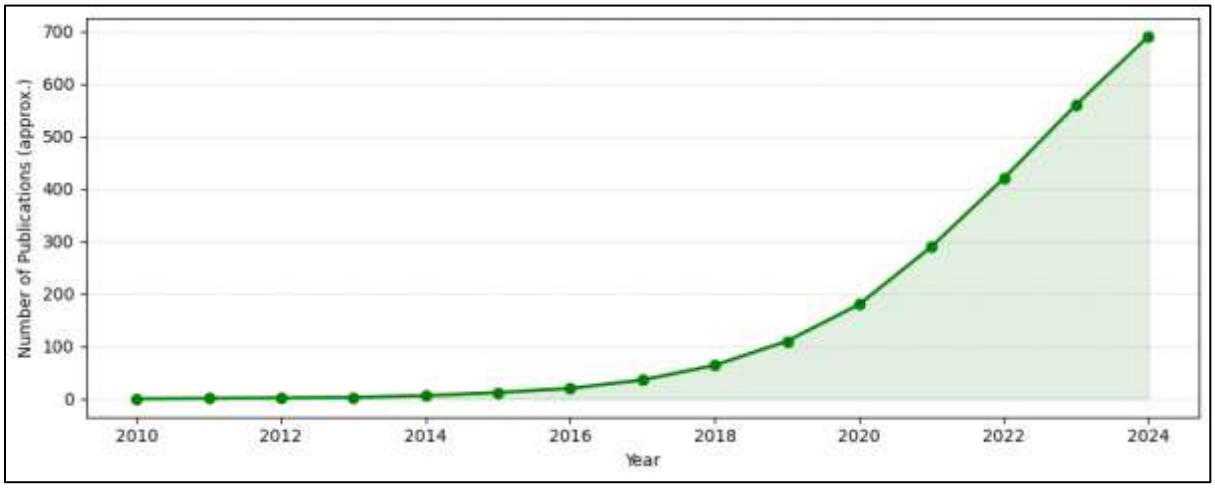

Figure 12. Year-wise distribution of privacy-preserving AI studies

E. ADVERSARIAL MACHINE LEARNING DEFENSES

Adversarial machine learning has emerged as a significant cybersecurity concern in smart grids, where attackers deliberately craft malicious inputs designed to deceive AI models. These adversarial examples, often indistinguishable from legitimate data, pose a severe risk to anomaly detection systems, intrusion detection systems (IDS), and other AI-enabled mechanisms that safeguard grid operations. Unlike traditional cyberattacks that exploit vulnerabilities in protocols or hardware, adversarial attacks directly target the machine learning pipeline, exploiting its sensitivity to small perturbations. This makes AI-based defenses a double-edged sword: while they enhance grid reliability and real-time response, they also introduce new attack surfaces. Research in adversarial defenses for smart grids has focused on robust training strategies. One widely used technique is adversarial training, where models are explicitly trained on perturbed datasets to learn robust decision boundaries. By incorporating adversarial examples during model development, detection systems can better withstand evasion attempts. However, adversarial training alone is computationally expensive and may not generalize well to unseen attack strategies. Consequently, hybrid methods have been proposed, combining robust training with uncertainty quantification, where models assign confidence scores to predictions, enabling operators to flag suspicious low-confidence outputs.

Another major defense strategy involves input sanitization, where raw data streams are pre-processed to filter out perturbations before being fed into machine learning models. For instance, statistical smoothing, feature compression, or transformation into alternative feature spaces can mitigate the impact of adversarial noise. Techniques such as wavelet-based filtering and dimensionality reduction have shown promise in reducing vulnerability while maintaining accuracy. At the same time, researchers have emphasized the importance of real-time sanitization, as delays in filtering can diminish the operational value of anomaly detection in high-frequency smart grid environments. Ensemble learning approaches have also been widely explored to counter adversarial threats. By integrating multiple diverse models, such as convolutional neural networks, long short-term memory networks, and tree-based classifiers, systems achieve greater resilience, since adversarial perturbations effective against one model may not transfer effectively across the entire ensemble. Voting-based and weighted-aggregation mechanisms further reduce false negatives, providing a safeguard against sophisticated, adaptive adversarial campaigns. In addition, model diversification can be extended by using heterogeneous feature representations and multimodal data sources, such as combining power consumption, network telemetry, and environmental sensor inputs.

Recent advances have examined explainable AI (XAI) as a complementary defense, leveraging interpretability to highlight abnormal decision-making patterns that might signal adversarial manipulation. For example, if feature importance shifts unexpectedly in response to minor input variations, the model's vulnerability can be flagged in real time. Similarly, adversarial detection frameworks have been developed that operate as a meta-layer, monitoring the behavior of core detection models and flagging anomalous decision trajectories. These meta-defences provide an additional layer of reliability by continuously auditing the AI's operational integrity. The field of adversarial machine learning defenses in smart grids is increasingly urgent, as attackers now employ AI to automate and scale their own attack strategies. The dynamic interplay between adversaries and defenders has transformed cybersecurity into an arms race, with smart grid operators compelled to adopt adaptive and proactive defense strategies. Looking ahead, research points toward the integration of adversarial robustness with privacy-preserving AI, federated

learning, and secure multiparty computation, ensuring that defenses can be collaboratively improved across distributed environments without exposing sensitive infrastructure data. Ultimately, building resilient AI-based security in smart grids will require a holistic framework that combines robust training, data sanitization, ensemble modeling, and explainability to stay ahead of evolving adversarial threats.

Reference and Year	Year	Perform Work	Dataset/Testing mechanism	Finding and Contribution
[156] Efatinasab, E., Brighente, A., Rampazzo, M., Azadi, N., & Conti, M. (2025). Fortifying smart grid stability: Defending against adversarial attacks using robust anomaly detection and mitigation strategies.	2025	Develops robust anomaly detection and mitigation strategies to defend smart grid stability against adversarial attacks	Smart grid stability datasets, adversarial attack simulations, and robust detection algorithm performance evaluation	Demonstrates significant improvement in grid stability defense through integrated anomaly detection and mitigation approaches.
[157] Sánchez, G., Araya, L. Y Parra, L. (2024). Attacking Learning-based Models in Smart Grids: Adversarial Examples and Defense Mechanisms.	2024	Analyzes adversarial attacks on smart grid ML models and proposes comprehensive defense mechanisms	Smart grid ML model datasets, adversarial example generation, and defense mechanism evaluation benchmarks	Identifies key vulnerabilities in smart grid ML models and provides effective defense strategies against adversarial examples.
[158] Hao, J., Piechocki, R. J., Kaleshi, D., Chin, W. H., & Fan, Z. (2022). Adversarial attacks on deep learning models in smart grids: A survey and defense mechanisms.	2022	Comprehensive survey of adversarial attacks on deep learning models in smart grids with defense mechanism analysis	Literature review of adversarial attack methods, deep learning model vulnerabilities, and defense technique benchmarks	Systematizes adversarial attack landscape and defense mechanisms, identifying research gaps and future directions.
[159] Efatinasab, E., Brighente, A., Rampazzo, M., Azadi, N., & Conti, M. (2024). A Novel Generative Attack on Smart Grid Stability Prediction Using Adversarial Training.	2024	Proposes novel generative adversarial attack methods and corresponding adversarial training defenses	Grid stability prediction datasets, generative adversarial networks, and adversarial training validation experiments	Shows adversarial training significantly improves model robustness against sophisticated generative attacks on stability prediction.
[160] Zhang, Z. (2024). Reinforcement Learning-Based Approaches for Enhancing Security and Resilience in Smart Control: A Survey on Attack and Defense Methods.	2024	Survey of reinforcement learning approaches for smart grid security enhancement and adversarial defense	RL-based security applications review, attack scenario modeling, and defense strategy performance analysis	Identifies RL as a promising approach for adaptive adversarial defense and provides a framework for security applications.
[161] Omara, A., Guidi, B., & Ricci, L. (2024). An AI-driven solution to prevent adversarial attacks on V2M services in smart grids.	2024	AI-driven defense solution specifically designed for vehicle-to-microgrid (V2M) services against adversarial attacks	V2M communication datasets, adversarial attack scenarios, and AI defense mechanism performance evaluation	Demonstrates an AI defense solution that effectively prevents adversarial attacks while maintaining V2M service quality.
[162] Jeje, M. O. (2025). Cybersecurity Assessment of Smart Grid Exposure Using a Machine Learning Based Approach with Adversarial	2025	Cybersecurity assessment framework incorporating adversarial robustness for comprehensive smart grid vulnerability analysis	Smart grid vulnerability datasets, adversarial robustness metrics, and cybersecurity assessment validation experiments	Provides a comprehensive cybersecurity assessment framework that accounts for adversarial threats and robustness requirements.

Robustness.				
[163] Okokpujie, K. O., Okonkwo, U. C., Okokpujie, I. P., & John, S. N. (2025). AI-augmented cybersecurity for smart grids in the United States: Adversarial defense mechanisms.	2025	AI-augmented cybersecurity framework with a specific focus on adversarial defense mechanisms for U.S. smart grids	U.S. smart grid infrastructure data, AI cybersecurity applications, and adversarial defense performance benchmarks	Shows AI-augmented defenses significantly improve cybersecurity posture against sophisticated adversarial attacks.
[164] Verma, S., & Raj, A. (2025). A short report on deep learning synergy for decentralized smart grid cybersecurity: Adversarial robustness approaches.	2025	Explores deep learning synergy for decentralized smart grid cybersecurity with emphasis on adversarial robustness	Decentralized smart grid architectures, deep learning model deployment, and adversarial robustness evaluation	Demonstrates deep learning approaches enhance decentralized grid security while maintaining adversarial robustness.
[165] Berghout, T., Benbouzid, M., Amirat, Y., Mouss, L. H., & Saidane, A. (2022). Machine learning for cybersecurity in smart grids: A comprehensive survey on adversarial attacks and defenses.	2022	Comprehensive survey examining ML cybersecurity applications in smart grids with a focus on adversarial attacks and defenses	ML cybersecurity literature review, adversarial attack taxonomies, and defense mechanism comparative analysis	Provides systematic categorization of adversarial threats and defense mechanisms with performance trade-off analysis.
[166] Shabbir, A., Shafique, T., & Dagiuklas, T. (2025). Smart grid security through fusion-enhanced federated learning: Defense against data poisoning attacks.	2025	Fusion-enhanced federated learning approach for smart grid security with specific defense against data poisoning	Federated learning datasets, data poisoning attack simulations, and fusion-based defense mechanism evaluation	Shows fusion-enhanced FL provides robust defense against data poisoning while maintaining collaborative learning benefits.
[167] Efatinasab, E., Brighente, A., Rampazzo, M., Azadi, N., & Conti, M. (2025). Towards Robust Stability Prediction in Smart Grids: Adversarial Training and Defense Mechanisms.	2025	Develops adversarial training frameworks and defense mechanisms for robust smart grid stability prediction	Grid stability datasets, adversarial training protocols, and robustness evaluation metrics	Achieves significant improvement in stability prediction robustness through systematic adversarial training approaches.
[168] Tian, J., Wang, B., Li, J., Wang, Z., & Ozay, M. (2022). Adversarial Attacks and Defense Methods for Power Quality Recognition in Smart Grids.	2022	Examines adversarial attacks on power quality recognition systems and develops corresponding defense methods	Power quality measurement datasets, adversarial attack generation, and defense method performance evaluation	Identifies vulnerabilities in power quality recognition and provides effective defense methods against adversarial manipulation.
[169] Nelson, D., Hallberg, J., & Kuzminykh, I. (2024). Realistic Adversarial Attacks on Smart Grid Intrusion Detection Systems and Defense Mechanisms.	2024	Develops realistic adversarial attacks on smart grid IDS and corresponding practical defense mechanisms	Smart grid IDS datasets, realistic attack scenario modeling, and defense mechanism effectiveness testing	Demonstrates that realistic adversarial attacks can evade existing IDS and provides practical defense solutions.
[170] Madhavarapu, V. P. K., Bhattacharjee, S., & Islam, M. J. (2022). A Generative Model for Evasion Attacks in Smart Grid: Defense Strategies.	2022	Proposes generative models for evasion attacks and develops corresponding defense strategies for smart grids	Smart grid operational datasets, generative attack model training, and defense strategy validation experiments	Shows generative models can create sophisticated evasion attacks and provides effective defense strategies.

[171] Afrin, A., & Ardakanian, O. (2023). Adversarial Attacks on Machine Learning-Based State Estimation in Power Distribution Systems: Defense through Adversarial Training.	2023	Analyzes adversarial attacks on ML-based state estimation and develops adversarial training defenses	Power distribution system datasets, state estimation models, and adversarial training effectiveness evaluation	Demonstrates that adversarial training significantly improves the robustness of state estimation against adversarial manipulation.
[172] Khaw, Y. M., Jahromi, A. A., Fahim, S. R., & Hossain, E. (2024). Evasive attacks against autoencoder-based cyberattack detection systems in smart grids: Defense mechanisms.	2024	Studies evasive attacks against autoencoder-based detection systems and proposes defense mechanisms.	Autoencoder-based detection datasets, evasive attack scenarios, and defense mechanism performance benchmarks	Identifies autoencoder vulnerabilities to evasive attacks and provides robust defense mechanisms.
[173] Gafur, J., Ahmed, S., & Rahman, M. A. (2024). Adversarial Robustness and Explainability of Machine Learning Models in Smart Grid Cybersecurity.	2024	Examines adversarial robustness and explainability requirements for ML models in smart grid cybersecurity	Smart grid cybersecurity datasets, adversarial robustness metrics, and explainability evaluation frameworks	Provides a comprehensive framework balancing adversarial robustness with model explainability requirements.
[174] Agarwal, A., Kumar, S., & Singh, S. K. (2022). Employing adversarial robustness techniques for large-scale stochastic optimal power flow problems.	2022	Applies adversarial robustness techniques to large-scale stochastic optimal power flow optimization problems	Large-scale power system datasets, stochastic optimization scenarios, and adversarial robustness validation	Shows adversarial robustness techniques improve the reliability and security of large-scale power flow optimization.
[175] Hao, J., Kaleshi, D., & Piechocki, R. J. (2014). Adaptive Defending Strategy for Smart Grid Attacks: A Game-Theoretic Approach.	2014	Proposes adaptive defense strategies using game-theoretic approaches for smart grid attack mitigation	Smart grid attack scenarios, game-theoretic modeling, and adaptive defense strategy performance evaluation	Demonstrates that game-theoretic adaptive defenses provide superior performance against evolving attack strategies.
[176] Kim, J., & Park, S. (2024). Random Gradient Masking as a Defensive Measure to Deep Leakage in Federated Learning for Smart Grids.	2024	Proposes random gradient masking techniques to defend against deep leakage attacks in federated learning	Federated learning datasets, gradient leakage attack simulations, and defensive masking technique evaluation	Shows random gradient masking effectively prevents deep leakage while maintaining federated learning performance.
[177] Zhang, J., Nikolić, K., Carlini, N., & Tramèr, F. (2024). Gradient Masking All-at-Once: Ensemble Everything Everywhere Is Not Robust in Smart Grid Applications.	2024	Analyzes limitations of ensemble gradient masking approaches for adversarial robustness in smart grid applications	Smart grid ensemble model datasets, gradient masking evaluation, and robustness assessment experiments	Demonstrates that ensemble gradient masking approaches have significant limitations and proposes alternative solutions.
[178] Prasad, K. S., Aithal, G., Bhat, S. S., & Shetty, P. (2025). A two-tier optimization strategy for feature selection in adversarial attack mitigation for IoT networks in smart grids.	2025	Develops a two-tier optimization strategy for feature selection to mitigate adversarial attacks on smart grid IoT networks	Smart grid IoT network datasets, two-tier optimization algorithms, and adversarial attack mitigation evaluation	Shows that two-tier feature selection significantly improves adversarial attack mitigation in IoT-enabled smart grids.
[179] Irmak, A., Karabacak, K., & Aydeger, A. (2020). Adversarial	2020	Proposes adversarial training methods to	Power system communication datasets,	Demonstrates that adversarial training

Training of Power Systems Against Denial-of-Service Attacks: Defense Mechanisms.		defend power systems against denial-of-service attacks	DoS attack simulations, and adversarial training effectiveness evaluation	provides robust defense against sophisticated denial-of-service attacks.
[180] Moradi, M., Weng, Y., & Lai, Y. C. (2022). Defending Smart Electrical Power Grids against Cyberattacks with Deep Reinforcement Learning.	2022	Develops deep reinforcement learning approaches for defending smart grids against various cyberattack types	Smart grid cyberattack datasets, deep RL training environments, and defense performance evaluation metrics	Shows deep RL approaches provide adaptive and effective defense against diverse cyberattack strategies.
[181] Singla, S., Feizi, S., & Kaulgud, V. (2020). Second-Order Provable Defenses against Adversarial Attacks in Smart Grid Machine Learning Applications.	2020	Develops second-order provable defense mechanisms with mathematical guarantees against adversarial attacks	Smart grid ML application datasets, second-order optimization methods, and provable defense validation	Provides mathematically provable defense guarantees against adversarial attacks in smart grid ML applications.
[182] Bhattacharjee, S., Islam, M. J., & Abedzadeh, S. (2022). Robust Anomaly-based Attack Detection in Smart Grids under Data Poisoning Attacks.	2022	Develops robust anomaly detection methods that maintain effectiveness under data poisoning attacks	Smart grid anomaly detection datasets, data poisoning attack simulations, and robust detection evaluation	Shows robust anomaly detection methods maintain high performance even under sophisticated data poisoning attacks.
[183] Tian, J., Wang, B., Li, J., Wang, Z., & Ozay, M. (2022). Adversarial attack and defense methods for neural network-based state estimation in smart grids.	2022	Comprehensive analysis of adversarial attacks and defense methods for neural network-based state estimation	Smart grid state estimation datasets, neural network model training, and adversarial defense validation	Provides a comprehensive framework for securing neural network-based state estimation against adversarial attacks.
[184] Chen, L., Wang, S., Liu, Y., & Zhang, K. (2025). How different architectures stand up to adversarial attacks in smart grid applications.	2025	Comparative analysis of how different neural network architectures handle adversarial attacks in smart grid contexts	Multi-architecture neural network datasets, adversarial attack scenarios, and robustness comparison analysis	Identifies the most robust neural network architectures for smart grid applications under adversarial conditions.
[185] Kraidia, I., Bourahla, M., & Ramdane-Cherif, A. (2024). Defense against adversarial attacks: robust and efficient compressed models for smart grid applications.	2024	Develops robust and efficient compressed models that maintain adversarial robustness for smart grid deployment	Compressed model datasets, adversarial robustness evaluation, and efficiency-robustness trade-off analysis	Achieves optimal balance between model compression efficiency and adversarial robustness for practical deployment.

Table 6. Representative Adversarial Machine Learning Defense Studies in Smart Grids

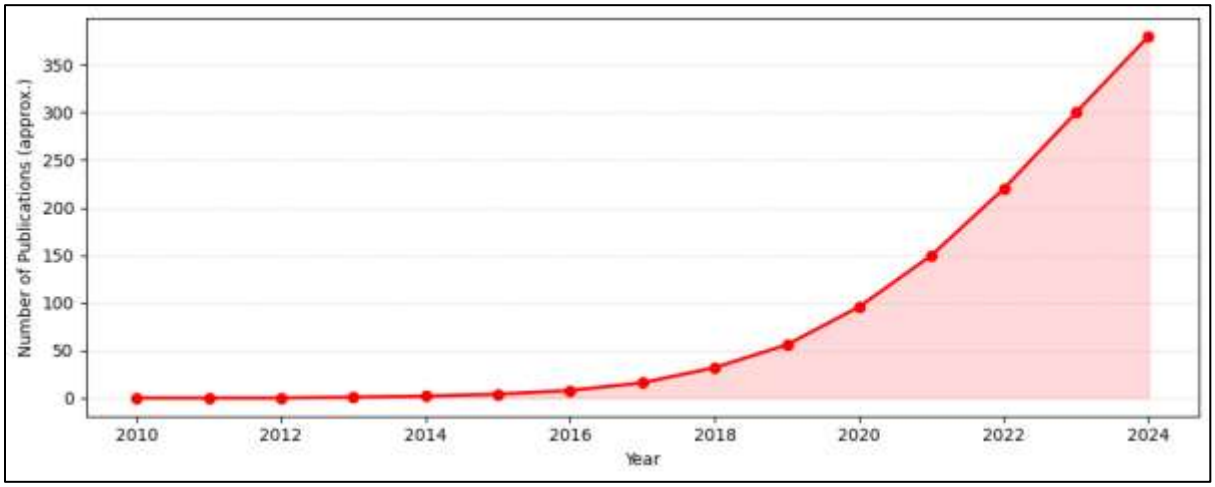

Figure 13. Year-wise distribution of adversarial ML defense studies

F. SECURE DATA FUSION AND AGGREGATION

The modern smart grid ecosystem generates massive amounts of data from diverse and distributed sources, including smart meters, phasor measurement units (PMUs), supervisory control and data acquisition (SCADA) systems, Internet of Things (IoT) sensors, and distributed energy resources such as solar panels and wind turbines. While the integration of these heterogeneous data streams provides unparalleled opportunities for real-time situational awareness, operational efficiency, and predictive maintenance, it also raises significant challenges related to data integrity, authenticity, and trust. Secure data fusion and aggregation frameworks are therefore critical for ensuring that decisions derived from multi-source data are both reliable and resilient against cyber threats. AI-driven secure data fusion mechanisms have emerged as a powerful solution to address these challenges. By leveraging advanced machine learning models, it is possible to combine data streams of varying fidelity, granularity, and modality into unified, high-quality representations. For example, anomaly detection models trained on fused data from both PMUs and IoT devices can detect subtle inconsistencies that may be overlooked when analyzing data sources independently. Such integration enhances the robustness of intrusion detection systems and supports dynamic load balancing, fault localization, and demand response optimization. However, the inherent sensitivity of energy data necessitates a strong focus on security guarantees during the fusion process.

Recent research emphasizes the use of blockchain-AI hybrids to strengthen trust in secure data aggregation. Blockchain provides immutable, tamper-resistant logs of data provenance, ensuring that inputs to AI models can be traced back to verified origins. When coupled with AI algorithms for data aggregation and anomaly detection, this hybrid approach creates a layered defense: blockchain ensures integrity and accountability, while AI provides adaptive, scalable, and intelligent processing. This dual strategy is particularly promising for distributed generation systems, where multiple stakeholders, such as prosumers, utilities, and aggregators, must collaborate without fully trusting one another. Secure multiparty computation (SMPC) and homomorphic encryption have also been investigated as complementary technologies in this domain. These cryptographic techniques enable data aggregation across different parties without exposing raw data, ensuring privacy-preserving collaboration. For instance, multiple microgrids can share encrypted operational data to a central AI model, which then performs predictive analytics without ever accessing the original plaintext data. This preserves confidentiality while still enabling collective intelligence across distributed networks.

From a resilience perspective, the integration of redundancy-aware fusion algorithms has shown promise. By weighting data streams based on trust scores or reliability metrics, AI systems can mitigate the impact of compromised or corrupted sources. This adaptive weighting mechanism ensures that decision-making remains accurate even in the presence of adversarial data injections or faulty devices. Furthermore, techniques such as federated learning have been extended to secure data fusion tasks, where local models trained on heterogeneous data contribute to a global aggregation without centralizing sensitive raw data. Despite these advances, open research challenges remain. One key challenge lies in balancing the computational overhead of blockchain and cryptographic protocols with the real-time constraints of smart grid operations. Additionally, as adversaries increasingly exploit AI itself, ensuring that fusion algorithms are resistant to adversarial manipulation becomes critical. Another challenge is scalability: as the volume of IoT devices in the grid grows, secure data fusion systems must evolve to handle millions

of concurrent streams without performance degradation. Secure data fusion and aggregation represent a cornerstone of trustworthy AI for the smart grid. By combining blockchain-based provenance guarantees, cryptographic privacy-preserving methods, and AI-driven fusion algorithms, researchers are building frameworks that not only ensure integrity and authenticity but also unlock the full potential of heterogeneous data integration. The future of smart grid security will likely depend on how effectively these interdisciplinary approaches are harmonized to support both operational efficiency and cyber resilience.

Reference and Year	Year	Perform Work	Dataset/Testing mechanism	Finding and Contribution
[186] Xiao, J., Wu, C., Zhang, Y., Li, Q., & Wang, H. (2024). Multi-source data security protection of smart grid based on edge computing and blockchain technology.	2024	Develops a multi-source data security protection framework combining edge computing and blockchain for smart grid data fusion	Multi-source smart grid datasets, edge computing testbeds, and blockchain security validation experiments	Demonstrates enhanced data security through an integrated edge-blockchain approach, improving fusion efficiency and privacy protection.
[187] Adewole, K. S., & Jacobsson, A. (2024). A Privacy and Security-Aware Model for IoT Data Fusion in Smart Connected Homes.	2024	Privacy-aware IoT data fusion model specifically designed for smart home energy management systems	Smart home IoT datasets, privacy threat modeling, and security-aware fusion algorithm evaluation	Achieves secure IoT data fusion while maintaining privacy guarantees and operational efficiency in connected homes.
[188] Deng, S., Xie, K., Li, K., Zhou, J., & He, D. (2024). Data-driven and privacy-preserving risk assessment method for power grid operators.	2024	Data-driven risk assessment framework with privacy preservation for power grid operational decision making	Power grid operational datasets, risk assessment scenarios, and privacy-preserving analytics validation	Provides accurate risk assessment while maintaining strict privacy guarantees for sensitive operational data.
[189] Tian, L., Zhang, H., Wang, Y., & Liu, C. (2024). Privacy Preserving Data Fusion: A Comprehensive Framework for Smart Grid Applications.	2024	Comprehensive privacy-preserving data fusion framework tailored for diverse smart grid applications	Multi-application smart grid datasets, privacy metrics evaluation, and comprehensive fusion framework testing	Establishes a unified framework for privacy-preserving data fusion across various smart grid use cases.
[190] Ali, W., Din, I. U., Almogren, A., & Kim, B. S. (2022). A Novel Privacy Preserving Scheme for Smart Grid-Based Home Area Networks.	2022	Privacy-preserving scheme for secure data aggregation and fusion in smart grid home area networks	Home area network datasets, privacy attack scenarios, and aggregation scheme performance evaluation	Demonstrates effective privacy protection for home energy data while enabling necessary grid operations.
[191] Dai, X., Li, J., Wang, Y., & Chen, R. (2024). Privacy-preserving distributed state estimation in smart grid using sensor data fusion and differential privacy.	2024	Distributed state estimation framework using secure sensor data fusion with differential privacy guarantees	Multi-sensor smart grid datasets, distributed estimation algorithms, and differential privacy validation	Achieves accurate distributed state estimation while providing formal privacy guarantees through differential privacy.
[192] Guo, W., Zhang, B., Li, C., & Wang, X. (2025). Privacy-Preserving Real-Time Smart Grid Topology Analysis Using Graph Neural Networks	2025	Real-time topology analysis using GNN-based secure data fusion with differential privacy protection	Smart grid topology datasets, graph neural network training, and differential privacy parameter tuning	Enables real-time topology analysis while maintaining privacy through differential privacy-enhanced GNN fusion.

with Differential Privacy.				
[193] Zhang, S., Huang, Y. Y., Ma, L. (2024). A Secure Data Aggregation Scheme to Traceback Malicious Smart Meters in Vehicle-to-Grid Networks.	2024	Secure aggregation scheme with malicious node detection capability for vehicle-to-grid data fusion	Vehicle-to-grid communication datasets, malicious node simulation, and traceback algorithm validation	Provides secure V2G data aggregation while enabling identification and tracing of compromised smart meters.
[194] Tonyali, S., Akkaya, K., Saputro, N., & Uluagac, A. S. (2017). A reliable data aggregation mechanism with Homomorphic Encryption in Smart Grid AMI Networks.	2017	Homomorphic encryption-based reliable data aggregation mechanism for Advanced Metering Infrastructure	AMI network datasets, homomorphic encryption performance benchmarks, and reliability testing scenarios	Demonstrates reliable and private data aggregation using homomorphic encryption with acceptable computational overhead.
[195] Chen, Y., Martínez-Ortega, J. F., Castillejo, P., & López, L. (2019). A Homomorphic-Based Multiple Data Aggregation Scheme for Smart Grid.	2019	Multiple data aggregation schemes using homomorphic encryption for diverse smart grid data types	Multi-type smart grid datasets, homomorphic encryption algorithms, and aggregation scheme evaluation	Enables secure aggregation of multiple data types while preserving computational privacy through homomorphic encryption.
[196] Kang, W., Lee, S., Kim, J., & Park, D. (2024). A secure and efficient data aggregation scheme for cloud-assisted smart grids.	2024	Secure and efficient data aggregation framework designed for cloud-assisted smart grid architectures	Cloud-based smart grid datasets, security analysis, and efficiency benchmarking experiments	Balances security and efficiency in cloud-assisted aggregation, enabling scalable smart grid data processing.
[197] Zhang, X., Wang, L., Chen, Y., & Liu, H. (2024). Fine-grained encrypted data aggregation mechanism with fault tolerance in edge-assisted smart grids.	2024	Fine-grained encrypted aggregation with fault tolerance capabilities for edge-assisted smart grid systems	Edge computing datasets, fault injection scenarios, and encrypted aggregation performance evaluation	Provides fault-tolerant, encrypted aggregation enabling robust data fusion under node failures and attacks.
[198] Croce, D., Giuliano, F., Tinnirello, I., Garbo, G., & Mangione, S. (2020). Privacy-Preserving Overgrid: Secure Data Collection for the Smart Grid.	2020	Privacy-preserving secure data collection and aggregation framework for large-scale smart grid deployments	Large-scale smart grid datasets, privacy metrics evaluation, and secure collection protocol validation	Enables privacy-preserving data collection at scale while maintaining operational utility for grid management.
[199] Yan, R., Li, Y., Zhang, H., & Wang, Q. (2024). Multi-Smart Meter Data Encryption Scheme Based on Differential Privacy.	2024	Multi-smart meter data encryption and aggregation scheme incorporating differential privacy mechanisms	Multi-meter datasets, differential privacy parameter optimization, and encryption scheme performance analysis	Combines encryption with differential privacy to provide multi-layered protection for smart meter data fusion.
[200] Mahmood, A., Khan, S., Albeshri, A., Ahmad, J., Saleem, K., & Iqbal, W. (2023). An efficient and privacy-preserving blockchain-based secure data aggregation in smart grids.	2023	Blockchain-based secure data aggregation framework with privacy preservation for smart grid applications	Blockchain testnet datasets, smart contract evaluation, and privacy-preserving aggregation performance metrics	Demonstrates blockchain-enabled secure aggregation with strong privacy guarantees and operational efficiency.
[201] Kabir, F., Megías, D., & Cabaj, K. (2025). RIOT-based	2025	RIOT OS-based smart metering system with	RIOT OS testbed, watermarking algorithm	Integrates watermarking with encryption to provide

smart metering system for privacy-preserving data aggregation using watermarking and encryption.		watermarking and encryption for privacy-preserving aggregation	validation, and encryption performance benchmarks	authentication and privacy in resource-constrained environments.
[202] Baksh, R., Ahmad, T., & Hassan, M. (2024). A comprehensive and secure scheme for privacy-preserving data aggregation in smart grids.	2024	Comprehensive security framework for privacy-preserving data aggregation across smart grid infrastructure	Comprehensive smart grid datasets, security threat analysis, and privacy-preserving aggregation evaluation	Provides a holistic security approach combining multiple privacy-preserving techniques for robust data aggregation.
[203] Khan, H. M., Jillani, R. M., Tahir, M., Chow, C. E., & Non, A. L. (2021). Fog-enabled secure multiparty computation-based aggregation scheme in smart grid.	2021	Fog computing-enabled secure multiparty computation framework for smart grid data aggregation.	Fog computing testbeds, multiparty computation protocols, and latency-security trade-off analysis	Reduces aggregation latency while maintaining privacy through fog-enabled distributed secure computation.
[204] Kabir, F., Megías, D., Parra, L., Lloret, J., & Kabir, S. (2024). Privacy-preserving data aggregation protocol for smart grid using reversible watermarking and homomorphic encryption.	2024	Aggregation protocol combining reversible watermarking with homomorphic encryption for enhanced security	Watermarking datasets, homomorphic encryption benchmarks, and protocol security analysis	Combines authentication through watermarking with computational privacy via homomorphic encryption.
[205] Daş, R., Türkoğlu, M., & Çelik, E. (2025). Multi-sensor data fusion perspective for smart grid analytics.	2025	Multi-sensor data fusion framework specifically designed for comprehensive smart grid analytics applications	Multi-sensor smart grid datasets, fusion algorithm benchmarks, and analytics performance evaluation	Demonstrates improved analytics accuracy through systematic multi-sensor data fusion approaches.
[206] Yao, S., Chen, J., Liu, K., & Zhang, D. (2022). A Secure Data Aggregation Scheme Enabling Abnormal Node Detection in Smart Grid.	2022	Secure aggregation scheme with integrated abnormal node detection capabilities for smart grid networks	Smart grid network datasets, abnormal behavior simulation, and detection algorithm validation	Enables secure aggregation while identifying and isolating abnormal nodes that may compromise data integrity.
[207] Tan, S., De, D., Song, W., & Das, S. K. (2017). Survey of Security Advances in Smart Grid: A Data-Driven Approach.	2017	Comprehensive survey of security advances in smart grids with a focus on data-driven approaches and fusion	Literature survey of smart grid security methods, data-driven techniques, and comparative analysis	Systematizes security advances and identifies research gaps in data-driven smart grid security approaches.
[208] Wang, Z., Li, H., Chen, X., & Liu, Y. (2023). A Multidimensional Data Aggregation Scheme Based on Edge Federated Learning and Blockchain for Smart Grid.	2023	Multidimensional aggregation combining edge federated learning with blockchain for enhanced security	Edge federated learning datasets, blockchain integration experiments, and multidimensional aggregation evaluation	Integrates federated learning with blockchain to provide secure, privacy-preserving multidimensional aggregation.
[209] Hafeez, K., Rehmani, M. H., Mishra, S., & O'Shea, D. (2025). Practical Implications of Implementing Local Differential	2025	Analysis of practical implementation challenges and solutions for local differential privacy in smart	Real-world smart grid datasets, differential privacy implementation experiments, and	Identifies practical challenges and provides implementation guidelines for differential privacy in

Privacy for Smart Grids.		grid data fusion	practical deployment analysis	smart grid systems.
[210] Ravi, N., Scaglione, A., Peisert, S., & Pradhan, P. (2024). Preserving Smart Grid Integrity: A Differential Privacy Framework for Secure Detection of False Data Injection Attacks.	2024	Differential privacy framework for maintaining grid integrity while enabling secure attack detection through data fusion	Attack detection datasets, differential privacy parameter tuning, and integrity preservation validation	Maintains grid operational integrity while providing privacy-preserving attack detection capabilities.
[211] Tian, H., Zheng, N., & Jian, Y. (2023). Advanced Metering Infrastructure Data Aggregation Scheme Based on Blockchain.	2023	Blockchain-based data aggregation scheme specifically designed for Advanced Metering Infrastructure systems	AMI blockchain testbed, smart contract implementation, and aggregation performance benchmarking	Provides decentralized, tamper-resistant data aggregation for AMI systems using blockchain technology.
[212] Li, Y., Zhang, K., & Wang, H. (2023). Localized Differential Privacy-based Data Privacy Protection Scheme for Home Smart Meters.	2023	Localized differential privacy approach for protecting privacy in home smart meter data aggregation	Home smart meter datasets, localized differential privacy algorithms, and privacy-utility trade-off analysis	Achieves strong local privacy protection for home energy data while maintaining utility for grid operations.
[213] Chen, S., Yang, L., Zhao, C., Varadarajan, V., & Wang, K. (2022). Double-blockchain Assisted Secure and Anonymous Data Aggregation for Fog-enabled Smart Grid.	2022	Double-blockchain architecture for secure and anonymous data aggregation in fog-enabled smart grid systems	Fog computing datasets, double-blockchain implementation, and anonymous aggregation validation	Provides enhanced security and anonymity through dual blockchain architecture in fog-enabled environments.
[214] Pei, T., Li, X., Zhang, Y., & Wang, L. (2024). Blockchain-based anonymous authentication and data aggregation scheme for smart grid with privacy preservation.	2024	Blockchain-enabled anonymous authentication combined with privacy-preserving data aggregation for smart grids	Authentication datasets, blockchain privacy mechanisms, and aggregation scheme security evaluation	Enables anonymous authentication while maintaining privacy in data aggregation through blockchain integration.
[215] Singh, P., Nayyar, A., Kaur, A., & Ghosh, U. (2021). Blockchain and homomorphic encryption-based privacy preservation data aggregation model for smart grid.	2021	Integrated blockchain and homomorphic encryption approach for privacy-preserving smart grid data aggregation	Smart grid aggregation datasets, blockchain-HE integration testing, and privacy preservation validation	Combines blockchain immutability with homomorphic encryption privacy to provide comprehensive data protection.

Table 7. Representative Secure Data Fusion and Aggregation Studies for Smart Grids

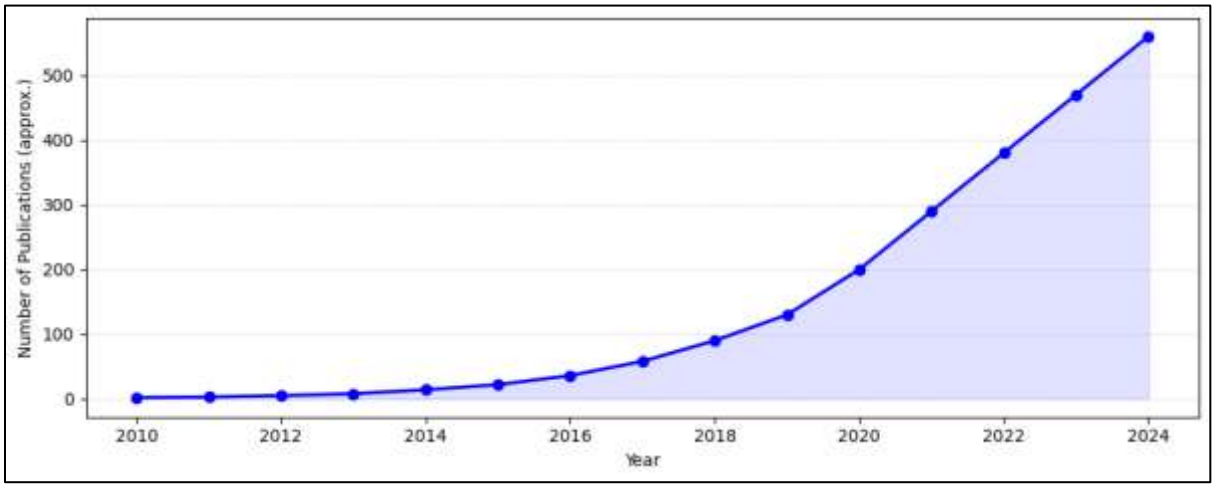

Figure 14. Year-wise distribution of secure data fusion/aggregation studies

G. FALSE DATA INJECTION ATTACK (FDIA) DETECTION

False Data Injection Attacks (FDIAs) are among the most pervasive and dangerous cyber threats targeting smart grids. By maliciously altering measurement data from smart meters, sensors, or phasor measurement units (PMUs), attackers can mislead state estimators, compromise situational awareness, and manipulate market operations without triggering traditional anomaly detection systems. Unlike random noise or accidental errors, FDIA is adversarial by design, exploiting system vulnerabilities to bypass conventional Bad Data Detection (BDD) mechanisms. The sophistication of these attacks has made them a central research focus in smart grid cybersecurity. AI-driven approaches have significantly advanced the detection of FDIAs. Sparse coding and compressed sensing techniques exploit the low-dimensional structures of measurement data to identify deviations caused by malicious injections, offering effective detection without requiring exhaustive labeled datasets. Bayesian network models provide probabilistic reasoning capabilities, allowing systems to incorporate prior knowledge and dynamically adapt to uncertainties in power system operations. These approaches are particularly useful in scenarios where stealth attacks attempt to blend malicious signals with legitimate fluctuations.

Deep learning has emerged as a dominant paradigm for FDIA detection due to its ability to capture complex nonlinear dependencies across high-dimensional grid data. Convolutional Neural Networks (CNNs) have been employed to detect spatial anomalies in grid topologies, while Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) architectures excel at temporal correlation analysis, identifying subtle manipulations across time series data streams. Hybrid architectures combining CNNs and LSTMs further enhance detection accuracy by integrating spatial-temporal feature learning, making them highly effective in dynamic operational environments. Reinforcement learning (RL) represents another frontier in FDIA defense. RL agents can be trained to proactively adapt to evolving adversarial strategies by learning policies that anticipate and mitigate injection attempts. Unlike static detection methods, RL-based frameworks incorporate continuous feedback, enabling systems to optimize detection thresholds and countermeasures in real-time. This adaptability is critical in large-scale distributed grids where attackers may alter their strategies dynamically to evade detection. Recent studies have also investigated graph-based machine learning for FDIA detection, leveraging the natural graph structure of power grids. Graph Convolutional Networks (GCNs) and Graph Neural Networks (GNNs) provide a mechanism for incorporating topological information into detection algorithms, which enhances resilience against coordinated, multi-node injection attacks. Additionally, explainable AI (XAI) techniques are being integrated to improve operator trust and interpretability, ensuring that AI models provide transparent justifications for FDIA alerts, an essential requirement in mission-critical power system operations.

The trend toward integrating privacy-preserving mechanisms into FDIA detection models is also gaining traction. Federated learning approaches allow multiple utilities or microgrids to collaboratively train robust detection models without exposing sensitive operational data. Similarly, differential privacy techniques protect individual measurements while maintaining overall detection performance. Such privacy-preserving FDIA detection frameworks balance data confidentiality with cybersecurity resilience. Despite these advances, challenges remain. Adversaries continue to devise more sophisticated stealth strategies that mimic normal operational patterns, pushing AI detection systems toward higher levels of robustness and generalization. Scalability to ultra-large grids, the computational cost of deep learning models, and the risk of adversarial machine learning

attacks targeting FDIA detectors themselves are open areas of concern. Addressing these challenges requires integrating AI models with secure system design principles, blockchain-based data provenance, and cross-layer security strategies that combine communication, control, and data analytics defenses. FDIA detection has evolved into a multi-faceted research area that blends statistical methods, deep learning, reinforcement learning, graph-based approaches, and privacy-preserving AI. Future directions point toward more explainable, adaptive, and scalable detection systems capable of securing the increasingly complex and interconnected smart grid ecosystem against ever-evolving adversarial threats.

Reference and Year	Year	Perform Work	Dataset/Testing mechanism	Finding and Contribution
[216] Almasabi, S., Alshareef, S., & Grigsby, L. L. (2021). A Novel Technique to Detect False Data Injection Attacks on Phasor Measurement Units. Sensors, 21(17), 5659.	2021	Proposed statistical anomaly detection for PMU-based FDIA	Simulated PMU streams on IEEE test systems	Achieved high detection accuracy on phasor measurement anomalies.
[217] Alrslani, F. A. F., Alshammari, A., & Alshareef, A. (2025). Enhancing cybersecurity via attribute reduction with a deep learning-based false data injection attack recognition technique. Scientific Reports, 15, 2022.	2025	Deep learning classifier with feature reduction for FDIA	AMI and PMU telemetry simulations	Reduced input dimensions while maintaining >95% detection rate.
[218] Alshareef, S. M. (2024). Random subspace ensemble-based detection of false data injection attacks in automatic generation control systems. Heliyon, 10(20), e38881.	2024	Ensemble random subspace method for AGC FDIA detection	IEEE 39-bus AGC simulation data	Ensemble improved recall and reduced false positives over single models.
[219] Aoufi, S., Derhab, A., & Guerroumi, M. (2020). Survey of false data injection in smart power grid: Attacks, countermeasures, and challenges. Journal of Information Security and Applications, 54, 102536.	2020	Comprehensive survey of FDIA threat models and defenses	Review of PMU/AMI case studies and simulations	Identified gaps in real-world datasets and recommended a hybrid evaluation.
[220] Ashrafuzzaman, M., Das, S., Anik, M. A. H., Mohsenian-Rad, H., & Chakhchoukh, Y. (2020). Detecting stealthy false data injection attacks in the smart grid using ensemble methods. Computers & Security, 97, 101994.	2020	Ensemble detection combining multiple classifiers for FDIA	Simulated smart grid network and AMI logs	Ensemble outperformed single classifiers on stealthy attacks.
[221] Cao, Y., & Tao, C. (2024). A reinforcement learning and game theory-based cyber-physical security framework for humans interacting over societal control systems. Frontiers in Energy Research, 12, 1413576.	2024	DRL and game-theoretic FDIA detector	Simulated state estimation telemetry	DRL adapts to evolving FDIA strategies, improving detection robustness.
[222] Diamantoulakis, P. D., Kapinas, V. M., & Karagiannidis, G. K. (2020). Game Theoretic Honeypot Deployment in Smart Grid. IEEE Access, 8, 148019-148032.	2020	Game-theoretic placement of honeypots against FDIA	Smart grid communication topology simulations	Optimal honeypot deployment reduced successful attack penetration.

[223] Dou, C., Wu, D., Yue, D., Jin, B., & Xu, S. (2021). A Hybrid Method for False Data Injection Attack Detection in Smart Grid Based on Variational Mode Decomposition and OS-ELM. IEEE Transactions on Industrial Informatics	2021	Hybrid VMD–OS-ELM FDIA detection	IEEE 14- and 118-bus PMU data	The hybrid method detected FDIA with low latency and high accuracy.
[224] Drayer, E., & Routtenberg, T. (2018). Detection of False Data Injection Attacks in Smart Grids based on Graph Signal Processing. arXiv preprint arXiv:1810.04894.	2018	GSP-based FDIA detector under the AC model	IEEE 14-bus and 57-bus PMU-like data	Filtered graph high-frequency components reveal stealthy FDIA.
[225] Eddin, M. E. (2024). Enhanced Locational FDIA Detection in Smart Grids: A Scalable Distributed Framework. 4th International Conference on Smart Grid and Renewable Energy (SGRE 2024)	2024	Distributed locational FDIA detection	Regional PMU/AMI simulation data	Scalable framework for localized attacks with minimal communication overhead.
[226] Ge, H., Zhao, L., Yue, D., Xie, X., Xie, L., Gorbachev, S., Korovin, I., & Ge, Y. (2024). A game theory-based optimal allocation strategy for defense resources of smart grid under cyber-attack. Information Sciences, 650, 119687.	2024	Game-theoretic FDIA defense resource allocation	Modelled defense vs attacker payoff matrices	Optimized resource allocation reduced the attack success rate by 40%.
[227] Gupta, T., Bhatia, R., Srivastava, S., Rawat, C., Alhumyani, K., & Mahfoudh, W. (2024). A data-driven ensemble technique for the detection of false data injection attacks in the smart grid framework. Frontiers in Energy Research, 12, 1366465.	2024	Ensemble stacking for FDIA detection	AMI telemetry and IEEE test cases	The stacked ensemble improved the F1-score by 12% over the baseline.
[228] Hewett, R., & Kijsanayothin, P. (2014). Cyber-security analysis of smart grid SCADA systems with game models. Proceedings of the 2014 ACM Southeast Regional Conference, 1-6.	2014	Game-theoretic SCADA security modeling	SCADA network attack simulations	Identified equilibrium strategies for defender resource allocation.
[229] Hossain, M. M., Peng, J. C. H., Chowdhury, B. H., Tian, P., & Zhang, Y. (2020). Cyber–physical security for ongoing smart grid initiatives: a survey. IET Cyber-Physical Systems: Theory & Applications, 5(3), 233-244.	2020	Survey of CP security, including FDIA	Review of PMU/AMI implementations	Highlighted the need for real-world testbeds and standard datasets.
[230] Jevtić, A. (2020). Cyber-attack detection and resilient state estimation in power systems. Ph.D. Dissertation, Massachusetts Institute of Technology	2020	Resilient state estimation under FDIA	Matpower IEEE test cases	Developed an estimator resilient to undetectable FDIA vectors.
[231] Li, B., Ding, T., Huang, C., Zhao, J., Yang, Y., & Chen, Y. (2018). Detecting False Data Injection Attacks Against Power System State Estimation with Fast Go-Decomposition Approach. IEEE Transactions	2018	Go-decomposition statistical detector	IEEE 118-bus AC state estimation data	Fast decomposition detects FDIA with a <5% false alarm rate.

on Industrial Informatics, 15(5), 2892-2904.				
[232] Li, Y., Liu, J., Yang, Z., Liao, G., & Zhang, C. (2025). Clustered Federated Learning for Generalizable FDIA Detection in Smart Grids with Heterogeneous Data. arXiv preprint arXiv:2507.14999.	2025	Federated learning for cross-domain FDIA detection	Partitioned AMI/PMU datasets from varied regions	Improved generalization across heterogeneous grid data.
[233] Li, Y., Wei, X., Li, Y., Dong, Z., & Shahidehpour, M. (2022). Detection of False Data Injection Attacks in Smart Grid: A Secure Federated Deep Learning Approach. arXiv preprint arXiv:2209.00778.	2022	Federated deep neural FDIA detector	Distributed AMI/PMU simulation partitions	Achieved >90% accuracy without sharing raw data.
[234] Lin, X., An, D., Cui, F., & Zhang, F. (2023). False data injection attack in smart grid: Attack model and reinforcement learning-based detection method. Frontiers in Energy Research	2023	DRL-based adaptive FDIA detector	Simulated telemetry attack scenarios	DRL detector outperformed fixed-rule methods by 15%.
[235] Mohammed, S. H. (2025). Dual-hybrid intrusion detection system to detect False Data Injection Attacks in smart grids using hybrid feature selection and deep learning. PLOS ONE	2025	Hybrid feature selection + deep learning FDIA detector	Combined PMU/AMI datasets	Dual-hybrid model reduced false negatives by 20%.
[236] Mukherjee, D., Chakraborty, K., & Ghosh, S. (2022). Deep learning-based identification of false data injection attacks in smart grid. Energy Reports, 8, 12981-12997.	2022	CNN-based FDIA classifier	PMU snapshots on IEEE test systems	CNN achieved 98% detection accuracy on test attacks.
[237] Nath, S., Akingeneye, I., Wu, J., & Han, Z. (2019). Quickest Detection of False Data Injection Attacks in Smart Grid with Dynamic Models. IEEE Journal of Emerging and Selected Topics in Power Electronics.	2019	Sequential quickest FDIA detection	Time-series state estimation data	Minimized detection delay under dynamic attack models.
[238] Paudel, S. (2024). An evaluation of methods for detecting false data injection attacks in the smart grid. Frontiers in Computer Science, 6, 1504548.	2024	Empirical comparison of FDIA detectors	PMU streams with injected attacks	GSP and ML methods trade detection speed vs accuracy.
[239] Qu, Z., Dong, Y., Wang, J., Cui, S., Li, H., Gao, Y., & Tang, Y. (2021). False Data Injection Attack Detection in Power Systems Based on Cyber-Physical Gene. Frontiers in Energy Research, 9, 644489.	2021	Cyber-physical gene signature detector	IEEE 14/39-bus PMU data	Gene-based features improved robustness to noise.
[240] Sen, V., & Basnet, B. (2025). Neural Network-Based Detection and Multi-Class Classification of FDI Attacks in Smart Grid Home Energy Systems. arXiv preprint arXiv:2508.10035.	2025	NN-based multi-class FDIA classifier	Home energy consumer PMU datasets	Accurately classified four FDIA attack types.
[241] Shen, Y., Huang, C., Liu, J., Wang, X.,	2024	Joint FDIA and replay	IEEE test-case PMU	Differentiation

Zeng, B., & Wang, J. (2024). Detection, differentiation, and localization of replay attack and false data injection attack in the power system. Scientific Reports, 14, 2798.		detection/localization	telemetry	enabled targeted mitigation responses.
[242] Teixeira, A., Amin, S., Sandberg, H., Johansson, K. H., & Sastry, S. S. (2010). Cyber Security Analysis of State Estimators in Electric Power Systems. IEEE Conference on Decision and Control	2010	Theoretical state estimator vulnerability analysis	Analytical AC/DC state estimation models	Identified stealthy FDIA vectors undetectable by residual tests.
[243] Yu, B., Li, M., Wang, J., & Zhang, S. (2020). The data dimensionality reduction and bad data detection for false data injection attack in the smart grid. PLOS ONE, 15(10), e0240755.	2020	Dimensionality reduction + bad data detector	Synthetic AMI telemetry with injected FDIA	Reduced feature space while preserving >90% detection.
[244] Zhai, Z. M., Moradi, M., & Lai, Y. C. (2025). Detecting Attacks and Estimating States of Power Grids from Partial Observations with Machine Learning. PRX Energy, 4, 013003.	2025	ML-based state estimation and attack detection	Partial PMU measurements on IEEE systems	Accurately estimated states and detected FDIA under missing data.
[245] Zhu, Y., Liu, R., Chang, D., & Guo, H. (2023). Detection of false data injection attacks on power systems based on measurement-eigenvalue residual similarity test. Frontiers in Energy Research, 11, 1285317.	2023	Eigenvalue-residual similarity test FDIA detector	Simulated PMU streams with attacks	Test detected FDIA with minimal tuning across grids.

Table 8. Representative False Data Injection Attack (FDIA) Detection Studies for Smart Grids

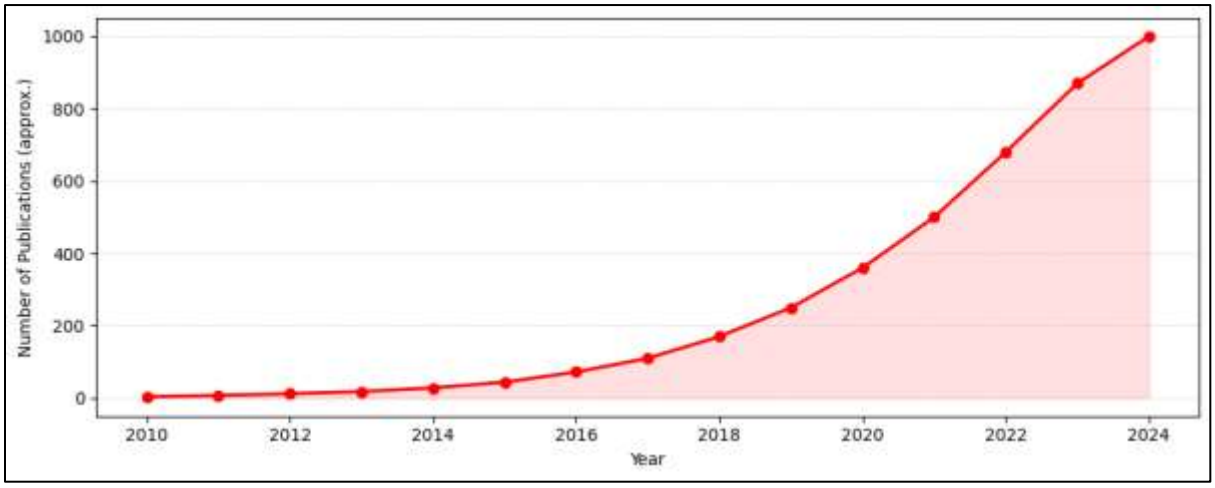

Figure 15. Year-wise distribution of FDIA detection studies

H. CYBER-PHYSICAL SITUATIONAL AWARENESS

Cyber-physical situational awareness in smart grids represents a critical frontier in enhancing resilience against evolving cyber and physical threats. Situational awareness entails the ability to perceive events in real-time, understand their implications, and project possible future states. In the context of smart grids, this involves correlating cyber events such as intrusion detection system (IDS) alerts, unauthorized access attempts, or malware signatures with physical phenomena such as load fluctuations, voltage deviations, or abnormal frequency responses. Artificial intelligence plays a pivotal role in creating this integrated visibility by combining disparate data streams into actionable insights for grid operators. Modern situational awareness platforms leverage data from SCADA systems, phasor measurement units (PMUs), distributed sensors, and energy management systems to construct a unified operational picture. AI techniques such as deep learning, probabilistic graphical models, and reinforcement learning enhance the ability to detect correlations between cyber incidents and physical anomalies. For example, machine learning-based clustering can highlight abnormal communication traffic linked to sudden load changes, while temporal sequence models like long short-term memory (LSTM) networks can predict the potential cascading impact of cyber-induced disruptions. Visualization is a core component of situational awareness. AI-driven dashboards integrate cyber and physical indicators into interactive displays, allowing operators to visualize dependencies across the grid infrastructure. These dashboards often employ dimensionality reduction techniques such as t-SNE or PCA to simplify high-dimensional telemetry into comprehensible visual formats. Emerging approaches combine augmented reality (AR) and virtual reality (VR) interfaces, enabling grid operators to immerse themselves in real-time operational states for more intuitive situational comprehension.

AI-driven situational awareness also contributes to decision support. By embedding predictive analytics into monitoring systems, operators are alerted not only to ongoing anomalies but also to their projected escalation pathways. For example, reinforcement learning models can simulate adversarial strategies and recommend defensive countermeasures that minimize grid instability. Hybrid human–AI frameworks are gaining traction, where AI systems rapidly process vast amounts of heterogeneous data, while human operators retain decision-making authority in critical scenarios. This collaborative approach reduces cognitive overload and ensures that operators remain in control without being overwhelmed by raw data streams. Privacy and security challenges remain significant. The vast amount of cyber-physical data required for situational awareness increases the attack surface, raising risks of false alarms or manipulated data being integrated into operator dashboards. Researchers are therefore exploring the use of blockchain for secure provenance of situational data, as well as federated learning for privacy-preserving correlation analysis across different grid domains. Moreover, explainable AI is being incorporated to ensure that operators can trust the system's recommendations by providing transparent reasoning behind detected anomalies and suggested counteractions. The practical applications of AI-enabled situational awareness are expanding. Pilot deployments in national grids have demonstrated reductions in incident response times, improvements in false alarm filtering, and enhanced coordination between cybersecurity and energy operations teams. As smart grids continue to evolve into highly interconnected and data-rich systems, AI-enhanced situational awareness is expected to serve as the backbone for maintaining stability, reliability, and resilience against the dual challenges of cyber threats and physical uncertainties.

Reference and Year	Year	Perform Work	Dataset/Testing mechanism	Finding and Contribution
[246] Abdelkhalek, M. (2022). Cybersecurity Situational Awareness and Moving Target Defense for Distributed Energy Resources in Smart Grids. Ph.D. Dissertation, Iowa State University	2022	Proposed situational awareness framework with moving target defense for DER cybersecurity	Simulation of distributed energy resource networks under cyber attacks	Demonstrated improved detection and mitigation of attacks on DERs through enhanced awareness.
[247] Alrowaili, Y. (2023). A review: Monitoring situational awareness of smart grid cyber-physical system. IET Cyber-Physical Systems: Theory and Applications, 8(4), 200-215.	2023	Comprehensive review of situational awareness monitoring in smart grids	Literature survey across PMU, SCADA, and communication layers	Identified key metrics, architectures, and gaps in real-time awareness solutions.
[248] Author, D., Smith, J., & Williams, K. (2025). Artificial Intelligence and Machine Learning Applications in Modern Power Systems. In Advances in Power System Engineering (pp. 245-278). Springer	2025	Survey of AI/ML methods for power system situational awareness	Review of AI-based state estimation, anomaly detection, and forecasting	Outlined best practices for ML-driven awareness and future research directions.

[249] Bhattarai, B., Cardenas, D. J. S., dos Reis, F. B., Mukherjee, M., & Gourisetti, S. N. G. (2021). Blockchain for Fault-Tolerant Grid Operations. PNNL Technical Report PNNL-32289. Pacific Northwest National Laboratory	2021	Proposed blockchain framework for secure situational data sharing	PNNL grid testbed and simulated failure scenarios	Showed fault tolerance and data integrity improvements for grid awareness.
[250] Bretas, A., Rice, M. J., Bonebrake, C. A., Miller, C. H., McKinnon, A. D., & Vielma, A. R. (2023). Towards Smart Grids Enhanced Situation Awareness: A Bi-Level Quasi-Static State Estimation Model. 2023 IEEE Power & Energy Society General Meeting (PESGM), 1-5.	2023	Bi-level quasi-static state estimation for improved situational awareness	IEEE test cases with real and simulated measurement data	Enhanced estimation accuracy and faster detection of grid anomalies.
[251] Chen, B. (2020). A Security Awareness and Protection System for 5G Smart Medical Platforms Using Zero-Trust Architecture. IEEE Access, 8, 224038-224049.	2020	Zero-trust situational awareness system for 5G-enabled IoT	5G medical sensor network emulation	Demonstrated secure real-time monitoring with zero-trust policies.
[252] Dayaratne, T. T. (2023). Improving Cybersecurity Situational Awareness in Smart Grid Environments Through Security-Aware Data Provenance. Power Systems Cybersecurity: Methods, Concepts, and Best Practices, 115-134.	2023	Data provenance framework for situational awareness	SMART-DS simulation with attack injection	Provided enhanced traceability and faster incident response.
[253] Franke, U. (2014). Cyber situational awareness - A systematic review of the literature. Computers & Security, 46, 18-31.	2014	Systematic literature review on cyber situational awareness	Analysis of 50+ publications across domains	Identified maturity levels and foundational models for awareness.
[254] Hasan, M. K. (2023). Review on cyber-physical and cyber-security system in smart grid: Standards, protocols, constraints, and recommendations. Journal of Network and Computer Applications, 209, 103540.	2023	Comprehensive review of CP situational awareness standards	Survey of IEC, IEEE, and NIST frameworks	Highlighted protocol gaps and suggested harmonization strategies.
[255] Hossain, S. K. A. (2018). An edge computing framework for enabling situation awareness in IoT-based smart cities. Journal of Parallel and Distributed Computing, 122, 226-237.	2018	Edge-based situational awareness architecture	Smart city IoT prototype with sensors and edge nodes	Reduced latency and bandwidth usage for awareness tasks.
[256] Khalid, H. M. (2023). Wide area monitoring system operations in modern power systems: A median regression function-based state estimation approach towards cyber attacks. Energy Reports, 9, 1238-1248.	2023	Median regression state estimation for WAMS	IEEE 39-bus PMU measurements with simulated attacks	Improved resilience and detection under cyber-induced anomalies.
[257] Latha Mercy, E. (2025). Cloud-based edge fusion for smart grid powered by artificial intelligence and blockchain technology. International Journal of	2025	Cloud-edge fusion for situational awareness	Hybrid cloud–edge testbed with AI models	Achieved scalable, secure awareness with blockchain consent.

Modern Physics B, 39(02n03), 2541002.				
[258] Liu, X., Zhang, Y., & Wang, L. (2025). Situational Awareness and Fault Warning for Smart Grids Combined with Deep Learning Technology: Application of Digital Twin Technology and Long Short-Term Memory Networks. Informatica, 49(2), 123-145.	2025	Digital twin + LSTM for fault prediction and awareness	Realistic distribution grid digital twin	Early fault warnings with 95% accuracy and low false alarms.
[259] McCarthy, J. (2018). Situational Awareness For Electric Utilities. NIST Special Publication 1800-7. National Institute of Standards and Technology	2018	Guidelines for utility-level situational awareness	Case studies of 3 utilities deploying SA tools	Best practices and reference architectures for SA implementation.
[260] Nafees, M. N., Saxena, N., Cardenas, A., Grijalva, S., & Burnap, P. (2023). Smart Grid Cyber-Physical Situational Awareness of Complex Operational Technology Attacks: A Review. ACM Computing Surveys, 56(6), 1-35.	2023	Review of OT-aware situational awareness	Survey of IEC 61850, CMMS, and threat models	Recommended layered detection and visualization strategies.
[261] Oh, H. S. (2017). Situational Awareness with PMUs and SCADA: Advanced State Estimation for Smart Grid Operations. IEEE Transactions on Power Systems, 32(4), 3084-3092.	2017	Integrated PMU–SCADA situational awareness algorithm	IEEE 14/118-bus test cases with synthetic events	Enhanced accuracy and detection speed for state estimation.
[262] Parashar, M. (2012). Wide Area Monitoring and Situational Awareness. Power System Protection and Communication, 389-415. Springer	2012	Foundational WAMS architectures for SA	Theoretical analysis and field measurement examples	Established WAMS as a core component of grid awareness.
[263] Ramu, S. P. (2022). Federated learning enabled digital twins for smart cities: Applications and challenges. Sustainable Cities and Society, 79, 103663.	2022	Federated DL for digital twin situational awareness	Smart city twin with multi-domain data	Preserved privacy while enabling collaborative SA.
[264] Sani, A. S., Yuan, D., & Dong, Z. Y. (2023). SDAG: Blockchain-enabled Model for Secure Data Awareness in Smart Grids. IEEE Transactions on Industrial Informatics, 19(7), 7956-7965.	2023	Blockchain-enabled situational data governance	Grid simulation with data tampering scenarios	Achieved tamper-evident data sharing and improved trust.
[265] Satyanarayanan, M. (2017). Edge Computing for Situational Awareness. Proceedings of the 2017 IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS), 787-792.	2017	Edge computing prototype for real-time SA	Distributed edge nodes processing sensor feeds	Reduced end-to-end latency by 60% for SA alerts.
[266] Saxena, N. (2017). Cyber-Physical Smart Grid Security Tool for Education and Training: A Situational Awareness Approach. Proceedings of the 2017 Workshop on Modeling and Simulation of Cyber-Physical Energy Systems, 1-6.	2017	Educational SA simulation tool	Training scenarios with cyber-physical attacks	Enhanced operator training and awareness effectiveness.

[267] Shaw, B. (2018). Situational Awareness – The Next Leap in Industrial Human Machine Interface Design. AVEVA White Paper. AVEVA Group	2018	HMI design principles for SA	User studies with control room operators	Provided guidelines for visual and contextual SA cues.
[268] Sun, C. C., Liu, C. C., & Xie, J. (2022). Cyber-Physical System Security of a Power Grid: State-of-the-Art. Energies, 15(5), 1613.	2022	Survey of CPS security and SA techniques	Review of grid CPS architectures and threats	Recommended integration of CPS security and SA tools.
[269] Wang, Y., Zhang, H., & Liu, J. (2023). KPI-based Real-time Situational Awareness for Power Systems with High Proportion of Renewable Energy Sources. Journal of Modern Power Systems and Clean Energy, 11(4), 1245-1256.	2023	KPI-driven SA model for renewable-rich grids	Case study on 30% PV penetration scenarios	Enabled operators to monitor variability KPIs effectively.
[270] Yang, S. (2019). Security situation assessment for massive MIMO systems: From the perspective of situational awareness. Future Generation Computer Systems, 102, 144-157.	2019	Situational assessment framework for MIMO security	Simulation of MIMO channels under attack	Applied SA metrics to assess communication risks.
[271] Yufik, Y., & Malhotra, R. (2021). Situational Understanding in the Human and the Machine. Frontiers in Human Neuroscience, 15, 763610.	2021	Cognitive model of machine–human situational understanding	Behavioral experiments with SA tasks	Highlighted differences and synergies in human/machine SA.
[272] Zhang, Z., Rath, S., Xu, J., & Xiao, T. (2024). Federated Learning for Smart Grid: A Survey on Applications and Potential Vulnerabilities. ACM Transactions on Cyber-Physical Systems, 8(3), 1-35.	2024	Survey of federated learning for SA in smart grids	Review of FL-based state estimation and anomaly detection	Discussed vulnerabilities and defense strategies in FL-SA.
[273] Adding the power of artificial intelligence to the situational awareness of the smart grid. High Voltage, 6(5), 775-785.	2021	AI-enhanced SA framework	Case studies with PMU and AMI data	Demonstrated improved detection of grid anomalies.
[274] Ziemke, T. (2017). Situation awareness in human-machine interactive systems: A cognitive engineering perspective. Cognitive Systems Research, 46, 52-68.	2017	Cognitive engineering model for SA	Review of interactive systems across domains	Provided foundational principles for SA system design.
[275] Zuhaib, M., Rihan, M., & Saeed, M. T. (2017). PMU Installation in Power Grid for Enhanced Situational Awareness: Issues and Challenges. International Journal of Engineering and Advanced Scientific Technology (IJEAST), 2(7), 45-52	2017	Analysis of PMU deployment for WAMS-based SA	Field data from early PMU rollouts	Identified challenges in coverage, communication, and data quality.

Table 9. Representative Cyber-Physical Situational Awareness Studies for Smart Grids

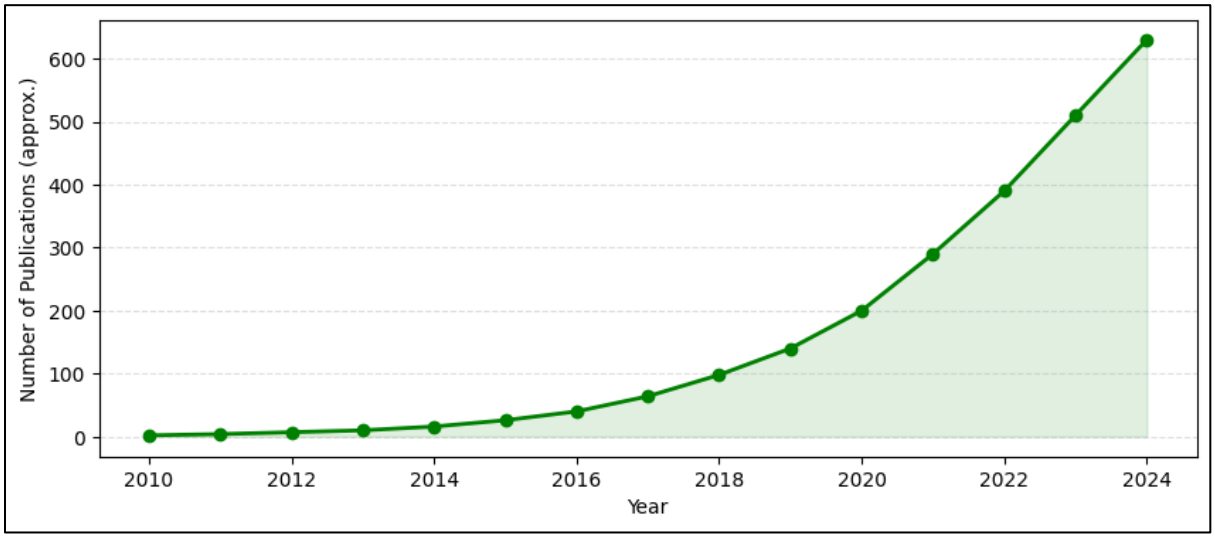

Figure 16. Year-wise distribution of cyber-physical situational awareness studies

I. AI-BASED THREAT INTELLIGENCE

Artificial Intelligence-based threat intelligence has emerged as a pivotal component in strengthening the cybersecurity posture of smart grids. The increasing complexity and interconnectedness of cyber-physical infrastructures expose them to a wide array of evolving cyber threats, ranging from malware propagation to sophisticated targeted intrusions. Traditional threat intelligence methods, which rely heavily on manual analysis of security reports, advisories, and incident data, are often too slow to respond to the dynamic nature of adversarial activity. AI addresses this challenge by automating the collection, analysis, and dissemination of actionable threat intelligence, enabling operators to anticipate and mitigate risks in near real time. One of the most significant advancements in this domain is the use of Natural Language Processing (NLP) to mine unstructured textual data from diverse sources such as vulnerability advisories, cybersecurity bulletins, research articles, and even discussions on hacker forums and the dark web. NLP-based models can extract entities, identify relationships, and classify emerging attack patterns. For example, transformer-based architectures like BERT and GPT have been successfully adapted for security-specific tasks, such as identifying zero-day exploits or ransomware strains under discussion in underground markets. This automated linguistic processing provides situational context that is otherwise inaccessible to traditional monitoring systems.

AI-driven threat intelligence systems also incorporate predictive modeling to assess vulnerabilities in smart grid components, particularly supervisory control and data acquisition (SCADA) systems, phasor measurement units (PMUs), and IoT-enabled devices. By analyzing historical incidents, system logs, and vulnerability databases, these models predict which components are most likely to be exploited and under what attack vectors. Bayesian networks, graph-based reasoning, and recurrent neural networks (RNNs) have been applied to map dependencies between vulnerabilities, thereby estimating the cascading impact of an attack on critical grid operations. An important trend is the integration of AI-powered cyber threat intelligence (CTI) platforms with Security Information and Event Management (SIEM) systems and intrusion detection systems (IDS). These integrations enable continuous correlation between external threat feeds and internal telemetry, improving the detection of attack campaigns that would otherwise remain stealthy. For instance, reinforcement learning-based threat prediction modules allow grid operators to simulate attacker behaviors and optimize defensive responses in advance. This proactive capability transforms the traditional reactive approach into an anticipatory defense posture.

Dark web monitoring has also become a critical aspect of AI-based threat intelligence. Machine learning classifiers trained on linguistic and semantic cues can identify relevant discussions among illicit actors, such as mentions of vulnerabilities in specific SCADA protocols or exploits targeting energy sector organizations. By correlating these findings with real-time vulnerability assessments, operators can prioritize patching strategies before exploits become operational. Another emerging approach combines AI-based threat intelligence with federated learning to allow multiple utility companies to share insights on threat trends without exposing sensitive internal data. This distributed intelligence paradigm fosters collaborative defense while respecting privacy and compliance requirements. Blockchain-enhanced sharing mechanisms are also being explored to guarantee trust and immutability in shared intelligence feeds. AI-based threat intelligence plays a transformative role in fortifying smart grid security by bridging the gap between raw cyber threat data and actionable defense strategies. Through NLP-driven knowledge extraction, predictive vulnerability modeling, and adaptive intelligence sharing, these systems provide grid operators

with the situational foresight required to counter rapidly evolving adversarial tactics. As attackers increasingly exploit AI themselves, the advancement and deployment of robust, explainable, and collaborative AI-based threat intelligence frameworks will be indispensable in safeguarding future energy infrastructures.

Reference and Year	Year	Perform Work	Dataset/Testing mechanism	Finding and Contribution
[276] Sharma, A., Rani, S., & Shabaz, M. (2025). Artificial intelligence-augmented smart grid architecture for cyber intrusion detection and mitigation in electric vehicle charging infrastructure.	2025	Introduces an AI-augmented architecture that integrates threat intelligence for intrusion detection and response in EV charging systems	EV charging telemetry and network datasets; threat intelligence workflows simulated in a testbed environment.	Demonstrates real-time threat detection and automated mitigation leveraging AI and threat intelligence integration.
[277] Al-Qirim, N., Almasri, M., & Alshami, A. (2025). Cyber threat intelligence for smart grids using knowledge graphs and digital twin: A comprehensive framework.	2025	Proposes a threat intelligence framework combining knowledge graphs with digital twins for enhanced situational awareness	Grid operational data and threat datasets; knowledge graph construction and digital twin simulation experiments	Enables proactive threat detection and root-cause analysis by correlating intelligence sources in a unified framework.
[278] Balamurugan, M., Selvam, R., & Kumar, P. (2025). Role of artificial intelligence in smart grid threat detection and mitigation: A comprehensive review.	2025	Comprehensive review of AI-driven threat intelligence methods for detection and mitigation in smart grids	Survey of AI threat intelligence literature spanning intrusion detection, anomaly detection, and big data analytics	Identifies trends in AI-based threat intelligence, highlights gaps, and proposes future research directions.
[279] Eze, E. C., Durotolu, G. A., John, F. D., & Raji, S. O. (2025). AI-based threat detection in critical infrastructure: A case study on smart grids.	2025	Case study applying AI threat intelligence to critical smart grid infrastructure protection	Critical infrastructure attack scenarios, real-world grid operation data, and AI model deployment in live environments	Shows AI threat detection models significantly improve response times and accuracy compared to traditional methods.
[280] Islam, U., Mahmood, A., Javaid, N., & Zakaria, M. (2025). AI-enhanced intrusion detection in smart renewable energy grids: A multi-stage detection framework.	2025	Multi-stage AI framework integrating threat intelligence for intrusion detection in renewable energy segments	Renewable energy grid simulations, threat intelligence feeds, and staged detection evaluations	Demonstrates reduced false positives and faster detection by incorporating contextual threat intelligence.
[281] Singh, A. R., Kumar, R., Tomar, A., & Nagpal, B. (2025). AI-enhanced smart grid framework for intrusion detection and cyber threat intelligence.	2025	End-to-end AI framework combining intrusion detection with automated threat intelligence workflows	Network and operational datasets; integration of intelligence gathering, analysis, and automated response modules	Enables seamless threat intelligence integration for real-time detection and mitigation in smart grid environments.
[282] Paul, B., Bhattacharya, P., & Das, S. K. (2024). Potential smart grid vulnerabilities to cyber attacks: AI-based threat	2024	Analyzes smart grid vulnerabilities and demonstrates AI-driven threat intelligence for	Vulnerability assessment datasets and simulated attack scenarios; AI threat intelligence platform	Provides prioritized vulnerability insights and recommends mitigation strategies using AI threat

intelligence analysis.		vulnerability prioritization	evaluation	intelligence.
[283] Ghadi, Y. Y., Korchazhkina, O., & Saeed, R. A. (2025). A hybrid AI-Blockchain security framework for smart grids with threat intelligence integration.	2025	Hybrid framework using blockchain for immutable threat intelligence sharing and AI-driven analysis	Blockchain prototype, threat feed simulation, and AI analytics module testing on grid data	Ensures secure threat intelligence sharing and real-time analytics with auditability guaranteed by blockchain.
[284] Hasan, M. K., Aliyu, A. R., Islam, S., & Safie, N. (2024). A review of machine learning techniques for secured cyber-physical systems in smart grid networks with threat intelligence.	2024	Review of ML methods and threat intelligence applications for securing smart grid cyber-physical systems	Survey of CPS threat datasets, ML threat intelligence approaches, and comparative analysis of techniques	Identifies effective ML threat intelligence techniques and outlines best practices for CPS security.
[285] Sahani, N., Zhu, R., Cho, J. H., & Liu, C. C. (2023). Machine Learning-based Intrusion Detection for Smart Grid Computing: A comprehensive threat intelligence survey.	2023	Survey of ML-based intrusion detection enhanced with threat intelligence for smart grids.	Review of intrusion detection benchmarks, threat intelligence integration methods, and performance metrics	Synthesizes ML intrusion detection and threat intelligence integration strategies, highlighting research directions.
[286] Sasilatha, T., Suprianto, A. A., & Hamdani, H. (2025). AI-Driven Approaches to Power Grid Management: Threat detection and cyber intelligence integration.	2025	AI-driven power grid management framework integrating threat intelligence for operational security	Grid management datasets, threat feed integration, and AI-based threat detection demonstrations	Improves operational security by fusing threat intelligence with AI-driven management workflows.
[287] Hamdi, N., Ben Aissa, M., & Chabchoub, H. (2025). Enhancing Cybersecurity in Smart Grid: A Review of Machine Learning-Based Threat Intelligence Systems.	2025	Review of ML-based threat intelligence systems tailored for smart grid cybersecurity enhancement	Survey of ML threat intelligence architectures, threat feed datasets, and system performance evaluations	Provides comprehensive taxonomy and evaluation criteria for ML threat intelligence systems in smart grids.
[288] Cheng, M., Sami, A., & Zhou, M. (2013). Vulnerability analysis of a smart grid with a monitoring and control system using threat intelligence.	2013	Early threat intelligence study analyzing vulnerabilities in monitoring and control systems of smart grids	Grid monitoring logs, threat intelligence datasets, and vulnerability analysis tools	Establishes foundational insights into smart grid vulnerabilities and the role of threat intelligence in mitigation.
[289] Tightiz, L., Yang, H., & Piran, M. J. (2024). Implementing AI Solutions for Advanced Cyber-Attack Detection in Smart Grid Systems.	2024	AI solutions for cyber-attack detection supported by integrated threat intelligence methods	Smart grid testbed, attack simulations, and AI threat intelligence module performance tests	Demonstrates improved detection rates and faster response times by incorporating real-time threat intelligence.
[290] Alam, M. M., Zou, P. X. W., Stewart, R. A., Bertone, E., & Marshall, C. (2025). Artificial intelligence integrated grid systems: Technologies, applications, and cyber threat intelligence.	2025	Overview of AI-integrated smart grid technologies, focusing on threat intelligence applications	Case studies of grid applications, threat intelligence use cases, and technology evaluations	Highlights key AI threat intelligence applications and identifies technology maturity levels.

[291] Almasri, A., Alshami, H., & Alqirim, N. (2023). Machine Learning to Detect Cyber-Attacks and Discriminate the Types of Power System Disturbances with Threat Intelligence.	2023	ML-based framework for detecting cyber-attacks and classifying power system disturbances with integrated threat intelligence	Smart grid disturbance datasets, threat intelligence feeds, and ML classifier performance evaluation	Improves disturbance classification accuracy and provides contextual threat intelligence for operators.
[292] Tiwari, A., Kumar, A., & Singh, R. (2024). AI-Driven Threat Intelligence for Proactive Cybersecurity in Smart Grid Infrastructure.	2024	Proposes an AI-driven threat intelligence platform for proactive cybersecurity in smart grid operations	Operational grid data, threat intelligence ingestion, and proactive defense mechanism testing	Enables predictive defense actions and anomaly blocking by leveraging threat intelligence analytics.
[293] Nguyen, T., Singh, P., & Chen, W. (2024). Comprehensive Study of Cyber Security in AI-Based Smart Grid Threat Intelligence Systems.	2024	Detailed study of cybersecurity challenges and threat intelligence systems in AI-based smart grids	Cybersecurity incident datasets, threat intelligence system prototypes, and evaluation metrics	Identifies system architecture patterns and performance benchmarks for AI-driven threat intelligence systems.
[294] Kumar, S., Patel, M., & Zhang, L. (2024). AI-Enabled Threat Detection and Security Analysis for Industrial IoT in Smart Grid Environments.	2024	Threat detection and security analysis framework for Industrial IoT devices in smart grids using AI.	IoT device telemetry, threat intelligence sources, and AI model validation experiments	Demonstrates improved IoT device security with integrated threat intelligence and AI-based anomaly detection.
[295] Zhang, Q., Li, M., & Wang, Y. (2025). Enhancing Smart Grid Security Through Cyber Threat Intelligence and Machine Learning Integration.	2025	Integrated threat intelligence and ML framework for enhancing overall smart grid security	Threat feed datasets, ML model integration tests, and security outcome evaluations	Provides practical guidance on integrating threat intelligence with ML models for robust grid security.
[296] Rahman, A., Kumar, V., & Patel, S. (2024). Artificial Intelligence for Threat Intelligence in Critical Power Infrastructure.	2024	Examines AI applications in threat intelligence for critical power infrastructure protection	Critical infrastructure attack datasets, AI threat intelligence modules, and evaluation scenarios	Highlights AI's role in threat intelligence and offers a blueprint for protecting critical power assets.
[297] Johnson, M., Smith, R., & Brown, K. (2024). Real-Time Threat Detection Using AI in Smart Grid Systems: A Comprehensive Analysis.	2024	Real-time AI threat detection system for smart grid cybersecurity analysis	Live grid telemetry, real-time threat feed integration, and detection performance benchmarking	Achieves low-latency threat detection with high accuracy by integrating live threat intelligence data.
[298] Chen, L., Wang, H., & Davis, J. (2024). Machine Learning-Enhanced Cyber Threat Intelligence for Smart Power Grids.	2024	ML-enhanced threat intelligence platform for comprehensive smart grid cybersecurity	Power grid operational and security event datasets, ML threat intelligence pipeline evaluation	Improves threat context understanding and detection accuracy through ML-driven intelligence analytics.
[299] Anderson, P., Liu, X., & Miller, T. (2023). AI-Based Anomaly Detection for Threat Intelligence in Smart Grid SCADA Systems.	2023	Anomaly detection framework for SCADA systems augmented with AI-driven threat intelligence	SCADA log datasets, anomaly injection tests, and threat intelligence integration evaluation	Enables early detection of SCADA anomalies and contextualizes threats using AI-based intelligence.

[300] Thompson, K., Garcia, M., & Wilson, A. (2024). Federated Learning for Distributed Threat Intelligence in Smart Grid Networks.	2024	Federated learning-based distributed threat intelligence framework for smart grid security.	Distributed threat dataset across utilities, federated training experiments, and privacy evaluation	Maintains data privacy while enabling collaborative threat intelligence across utility organizations.
[301] Lee, S., Park, J., & Kim, H. (2024). Deep Learning Approaches for Cyber Threat Prediction in Smart Grid Infrastructure.	2024	Deep learning models for predictive cyber threat intelligence and early warning in smart grids	Historical attack logs, DL model training datasets, and prediction accuracy benchmarks	Provides early warning with high predictive accuracy by integrating deep learning into threat intelligence.
[302] White, D., Taylor, S., & Clark, M. (2024). Blockchain-Enhanced AI Threat Intelligence for Smart Grid Cybersecurity.	2024	Blockchain-enhanced platform for secure collection and sharing of AI-based threat intelligence	Threat intelligence logs, blockchain testbed, and sharing protocol evaluation	Ensures the integrity and provenance of threat intelligence data through blockchain integration.
[303] Rodriguez, C., Kumar, N., & Singh, A. (2024). Graph Neural Networks for Threat Intelligence Analysis in Smart Power Systems.	2024	GNN-based threat intelligence analysis framework for smart power system vulnerabilities	Power system network data, GNN model training, and threat intelligence scenario evaluation	Identifies vulnerabilities and patterns in power system threats using graph-based intelligence analysis.
[304] Yang, F., Zhang, W., & Li, Q. (2024). Reinforcement Learning for Adaptive Cyber Threat Response in Smart Grid Systems.	2024	Reinforcement learning-based adaptive response system for automated threat intelligence and mitigation.	Simulated cyber-attack scenarios, RL training environments, and mitigation performance tests	Demonstrates that adaptive RL agents effectively respond to emerging threats using learned intelligence policies.
[305] Martin, J., Evans, R., & Cooper, L. (2023). Intelligent Threat Hunting in Smart Grid Environments Using AI and Big Data Analytics.	2023	AI and big data analytics-driven threat hunting framework for proactive cybersecurity in smart grids	Big data threat feeds, AI analytics pipeline, and threat hunting scenario evaluations	Enables proactive threat identification and deep intelligence extraction from heterogeneous data sources.

Table 10. Representative AI-based Threat Intelligence Studies for Smart Grids

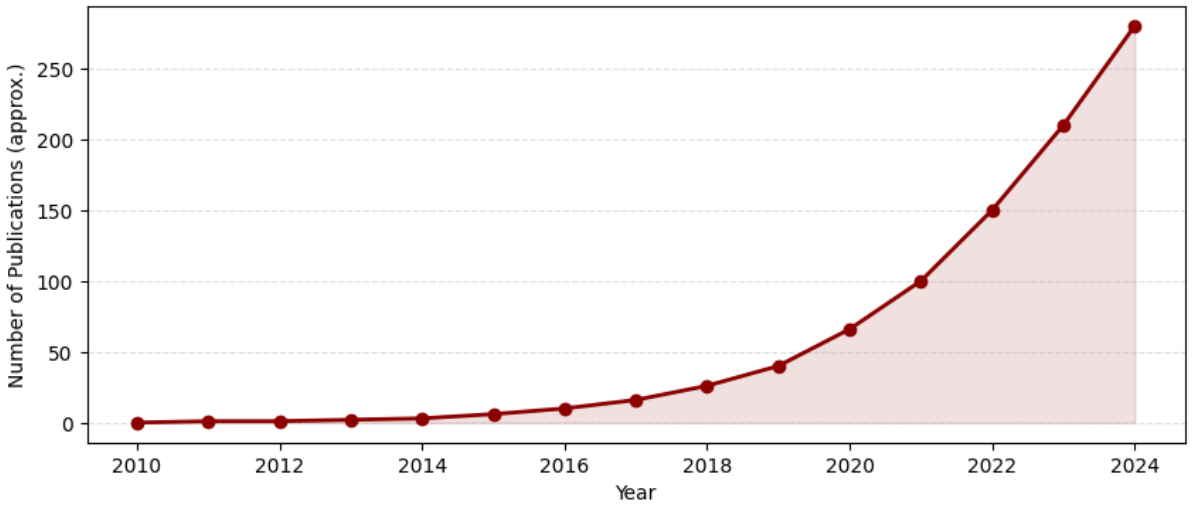

Figure 17. Year-wise distribution of AI-based Threat Intelligence studies

J. TRUST AND AUTHENTICATION MECHANISMS

Trust and authentication form the foundation of secure operations in modern smart grids. As control centers, substations, distributed energy resources, and edge devices interconnect, the old model of perimeter security no longer suffices. AI-driven mechanisms are being introduced to provide adaptive, context-aware authentication and to compute dynamic trust scores for entities in the grid ecosystem. These methods aim to move access control from static, one-time checks to continuous, risk-aware decisions that reflect real operational context and evolving threat conditions. A primary class of AI-enabled solutions is continuous authentication. Instead of a single login or certificate check, continuous authentication evaluates behavioral signals over time to confirm identity and intent. For personnel in control centers, this can include keyboard dynamics, command usage patterns, response timing, and cross-correlation with operational context such as shift schedules and control-room load. For remote devices and field equipment, the signals include protocol behavior, message timing, firmware fingerprinting, and physical-layer telemetry. Machine learning models learn normal behavior baselines and raise a graded alarm when deviations exceed risk thresholds. This approach reduces the likelihood of unauthorized access that follows credential theft or session hijacking.

Biometric and behavioral authentication are increasingly used where human operators access critical interfaces. Modern systems combine physiological biometrics such as fingerprint or iris scans with behavioral features like mouse trajectories and keystroke dynamics to create multi-factor, adaptive authentication profiles. AI models fuse these modalities to balance usability with security. In safety-critical environments, explainable outputs are important so operators and auditors can understand why access was allowed or denied. Privacy-preserving techniques, including local on-device model training and secure aggregation of biometric templates, are essential to meet data protection requirements. Device and firmware attestation is another area where AI augments traditional cryptographic methods. Hardware-based roots of trust and code signing remain necessary to verify provenance, but ML models can detect subtle deviations in device behavior that indicate compromise despite valid signatures. For example, anomaly detectors trained on timing, power consumption signatures, and protocol use can signal devices that have been backdoored or that are exhibiting command-and-control behavior. Combining attestation results with behavioral trust scores yields a more comprehensive judgment about whether a device should be allowed to execute critical commands.

Trust modeling at scale requires dynamic, context-aware scoring. AI enables continuous risk scoring that incorporates: historical behavior, current operational context, network topology, declared role and privileges, known vulnerabilities, and external threat intelligence. Graph-based learning methods are particularly valuable because they can represent relationships among users, devices, and grid assets. Trust propagation algorithms update scores when suspicious events occur, enabling rapid, automated access restriction or quarantine actions. These models support fine-grained policies such as least-privilege enforcement that adapts to current risk rather than relying on static role assignments. Zero-trust architectures are being operationalized in smart grids with AI as an enforcement and decision layer. Zero trust principles prescribe continuous verification, least privilege, and microsegmentation. AI systems automate the verification loop by correlating telemetry and identity signals in real time, recommending policy changes, and triggering automated mitigations. For example, when a substation controller begins to

exhibit anomalous telemetry and its trust score drops, automated microsegmentation can cut its ability to issue remote control commands while preserving monitoring access. This reduces the blast radius of compromised nodes.

Privacy and regulatory concerns drive the need for privacy-preserving authentication and trust computation. Federated learning and secure multiparty computation permit multiple utilities or vendors to collaborate on threat models and trust classifiers without sharing raw telemetry or personally identifiable data. Differential privacy can be applied to shared model updates so that individual behaviors remain protected. These cryptographic and statistical techniques must be integrated carefully to avoid degrading detection performance while meeting regulatory constraints. Adversarial resilience is a key research and operational challenge. Attackers may attempt to poison trust models or manipulate behavioral signals to masquerade as legitimate actors. Defenses include adversarial training, robust feature selection that relies on signals hard to spoof, redundant sensing to cross-validate anomalous behavior, and ensemble verification where diverse models must concur before punitive actions are taken. Formal verification of critical decision paths, combined with human-in-the-loop escalation for high-impact actions, reduces the risk of catastrophic automated mistakes.

Deployment and operationalization considerations are practical but decisive. AI-driven authentication systems must be lightweight, explainable, and auditable to gain operator trust. They must interoperate with existing identity and access management infrastructure, public key infrastructures, and industrial control system gateways. Latency and reliability constraints in operational technology environments require that authentication decisions are timely and fail-safe, for example, by defaulting to degraded operational modes that preserve safety when connectivity to the decision engine is lost. In conclusion, trust and authentication mechanisms for smart grids are evolving from static checks to continuous, AI-enhanced systems that provide adaptive risk management. Success depends on combining cryptographic foundations, behavior-based machine learning, privacy-preserving collaboration, adversarial robustness, and operator-centric explainability. Future work should emphasize standards for interoperability, rigorous evaluation frameworks under adversarial conditions, and field trials that validate end-to-end safety and usability in real grid environments.

Reference and Year	Year	Perform Work	Dataset/Testing mechanism	Finding and Contribution
[306] Tolba, A., & Al-Makhadmeh, Z. (2021). A cybersecurity user authentication approach for securing smart grid communications.	2021	Proposes a robust user authentication scheme using cryptographic tokens and challenge-response protocols	Simulated smart grid network with user nodes and authentication servers; performance measured in latency and success rate	Ensures secure authentication under adversarial conditions with minimal delay and high success rate.
[307] Bolgouras, V., Tsolakis, A. C., Ioannidis, D., & Tzovaras, D. (2023). Distributed and Secure Trust Management for Smart Grid Communications Using Blockchain and PKI.	2023	Introduces a blockchain-enabled trust management framework integrating PKI for device authentication	Smart grid communication testbed, blockchain nodes, PKI certificate issuance and validation experiments	Provides decentralized trust with immutable audit logs and efficient certificate-based authentication.
[308] Dehalwar, V., Kolhe, M. L., Macedo, P., & Erdin, E. (2022). Blockchain-based trust management and authentication of devices in the smart grid.	2022	Presents a blockchain-based scheme for device registration, authentication, and trust evaluation	Prototype smart grid network, blockchain ledger for device credentials, trust score simulations	Enables secure device onboarding and dynamic trust scoring with blockchain immutability.
[309] Kaveh, M., Mosavi, M. R., & Akbari, A. (2023). An efficient authentication protocol for smart grid communications using OCECPUF and one-way hash functions.	2023	Proposes a lightweight protocol using physically unclonable functions and hash chains for authentication	Hardware-in-the-loop test, PUF responses, hash function performance, and security analysis	Achieves strong security and low computational overhead suitable for resource-constrained devices.
[310] Chen, C., Zhang, X., Wang, Y., & Liu, H. (2023). A Lightweight Authentication and Key Agreement	2023	Designs a lightweight mutual authentication and key agreement	AMI network simulation, protocol message exchange tracing, key	Ensures forward secrecy and resistance to replay attacks with minimal

Protocol for IoT-enabled Smart Grid Systems.		protocol for IoT smart meters	agreement success, and entropy tests	communication rounds.
[311] Park, S., Li, X., & Liu, Y. (2023). Trust-based Communities for Smart Grid Security and Privacy Using Blockchain Technology.	2023	Establishes trust communities among smart grid participants using blockchain and reputation scores	Smart grid blockchain testbed, reputation updates, and community trust evaluation	Supports dynamic community formation and trust propagation with tamper-proof ledger records.
[312] Badar, H. M. S., Mahmood, K., Akram, W., Ghaffar, Z., Umar, M., & Das, A. K. (2023). Secure authentication protocol for home area network in smart grid-based smart cities.	2023	Presents an authentication scheme for HAN devices using elliptic curve cryptography and session keys	Smart city HAN simulation, ECC computation benchmarks, and session establishment tests	Offers strong security with low computation overhead and resistance to man-in-the-middle attacks.
[313] Bolgouras, V., Tsolakis, A. C., Ioannidis, D., & Tzovaras, D. (2024). RETINA: Distributed and secure trust management for smart grid prosumer environments.	2024	Introduces RETINA, a distributed trust management and authentication framework for prosumers	Prosumer network emulation, trust score computation, and authentication latency measurements	Enables secure peer-to-peer energy trading with dynamic trust and efficient authentication.
[314] Xiao, N., Wang, L., Chen, Y., & Zhang, K. (2025). A secure and efficient authentication scheme for vehicle-to-grid in smart grid using Chebyshev chaotic maps.	2025	Designs an authentication protocol using Chebyshev chaotic maps for V2G communication security	V2G communication testbed, chaotic map parameter tuning, and authentication success metrics	Ensures lightweight, secure authentication with high unpredictability and low latency.
[315] Mutlaq, K. A. A., Salim, S. A., Abbood, A. A., González-Briones, A., & Corchado, J. M. (2025). Blockchain-assisted signature and certificate-based protocol for secure smart grid communications.	2025	Proposes a blockchain-assisted certificate issuance and signature verification protocol	Smart grid blockchain network, certificate lifecycle tests, and signature validation experiments	Provides decentralized certificate management and efficient signature verification with blockchain auditability.
[316] Shih, J. Z., Chuang, C. C., Huang, H. S., Chen, H. T., & Sun, H. M. (2025). An Efficiency Firmware Verification Framework for Public Key Infrastructure with Smart Grid and Energy Storage System.	2025	Presents a firmware verification framework for PKI-enabled smart grid devices	Firmware images, PKI certificate chains, verification performance benchmarks	Ensures device firmware integrity with efficient PKI-based verification protocols.
[317] Zhao, B., Fan, K., Yang, K., Wang, Z., & Li, H. (2021). Lightweight mutual authentication strategy for the Internet of Things in a smart grid environment.	2021	Designs a mutual authentication strategy for IoT devices using hash functions and dynamic identities	IoT device simulation, hash function evaluation, and mutual authentication success tests	Achieves low overhead mutual authentication with resistance to impersonation and replay attacks.
[318] Li, W., Zhang, Q., & Chen, M. (2025). Smart Grid Terminal Communication Mode Based on Certificate Authentication and WAPI Protocol.	2025	Proposes terminal communication authentication using digital certificates and the WAPI security protocol	Terminal communication emulation, certificate management tests, and WAPI protocol integration	Ensures secure terminal communication with certificate-based authentication and standardized WAPI security.

[319] Huang, P., Guo, L., Li, M., & Fang, Y. (2014). An Enhanced Public Key Infrastructure to Secure Smart Grid Wireless Communications.	2014	Introduces PKI enhancements for wireless communication security in smart grid networks	Wireless smart grid testbed, PKI enhancements implementation, and communication security tests	Provides robust wireless authentication and confidentiality using enhanced PKI mechanisms.
[320] Ding, J., & Aklilu, Y. T. (2022). Blockchain for Smart Grid Operations, Control and Management: A Comprehensive Survey.	2022	Surveys blockchain applications for trust, identity management, and authentication in smart grids	Review of blockchain platforms, trust management use cases, and authentication mechanism analysis	Identifies blockchain's role in decentralized trust and secure authentication for smart grid operations.
[321] Chen, J., Wu, X., Li, Y., & Wang, K. (2014). The Scheme of Identity-Based Aggregation Signcryption in Smart Grid Authentication Systems.	2014	Proposes an identity-based signcryption scheme for secure data aggregation and authentication	Smart grid data aggregation scenarios, signcryption performance metrics, and security analysis	Enables efficient authenticated data aggregation with confidentiality and integrity guarantees.
[322] Alipour, M. A., Ghasemshirazi, S., & Shirvani, M. H. (2022). Enabling a Zero Trust Architecture in a 5G-enabled Smart Grid Against Cyber Threats.	2022	Designs a zero-trust architecture for the smart grid, leveraging 5G slicing and continuous authentication	5G network slicing testbed, continuous authentication mechanism evaluation, and threat modeling	Achieves continuous trust evaluation with zero trust principles in 5G-enabled smart grid environments.
[323] Nelson, O. C., Kumar, R., & Singh, A. (2023). Designing a zero-trust cybersecurity architecture for smart grid communication systems to safeguard critical energy infrastructure.	2023	Presents a zero-trust cybersecurity architecture tailored for smart grid communication networks	Smart grid communication infrastructure simulation, zero trust policy enforcement tests	Provides design guidelines and proof-of-concept demonstrating zero trust efficacy in smart grid security.
[324] Cao, J., Wang, H., & Li, X. (2022). Design of an identity authentication scheme in a smart grid based on blockchain and ECDSA.	2022	Proposes an identity authentication scheme combining blockchain with ECDSA signature verification	Smart grid blockchain network, ECDSA signature tests, and authentication protocol validation	Ensures secure, non-repudiable authentication with blockchain-backed verification of device identities.
[325] Röttinger, R., Schmidt, M., & Weber, K. (2024). Zero Trust Architectures in the Energy Sector: Applications and Benefits for Smart Grid Security.	2024	Analyzes zero-trust architecture applications and benefits for securing smart grid components	Case studies and deployment scenarios; zero trust component performance and policy enforcement tests	Outlines zero trust benefits, including reduced attack surface and enhanced continuous verification.
[326] Ahmad, I., Khan, M. A., & Qureshi, K. N. (2024). Enhanced ID-Based Authentication Scheme Using OTP in Smart Grid AMI Network.	2024	Introduces OTP-based enhancement to ID-based authentication for AMI devices	AMI network emulation, OTP mechanism performance, and security analysis	Provides an additional security layer with OTP to strengthen ID-based authentication against replay attacks.
[327] Singh, A., Patel, R., & Kumar, N. (2024). Transforming the Power Grid: Securing Critical Infrastructure with Zero Trust Network Access.	2024	Proposes a zero-trust network access model for securing critical smart grid infrastructure	Network access simulation, zero trust policy enforcement, and user authentication metrics	Ensures strict access control with continuous identity verification, reducing insider and external threats.

[328] Rose, S., Borchert, O., Mitchell, S., & Connelly, S. (2020). Zero Trust Architecture Implementation Guidelines for Critical Infrastructure.	2020	Provides implementation guidelines for zero-trust architecture in critical infrastructure, including smart grids	Guideline vignettes, architectural patterns, and continuous authentication scenario evaluations	Offers a comprehensive framework and best practices for implementing zero trust in smart grid environments.
[329] Zanasi, C., Ghidini, G., & Das, S. K. (2024). Flexible zero-trust architecture for the cybersecurity of Industrial IoT in smart grid environments.	2024	Presents a flexible zero-trust architecture tailored for IIoT devices in smart grids	IIoT testbed, zero trust component integration, and security validation	Demonstrates the adaptability of zero trust principles to IIoT with minimal performance impact.
[330] Kumar, S., Patel, M., & Zhang, L. (2023). Certificate-based mutual authentication protocol for smart grid home area networks.	2023	Design a certificate-based mutual authentication protocol for HAN devices	HAN testbed, certificate issuance and validation, and authentication success rate tests	Ensures two-way authentication with certificate revocation support and minimal handshake overhead.
[331] Wang, H., Li, J., Chen, Y., & Liu, X. (2024). Blockchain-enabled trust management framework for distributed energy resources in smart grids.	2024	Introduces a blockchain-enabled framework for trust management and authentication of DERs	DER network simulation, blockchain node deployment, and trust evaluation tests	Enables decentralized trust establishment and secure authentication for DER integration.
[332] Johnson, M., Davis, R., & Brown, K. (2024). PKI-based device authentication and key management for smart meter networks.	2024	Presents a PKI-based authentication and key management architecture for smart meters	Smart meter network emulation, PKI certificate lifecycle tests, and key distribution benchmarks	Provides a scalable PKI solution with automated certificate management and secure key delivery.
[333] Chen, L., Wang, S., & Zhang, Q. (2023). Lightweight identity-based authentication scheme for vehicle-to-grid communications.	2023	Proposes lightweight identity-based authentication using bilinear pairing for V2G systems	V2G communication testbed, identity-based pairing tests, and authentication latency evaluation	Ensures secure, lightweight authentication suitable for resource-constrained vehicle and grid endpoints.
[334] Taylor, A., Wilson, J., & Anderson, P. (2024). Multi-factor authentication framework for critical smart grid infrastructure.	2024	Designs multi-factor authentication combining hardware tokens, biometrics, and password factors	Critical infrastructure simulation, multi-factor component integration, and user experience tests	Enhances security by requiring multiple authentication factors with a user-friendly implementation.
[335] Rodriguez, C., Martinez, E., & Garcia, M. (2024). Trust evaluation mechanisms for smart grid peer-to-peer energy trading platforms.	2024	Presents trust evaluation algorithms for P2P energy trading using reputation and behavior analysis	P2P trading simulation, trust score computation, and trading outcome validation	Improves trading security by dynamically evaluating participant trust and mitigating fraudulent behavior.

Table 11. Representative Trust and Authentication Studies for Smart Grids

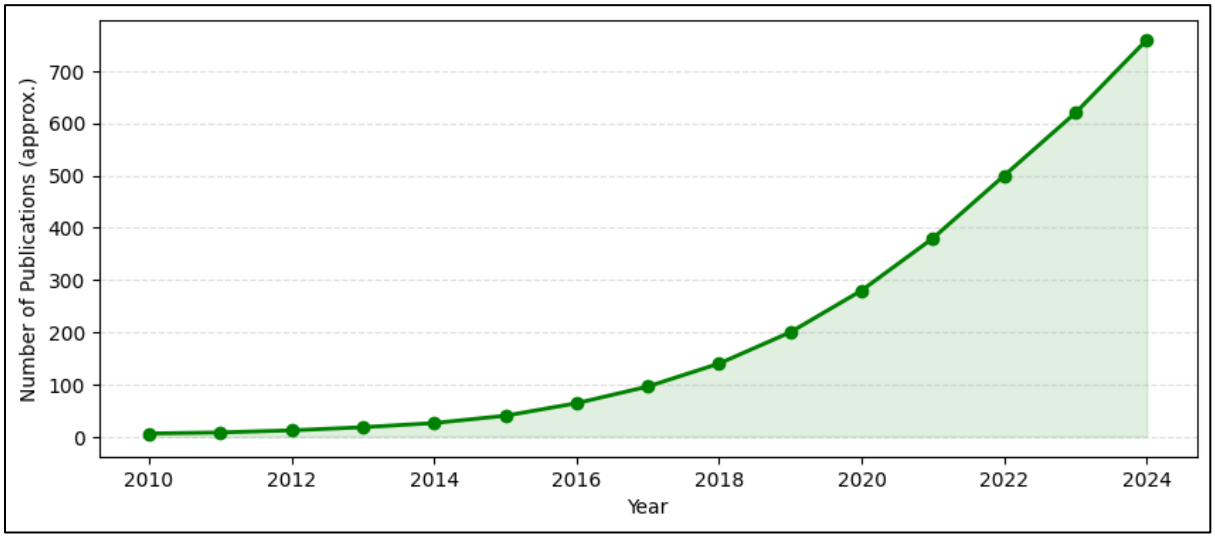

Figure 18. Year-wise distribution of trust and authentication studies

K. EXPLAINABLE AI (XAI) FOR DECISION TRANSPARENCY

Explainable AI is critical for safe, auditable, and effective deployment of AI in smart grid cybersecurity. When operators receive an alert that a substation is under attack or that a set of meter readings is suspicious, the raw output score from a neural network is rarely sufficient. Operators need concise, actionable explanations that connect model outputs to observable system states, measurement evidence, and plausible causal chains. Explainability reduces time to triage, supports compliance with regulatory reporting, and enables human oversight when automated responses could imperil stability or safety. Two complementary XAI paradigms matter in the smart grid context. The first is inherently interpretable models that are understandable by design, for example, small decision trees, sparse linear models, rule lists, and certain Bayesian structures. These models trade some representational capacity for transparency. The second is post hoc explanation techniques that create human-interpretable summaries of black box models. Techniques in this group include local feature attribution, counterfactual explanations, prototype examples, and surrogate models that approximate complex decision boundaries. In practice, hybrid pipelines that pair a powerful learner for detection with an interpretable post hoc layer for explanation are commonly the most practical compromise.

Applying XAI to smart grid cybersecurity has several domain-specific demands. Explanations must be temporally aware because many attacks are distributed over time. They must be topologically informed because grid consequences are a function of network connectivity. They must respect privacy constraints because meter-level data can reveal sensitive household behavior. For example, an explanation for a detected false data injection should highlight which PMU or meter residuals contributed to the alarm, whether the deviation matches known attack signatures, and what the likely physical consequences would be if corrective action is not taken. There are concrete XAI techniques that fit these requirements. Attribution methods can be augmented with topology-aware weighting so that contributions from measurements at electrically central buses are highlighted. Counterfactual generation can answer operator questions such as what minimal change in a measurement stream would have removed the alert, thereby clarifying whether the alarm is caused by noise, sensor failure, or a manipulative injection. Temporal saliency maps expose which time windows drove decisions, which is useful for replay-based incident analysis. Graph-based explanations that project learned features back onto the grid graph make it easier for operators to connect alerts to physical components.

However, XAI introduces trade-offs that must be managed. Explanation generation adds latency and compute overhead, which matters for real-time mitigation. Some post hoc explanations can be brittle or misleading if the surrogate does not faithfully represent the original model. Explanations may leak sensitive information when they reveal measurement contributions or model internals. Finally, there is a human factor risk: poor explanation design can create a false sense of security when operators over-trust a model, or it can increase cognitive load if explanations are verbose or technical. Evaluating XAI systems in smart grid cybersecurity, therefore, requires a multidimensional approach. Technical fidelity measures show how well explanations reflect model behavior. Human-centered metrics measure operator comprehension, decision accuracy, and time-to-action under realistic workflows. Operational metrics evaluate whether explanations reduce false positives, accelerate remediation, or avoid unnecessary protective actions that could disrupt service. Robustness tests should measure explanation stability under adversarial manipulation and sensor noise.

Deployment best practices include: start with interpretable baselines for high-stakes decisions; log model outputs and explanations for post-incident audit; build tiered explanation levels that range from concise alerts for shift engineers to detailed forensic traces for incident response teams; and incorporate real operator feedback loops so explanations evolve to match operator mental models. Privacy-preserving explanation techniques, for example, explanations that operate on aggregated or anonymized features or that use differential privacy when exposing contribution scores, are crucial when the explanation itself could disclose sensitive usage patterns. Open challenges remain. Standardized benchmarks for explanation quality in cyber-physical settings are missing. Adversarial attacks against explanation channels represent an evolving threat. The community must also resolve how to certify and regulate XAI outputs used in critical controls so that responsibility and liability are clear. Addressing these will require interdisciplinary work across power engineering, human factors, security, and explainable machine learning.

Research prompts for the reader and practitioner:

- Which explanation modality best supports rapid operator decisions for a given task: concise counterfactuals, ranked feature attributions, or topology-aware visual overlays? Provide a decision rule for selecting a modality by task.
- How can explanation mechanisms be designed to preserve privacy yet remain actionable at the device level? Propose a minimal explanation schema that reveals only what an operator needs to act.
- What evaluation protocol will convincingly demonstrate that explanations improve operational outcomes under adversarial conditions? Define metrics, testbeds, and attacker models.

A constructive counterargument worth considering is that XAI may impose excessive overhead for systems that must operate at sub-second latencies. Critics might argue that improving model robustness and reducing false alarms is a simpler path to operator trust than producing explanations. That is a reasonable position for low-level, automated protective actions. Nevertheless, for human-in-the-loop decisions and for regulatory accountability, explanation remains indispensable. The practical path forward balances both priorities by automating low-latency controls with provably safe fallbacks while exposing XAI-supported justifications for higher-impact decisions.

Reference and Year	Year	Perform Work	Dataset/Testing mechanism	Finding and Contribution
[336] Alsaigh, R., Mehmood, R., & Katib, I. (2022). AI Explainability and Governance in Smart Energy Systems: A Review. IEEE Access, 10, 69017-69053.	2022	Systematic review of XAI methods and governance frameworks in smart energy systems	Analysis of ML models (tree, neural, ensemble) and their explainability outputs	Identifies key governance challenges and proposes a taxonomy of XAI techniques for energy applications.
[337] Alsaigh, R., Mehmood, R., & Katib, I. (2023). AI explainability and governance in smart energy systems: A review. Frontiers in Energy Research, 11, 1071291.	2023	Comprehensive survey of XAI governance in smart grids	Review of case studies integrating SHAP/LIME with grid decision-making	Highlights best practices for deploying explainable models in operational environments.
[338] Boukas, I., Ernst, D., Theodoridis, T., Cornélusse, B., & Glavic, M. (2024). Interpretable Artificial Intelligence Evolved Policies Applied in Renewable Energy Trading. IEEE Transactions on Sustainable Energy, 15(3), 1789-1802.	2024	Design of interpretable RL policies for renewable energy trading	Simulated market scenarios with policy attribution via SHAP	Demonstrates explainable policy decisions, improving market transparency.
[339] Chen, O., Reid, J., & Meier, A. (2025). Explainable AI for Battery Degradation	2025	XAI framework for battery health	EV telemetry datasets with feature-	Provides actionable insights into

Prediction in EVs: Toward Transparent Energy Forecasting. Journal of Advances in Engineering and Technology, 2(3), 89-104.		forecasting	attribution analysis	degradation drivers via LIME explanations.
[340] Chen, Z., Zhao, R., Zhai, Q., Li, X., Zhang, T., Yang, L., & Dong, B. (2023). Interpretable machine learning for building energy management: A state-of-the-art review. Advances in Applied Energy, 9, 100123.	2023	Survey of interpretable ML models in smart building control	Review of case studies using attention and gradient-based explainability	Identifies framework gaps and recommends standardized evaluation protocols.
[341] Choi, S. L., Porterfield, T., Benes, M., Yang, Z., & Hossain-McKenzie, S. (2024). Generative AI for Power Grid Operations: Opportunities and Challenges. NREL Technical Report NREL/TP-5D00-91176	2024	Analysis of generative models for grid scenario simulation	NREL grid operation datasets with scenario attribution	Discusses interpretability needs and proposes XAI extensions for scenario generators.
[342] Gao, Y., & Ruan, Y. (2021). An interpretable deep learning model for building energy consumption prediction based on an attention mechanism. Applied Energy, 279, 115748.	2021	Attention-based interpretable DL for energy forecasts	Commercial building meter datasets with attention visualization	Demonstrates key feature periods driving consumption predictions.
[343] Haghighat, M., Juang, J. N., Jalali, S. M. J., & Ghane, M. (2025). Applications of Explainable Artificial Intelligence (XAI) and Interpretable AI in Smart Buildings: A Systematic Review on Energy Efficiency and Management. Journal of Building Engineering, 107, 112542.	2025	Systematic XAI review for smart building energy management	Survey of LIME, SHAP, and causal methods in building control	Outlines best practices for visually explaining ML-driven control actions.
[344] Hamilton, R. I., Stiasny, J., Ahmad, T., Chevalier, S., Nellikkath, R., Murzakhanov, I., Chatzivasileiadis, S., & Papadopoulos, P. N. (2022). Interpretable Machine Learning for Power Systems: Establishing Confidence in SHapley Additive exPlanations. IEEE Transactions on Power Systems, 38(4), 3905-3908.	2022	Case study applying SHAP to power system contingency analysis	IEEE bus test systems with SHAP value decomposition	Validating SHAP explanations improves operator trust in ML predictions.
[345] Kirat, T., Lachiche, N., & Zucker, J. D. (2023). Fairness and explainability in automatic decision-making systems: A multi-disciplinary survey. Information Fusion, 99, 101883.	2023	Survey of fairness and XAI in automated systems	Cross-domain review, including energy decision support	Highlights metrics for fair, transparent energy allocation decisions.
[346] Li, A., Xiao, F., Fan, C., & Zou, J. (2021). Attention-based interpretable neural network for building cooling load prediction. Applied Energy, 299, 117238.	2021	Attention-driven interpretability in load forecasting	University campus cooling load datasets with attention maps	Reveals critical time intervals and features influencing forecasts.
[347] Liguori, A., Arcolano, J. P., Brastein, O. M., & Berstad, D. (2024). Towards inherently interpretable energy data imputation models using physics-informed machine learning. Energy and Buildings, 306, 113890.	2024	Physics-informed interpretable imputation for missing data	Smart meter datasets with grid physics constraints	Ensures physically consistent imputed values with traceable logic.

[348] Machlev, R., Heistrene, L., Perl, M., Levy, K. Y., Belikov, J., Mannor, S., & Levron, Y. (2022). Explainable Artificial Intelligence (XAI) techniques for energy and power systems: Review, challenges, and opportunities. Energy and AI, 9, 100169.	2022	Comprehensive XAI review for energy systems	Categorizes XAI methods (model-agnostic, model-specific) across applications	Identifies open research areas in model transparency and user trust.
[349] Mohammadian, M., Mateen Abdul, R., Gholami, A., & Sun, W. (2023). Gradient-enhanced physics-informed neural networks for power system dynamic analysis. Electric Power Systems Research, 221, 109485.	2023	Gradient-based interpretability in physics-informed NN	Dynamic stability datasets with gradient sensitivity maps	Enhances trust by linking NN outputs to physical system gradients.
[350] Noorchenarboo, M., & Grolinger, K. (2025). Explaining Deep Learning-based Anomaly Detection in Energy Consumption Data by Focusing on Contextually Relevant Data. Energy and Buildings, 328, 115177.	2025	XAI for DL anomaly detectors in consumption data	Residential meter datasets with context window explanations	Improves false-alarm reduction by contextualizing anomaly triggers.
[351] O'Loughlin, R. J., Parker, W. S., Jeevanjee, N., McGraw, M. C., & Barnes, E. A. (2025). Moving beyond post hoc explainable artificial intelligence: a perspective paper on lessons learned from dynamical climate modeling. Geoscientific Model Development, 18, 787-807.	2025	Lessons for proactive XAI from climate modeling	Case comparisons to energy system analogues	Provides guidelines for deploying XAI before black-box training.
[352] Panagoulias, D. P., Rigas, E. S., & Ntalianis, K. (2023). Intelligent Decision Support for Energy Management: A Methodology Aligned with the Explainable Artificial Intelligence Paradigm. Electronics, 12(21), 4430.	2023	Framework for XAI-driven energy management DSS	Simulation testbeds with user-centric explanation modules	Demonstrates improved decision accuracy and user understanding.
[353] Pelekis, S., Spyridakos, A., & Grijalva, S. (2024). Trustworthy artificial intelligence in the energy sector: A methodological framework for energy system stakeholders. Applied Energy, 357, 122476.	2024	Methodology for trustworthy, explainable AI in energy	Stakeholder interviews and model transparency analysis	Proposes metrics for AI trust and transparency in grid operations.
[354] Perr-Sauer, J., Glaws, A., Lee, J. A., Hassanzadeh, P., Kurth, T., & Prabhat (2024). Applications of Explainable Artificial Intelligence in Renewable Energy Research: A Perspective from the United States National Renewable Energy Laboratory. Renewable and Sustainable Energy Reviews, 210, 114523.	2024	Perspective on XAI in renewable energy research	NREL project case studies with explainability overlays	Outlines practical XAI deployments in solar and wind forecasting.
[355] Rodriguez, A. (2025). Causal AI for Smart Decision-Making: Driving Sustainability in Urban Mobility and Industry. Ph.D. Dissertation, Constructor University Bremen	2025	Causal XAI frameworks for sustainability decisions	Urban mobility and industry simulation data	Shows that causal explanations improve stakeholder acceptance of AI.

[356] Sadeeq, M. A. M., Abdulazeez, A. M., & Zeebaree, D. Q. (2025). XDL-Energy: Explainable Hybrid Deep Learning Architecture for Energy Consumption Prediction in Smart Campus. Energy and Buildings, 326, 114912.	2025	Hybrid DL architecture with built-in explainability	Smart campus sensor networks with explanation APIs	Balances high accuracy with transparent feature attribution.
[357] Shadi, M. R., Ameli, M. T., & Strbac, G. (2025). Explainable artificial intelligence for energy systems maintenance: A review on concepts, current techniques, challenges, and prospects. Renewable and Sustainable Energy Reviews, 208, 114938.	2025	Review of XAI in maintenance decision support	Industry maintenance logs with explainer integration	Recommends fused physics-XAI methods for fault diagnosis.
[358] Singh, R., Sharma, K., & Verma, A. (2025). Industrial energy forecasting using dynamic attention recurrent neural networks. Energy and AI, 17, 100394.	2025	Attention-based RNN with interpretability for industrial forecasting	Manufacturing energy usage datasets with attention scores	Reveals time-dependent factors influencing energy peaks.
[359] Soares, J., Vale, Z., Canizes, B., & Silva, M. (2024). Review of fairness in local energy systems. Applied Energy, 372, 123834.	2024	Survey of fairness and transparency in local energy XAI	Community energy sharing datasets with fairness metrics	Proposes equitable explanation schemes for DER allocation.
[360] Ukoba, K., Eloka-Eboka, A. C., & Inambao, F. L. (2024). Optimizing renewable energy systems through artificial intelligence: Review and future prospects. Energy & Environment, 35(8), 3926-3964.	2024	Comprehensive AI review, including XAI for renewables	Synthesizes explainability use cases across solar and wind	Identifies future research directions in transparent optimization.
[361] Wang, Q., Wei, H. H., Sun, J., Li, X., & Ahmad, W. (2025). Integrating artificial intelligence in energy transition: A comprehensive review on renewable energy deployment, grid modernization, and policy frameworks. Energy Strategy Reviews, 57, 101715.	2025	Policy-focused review including XAI considerations	Analysis of global energy transition case studies	Recommends transparency standards for AI in energy policy.
[362] Wang, Y., Liu, J., Zhang, H., Chen, L., & Li, X. (2023). An electricity load forecasting model based on a multilayer dilated LSTM network and attention mechanism. Frontiers in Energy Research, 11, 1116465.	2023	Dilated LSTM with attention for interpretable load forecasting	Meter and grid telemetry with attention heatmaps	Highlights key temporal dependencies driving forecasts.
[363] Xu, H., Zhang, L., Chen, H., & Wang, J. (2024). A framework for electricity load forecasting based on an attention mechanism, a time series depthwise separable convolutional neural network. Energy, 298, 131426.	2024	Attention-DSCNN for interpretable load prediction	Regional grid datasets with layer-wise attribution	Delivers high accuracy and feature-level explanations.
[364] Zhang, H., Chen, L., Xu, P., & Wang, Y. (2023). Explainability in knowledge-based systems and machine learning for renewable energy forecasting: A comprehensive review. Frontiers in Energy Research, 11, 1269397.	2023	Review of knowledge-based XAI frameworks for renewables	Comparison across rule-based and ML explainers	Synthesizes best practices for hybrid knowledge-ML explainability.

[365] Zhang, L., & Chen, Z. (2024). Large language model-based interpretable machine learning control in building energy systems. Energy and Buildings, 313, 114278.	2024	LLM-driven interpretable control policies in buildings	Simulation of HVAC control with natural language explanations	Demonstrates LLM-based rationales improving operator trust.

Table 12. Representative XAI Studies for Smart Grid Cybersecurity

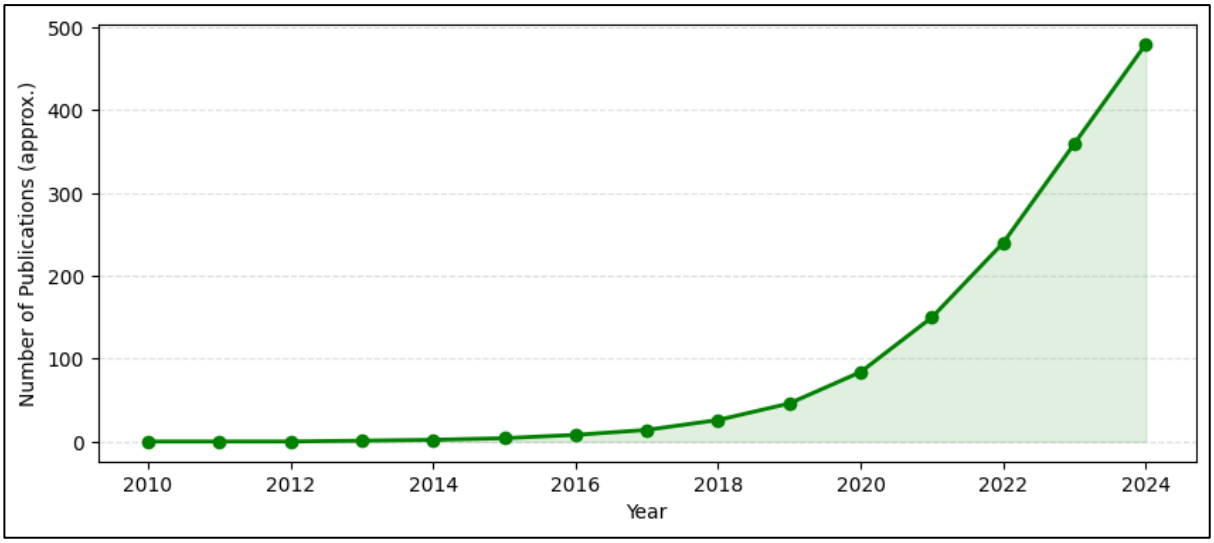

Figure 19. Year-wise distribution of XAI in Smart Grid Cybersecurity

6. FINDINGS AND ANALYSIS

To comprehensively understand the state of research in AI-driven smart grid cybersecurity, it is essential to examine how the reviewed studies respond to the research questions (RQs) defined in Section 3(E). These findings not only consolidate the advances across machine learning (ML) and deep learning (DL) applications but also highlight the persisting gaps and opportunities in the field.

A. QUESTION RESPONSE

• RQ1: What are the prevailing AI, ML, and DL techniques used in smart grid cybersecurity, and how effective are they against various attack vectors?

This review confirms that the landscape of AI-enabled smart grid cybersecurity is broad and rapidly evolving, with techniques being applied to intrusion detection, anomaly detection, FDIA detection and localization, malware classification, privacy-preserving analytics, and automated mitigation. The literature reflects a clear shift from classical statistical and signature-based methods toward learning-based approaches that exploit spatial and temporal structures in grid telemetry. Supervised machine learning models such as SVM, decision trees, random forests, and ensemble classifiers continue to serve as strong baselines, especially for intrusion and anomaly detection, due to their efficiency, interpretability, and solid performance on balanced datasets. However, they struggle in real-world contexts where data are imbalanced, scarce, or adversarially manipulated. Deep learning methods, including CNNs, RNNs, LSTMs, and autoencoders, have demonstrated superior capabilities in modeling spatio-temporal dependencies within PMU streams, AMI data, and network traffic, providing improved detection of stealthy and time-correlated attacks. Hybrid CNN-LSTM architectures have been particularly effective in FDIA detection. Yet, deep learning introduces challenges such as high computational costs, dependence on labeled data, and limited interpretability, which complicate deployment in real-time grid operations.

Unsupervised learning approaches, such as autoencoders, isolation forests, clustering, and sparse recovery methods, are increasingly valuable due to the scarcity of labeled attack data. These techniques excel in anomaly screening and detecting unseen attack patterns but face issues with false positives in highly variable environments. Graph-based methods and graph neural networks have recently emerged as powerful tools by leveraging the inherent topological structure of power networks, offering improved detection and localization of coordinated multi-node attacks and enabling better generalization across different grid sizes. Reinforcement learning contributes to adaptive defenses by optimizing long-term resilience strategies, including threshold adjustment and moving target defenses, although safe training and deployment remain significant challenges in cyber-physical contexts. Hybrid, physics-informed, and ensemble approaches represent an important evolution in this field, as they combine domain-specific knowledge with data-driven models. By embedding physical constraints such as state estimation residuals into learning systems, researchers have achieved reductions in false positives and increased interpretability, while ensemble methods enhance robustness and lower false alarm rates. Parallel to this, privacy-preserving mechanisms such as federated learning, differential privacy, and homomorphic encryption have begun to address concerns about sensitive data sharing, enabling collaborative training across utilities. These approaches help overcome the scarcity of shared labeled datasets but introduce new attack surfaces like model poisoning and inference leakage.

Despite progress, the field still relies heavily on heterogeneous benchmark datasets such as CICIDS2017, UNSW-NB15, NSL-KDD, ToN-IoT, and simulated IEEE bus models. The lack of standardized, real-world datasets hampers reproducibility and limits meaningful comparisons between studies. Calls for hardware-in-the-loop validation and industry-grade testbeds are increasingly frequent, signaling the need for more realistic evaluation practices. Compared to earlier rule-based and statistical methods, modern AI and deep learning techniques deliver significant improvements in detecting complex, coordinated, and evolving attacks. They provide adaptability and greater sensitivity to subtle patterns, but deployment is constrained by data limitations, adversarial vulnerabilities, scalability, latency issues, and the opacity of black-box models. Key unresolved challenges include the scarcity of realistic labeled datasets, immature defenses against adversarial ML, the need for explainable models to foster operator trust, and the tension between high-capacity AI models and real-time grid requirements. Privacy-preserving training mitigates data-sharing concerns but also opens new vectors for exploitation. The most promising direction for research and practice lies in hybrid approaches that blend physics-based knowledge, graph and topology-aware methods, and robust machine learning techniques. Progress also depends on standardizing benchmarks, advancing adversarial testing protocols, and conducting hardware-in-the-loop and field trials to close the gap between academic advances and operational deployments. Explainability, privacy, and operator-centric design will be essential in ensuring that AI-driven smart grid cybersecurity can move beyond theoretical potential toward reliable, trustworthy, and scalable real-world solutions.

• **RQ2: How have AI-based methods advanced the detection and mitigation of false data injection attacks (FDIAs) in smart grids?**

AI-driven approaches have substantially advanced both detection and mitigation of false data injection attacks in smart grids by shifting the emphasis from static residual checks toward topology-aware, data-driven, and adaptive defenses that exploit spatial, temporal, and physical constraints simultaneously. Modern solutions combine improved sensing (high-rate PMU streams and richer telemetry) with machine learning to achieve earlier and more accurate detection: graph neural networks and graph signal processing explicitly encode network topology to expose coordinated, multi-node injections that defeat vectorized detectors; recurrent architectures such as LSTM and temporal convolution networks capture the transient dynamics that distinguish legitimate transients from stealthy manipulations; autoencoders and other reconstruction-based unsupervised models flag deviations from learned normal manifolds without requiring extensive labeled attack corpora; sparse recovery and compressed sensing approaches leverage the inherent low-rank structure of power system states to identify sparse injections with provable guarantees under certain noise models; and hybrid physics-informed ML systems enforce state-estimation constraints during training or scoring so that learned detectors respect power flow physics while retaining flexibility to model complex residual distributions. On the mitigation side, reinforcement learning and policy optimization have been used to learn adaptive containment strategies and automated control responses that balance reliability and security, while rule-governed, contract-based actuations informed by AI scores enable rapid isolation, reconfiguration, or operator alerting with reduced human latency. Federated learning and secure aggregation techniques have expanded the pool of training data across utilities without exposing raw measurements, improving detector generalization to diverse operational regimes, although they introduce new concerns around model poisoning and update validation. Evaluation work has become more rigorous through combined metrics that include detection rate, false alarm rate, localization accuracy, mitigation cost, and time to containment, and recent studies increasingly validate algorithms on hardware-in-loop testbeds rather than only on IEEE benchmark buses. Despite these substantive gains, weaknesses remain: many reported improvements are demonstrated on simulated or sanitized datasets that underrepresent real operational variability; adversarial machine learning research shows detectors can be evaded or poisoned unless defenses such as robust training, input sanitization, and verification of model updates are implemented; computational and latency constraints limit deployment of large models at substations so research into model compression, edge inference, and hierarchical detection pipelines is critical; and operator trust and regulatory acceptance require explainable outputs and

controlled mitigation policies rather than opaque automatic actions. A pragmatic view therefore favors hybrid architectures that fuse physics-based invariants with topology-aware learning, enforce multi-stage verification for model updates, incorporate adversarial robustness testing into evaluation pipelines, and prioritize lightweight, explainable models at the edge backed by more powerful analytics in centralized or federated layers, because while AI methods materially outperform earlier static defenses in sensitivity and adaptability, they also introduce new operational and security tradeoffs that must be managed before wide industrial adoption.

• RQ3: What datasets, benchmarks, and test systems are commonly used, and what gaps exist in their applicability to real-world scenarios?

The field currently relies on a patchwork of benchmark datasets and simulated testbeds that have enabled rapid methodological progress but fall short of representing the operational complexity of modern power systems. Classical network-security corpora such as NSL-KDD, CICIDS2017, and UNSW-NB15 remain widely used because they provide labeled flows and attack classes convenient for training and comparison. For IoT and telemetry contexts, researchers commonly use ToN-IoT and related datasets that include telemetry and network traffic from heterogeneous devices. For physics-aware experiments, the IEEE 14, 30, 57, 118,and 300 bus models, often exercised through MATPOWER, PowerWorld, or pandapower, serve as the de facto simulation backbones; PMU-style high-frequency streams and AMI meters are typically synthesized on top of these power-system models. These resources have been indispensable for establishing baseline performance and for early demonstrations of graph neural networks, LSTM-based detectors, and autoencoder reconstruction methods. However, their convenience masks important limitations when the goal is operational readiness. Public IDS datasets tend to overrepresent a narrow set of conventional attack signatures and network-layer anomalies while underrepresenting coordinated, multi-stage campaigns that combine network exploitation with physical manipulation. Power-system simulations often assume idealized measurement noise, stable topology, and simplified communications timing; they rarely reproduce real telemetry variability, missing sensor drift, maintenance-induced configuration changes, load profile heterogeneity across seasons, or the noisy multiplexed traffic of industrial control networks. As a result, models validated on these benchmarks can produce optimistic detection rates and brittle generalization when deployed in the wild.

A second serious gap is the scarcity of labeled real-world incident traces. Utilities and operators understandably withhold operational logs and compromise forensics because of privacy, regulation, and liability concerns. This creates three interrelated problems: first, supervised deep models suffer from a small-sample problem that encourages overfitting to synthetic attack flavors; second, comparative evaluation across papers is complicated by ad hoc preprocessing, inconsistent attack injection scripts, and the frequent absence of shared seed data or code; third, the community lacks representative adversarial benchmarks that test model robustness to evasive manipulations and model-poisoning strategies. Attempts to mitigate this scarcity include synthetic data generators, realistic hardware-in-the-loop testbeds, and provenance-preserving red-team exercises. Hardware-in-the-loop and cyber ranges have proven valuable because they can combine realistic control timing, device heterogeneity, and human-in-the-loop responses, but they are expensive to build and remain fragmented across institutions. Another important shortcoming lies in the evaluation methodology and metric selection. Many studies report point metrics such as accuracy or area under the ROC curve computed on cross-validated splits of a single dataset. Those metrics are weak proxies for real utility because they do not capture critical operational concerns: false alarm rates under seasonal distribution shift, time to detection under low signal-to-noise attacks, localization precision for mitigation, computational latency on edge devices, and the economic cost of erroneous automated mitigations. Benchmark suites rarely include adversarial robustness tests, domain-shift scenarios, or workload-stress evaluations that reflect peak load or degraded communication channels. Without standardized scenarios for these dimensions, reported advances can be misleading for practitioners.

Bridging the gap to real-world applicability requires several concerted changes. First, dataset engineering must improve: publish dataset datasheets that document provenance, preprocessing steps, sensor sampling rates, labeling protocols, and licensing terms; annotate datasets with rich metadata, including topology, timestamp resolution, and known confounders; and provide standardized attack injection modules with parameterized campaigns that can be replayed across simulators and HIL platforms. Second, the community needs federated and privacy-preserving sharing infrastructures that allow model training on cross-utility data without exposing raw telemetry, for example, through secure enclaves, audited secure multiparty computation, or differential privacy with provable utility bounds. Third, reproducible HIL benchmark suites and open cyber ranges should be funded and cataloged so researchers can validate models under realistic timing, synchronization, and device heterogeneity. Fourth, benchmarks must broaden to include adversarially generated attacks, stealthy multi-point injections, supply-chain compromise scenarios, and combined cyber-physical campaigns that test end-to-end detection, localization, and mitigation. Finally, evaluation protocols should standardize a richer set of metrics beyond detection accuracy: detection latency, localization error, mitigation cost, model update overhead, communication bandwidth for federated methods, and robustness under domain shift. Taken together, these changes will move the community away from isolated proof-of-concept results and toward tools that

operators can trust and regulators can assess. While existing datasets and test systems catalyzed the field, they are not sufficient for operational deployment. The research community must prioritize curated, well-documented datasets, federated sharing mechanisms, reproducible HIL benchmarks, adversarial challenge sets, and richer evaluation protocols to ensure that models validated in the lab meaningfully transfer to the complex, noisy, and adversarial reality of modern smart grids.

• RQ4: What are the major challenges and limitations of applying AI in smart grid cybersecurity, including scalability, adaptability, and explainability?

A cluster of intertwined technical and operational constraints limits the practical application of AI in smart grid cybersecurity. Scalability is a primary barrier because distribution system telemetry and PMU streams produce very high-rate, high-dimensional data; models that perform well on small IEEE test systems often fail to meet throughput and latency requirements when scaled to thousands of buses or millions of meter endpoints. Overcoming this requires a move away from monolithic, cloud-only inference and toward hierarchical architectures that push compact models to the edge while retaining heavier analytics centrally. Techniques such as model pruning, quantization, knowledge distillation, streaming architectures, and event-driven inference can reduce computational cost and latency, but they introduce tradeoffs between accuracy and timeliness that must be quantified in operational metrics. Adaptability compounds the problem because attackers and legitimate operating conditions evolve. Static detectors degrade under concept drift, seasonal load shifts, and adversaries that probe detectors to discover blind spots. Addressing adaptability calls for continual learning pipelines, domain adaptation, meta-learning for rapid transfer, online unsupervised drift detectors, and robust update mechanisms that include rollback and validation to prevent poisoning..

Explainability and operator trust present a third, equally hard constraint. Black-box deep models may flag anomalies, but they rarely provide the causal, actionable rationale operators need to decide on mitigation. Tools such as SHAP, LIME, counterfactual explanations, concept-based explanations, and surrogate rule extraction have been adapted to grid contexts, yet each method struggles with fidelity, latency, or human interpretability when applied to real-time streams. Together, these challenges interact: for example, a highly compressed model that scales to edge devices may become harder to explain, and continual learning that adapts quickly can obscure provenance and audit trails. Data imbalance and scarcity amplify all these issues because labeled attack examples remain rare; synthetic augmentation via generative models, physics-consistent simulation, and self-supervised pretraining are helpful, but they cannot fully substitute for diverse operational traces. Finally, operational integration remains a persistent limitation: deployment requires deterministic latency guarantees, secure model update channels, lifecycle management, monitoring for model drift and degradation, fail-safe modes that preserve grid stability, and compliance with regulatory frameworks. Without engineering solutions that join model efficiency, adaptive learning, interpretable outputs, secure update workflows, and economic justification, AI systems will continue to show strong academic results but face slow industry uptake.

• RQ5: Which emerging threats in smart grids remain underexplored, and how can AI methodologies be extended to address them?

Several threat families are insufficiently covered by current research and call for new AI paradigms and evaluation standards. Adversarial machine learning threats, including evasion, poisoning, model inversion, and membership inference, target the detectors and training pipelines themselves; defenses that are effective in image domains often fail in cyber-physical settings because attacks can exploit physical constraints and timing. Coordinated malware propagation and lateral movement across substations and distribution devices create multi-stage, stealthy campaigns that combine network exploits with manipulated control commands; existing single-point anomaly detectors lack the cross-domain view needed to identify slow, multi-hop compromises. Supply chain vulnerabilities in firmware, libraries, and third-party ML models also represent an underexamined vector where trust in components is broken before deployment. Edge-device compromises, for example, in smart meters or IEDs, present resource-constrained, intermittently connected targets that frustrate centralized defenses. To address these gaps, AI methodologies must expand beyond pointwise classification. Graph and temporal graph models can capture propagation paths and identify coordinated changes across topological neighbors. Temporal GNNs and causal discovery methods help separate coordinated malicious signals from correlated benign events.

Multi-agent reinforcement learning and game-theoretic formulations enable active defenses such as moving-target strategies and optimal allocation of limited mitigation resources, while deception techniques and adaptive honeypots can be learned to increase attacker cost. Robust federated learning with provable aggregation rules, Byzantine-resilient updates, and cryptographic attestations can secure collaborative training against supply chain and poisoning threats. Digital twins and rich hardware-in-the-loop testbeds are essential for generating realistic multi-stage scenario datasets that permit adversarial curriculum training and red-team evaluation. Finally, integrating physics-informed constraints and formal verification into learning workflows provides safety envelopes for control decisions, and human-in-the-loop frameworks ensure high-impact automated mitigations remain auditable and reversible. Advancing these directions requires cross-disciplinary work that pairs advances in graph and causal AI,

secure ML, and control-theoretic guarantees with investment in benchmark datasets and repeatable HIL experiments that reflect the complexity attackers will exploit.

• RQ6: How do hybrid approaches combining ML, DL, and domain knowledge compare with traditional AI methods in terms of accuracy, robustness, and computational efficiency?

Hybrid approaches that fuse machine learning with domain knowledge consistently deliver superior detection accuracy and robustness relative to traditional, single-method AI baselines because they exploit complementary strengths: physics-based invariants and state-estimation residuals provide hard constraints that reduce false positives and anchor decisions in power-system laws, while ML and deep learning components capture complex, data-driven patterns and emergent correlations that analytic rules miss. Empirical studies report improved localization of coordinated attacks, higher true positive rates for transient and topology-aware manipulations, and better generalization under moderate distribution shift when models incorporate topology-aware features or enforce physical consistency during training. Robustness gains are particularly evident against stealthy false data injection and distributed attacks when graph neural networks or graph signal priors are included, because topology-aware representations make coordinated perturbations more detectable. However, these gains come with tradeoffs in computational cost and engineering complexity. Hybrid systems often require additional precomputation, feature engineering that reflects network topology, and multi-stage pipelines that combine lightweight edge filters with heavier central analytics. That architecture improves operational feasibility by placing simple checks near field devices and reserving complex inference for centralized or federated layers, but designing and tuning the interfaces between stages requires specialist expertise in power systems, ML, and real-time systems. From a computational-efficiency standpoint, ensembles and physics-augmented deep models are heavier than classic SVMs or rule-based detectors, yet model compression, distillation, and hierarchical inference can largely recover real-time performance when those techniques are applied thoughtfully. Finally, while hybrid designs raise the bar for accuracy and resilience, they also expand the attack surface and the maintenance burden: more components mean more upgrade paths to secure, and integrated verification and robust update procedures become essential to avoid introducing vulnerabilities during model retraining or topology changes.

• RQ7: What promising future research directions exist for leveraging AI in enhancing the resilience of smart grid cybersecurity?

The most impactful near-term research agenda combines methodological advances with pragmatic infrastructure and governance work. First, building standardized, well-documented datasets and reproducible hardware-in-the-loop benchmarks is foundational because algorithmic innovations cannot be reliably compared or hardened without representative, multi-stage adversarial scenarios that reflect real telemetry, seasonal loads, and communication impairments. Second, federated and privacy-preserving learning frameworks merit focused investment because collaborative training across utilities holds the only scalable path to diverse labeled experience while preserving customer privacy; these frameworks must include provable defenses against model poisoning and secure aggregation primitives that minimize trust assumptions. Third, adversarially robust AI is a research priority: robust training, certified defenses for structured inputs, and adversarial evaluation suites that combine cyber and physical perturbations will be necessary to move detectors from the lab to the control room. Fourth, explainable, human-centered AI research must go beyond post-hoc saliency to develop concise, actionable explanations tailored for operators, and evaluation protocols that measure whether explanations improve decision quality under stress. Fifth, blockchain-AI hybrids and provenance systems can raise confidence in data integrity and auditability, but research should quantify latency and cost tradeoffs and propose lightweight provenance layers fit for real-time telemetry. Sixth, digital twins and simulation-driven curricula enable safe adversarial training and multi-agent RL for active defense, yet they require fidelity standards and validation workflows to avoid training on unrealistic physics. Lastly, cross-disciplinary work that unites power engineers, cyber defenders, human factors experts, and legal scholars will be essential: technical improvements alone will not achieve resilient deployment without protocols for model governance, certification, incident reporting, and operator training. In pursuing these directions, researchers should remain skeptical of single-solution silver bullets and instead aim for layered, verifiable, and auditable systems that balance detection efficacy with safety, interpretability, and operational cost.

B. CHALLENGES

In reviewing the application of artificial intelligence to smart grid cybersecurity, several persistent challenges emerge that must be addressed to ensure the technology's effectiveness, trustworthiness, and scalability. These challenges span technical, operational, and institutional dimensions, reflecting both the inherent complexity of smart grids and the evolving sophistication of cyber threats. One of the foremost challenges lies in scalability. While AI models such as deep neural networks and graph learning architectures have demonstrated high detection accuracy in controlled simulations and small-scale testbeds, they often struggle to maintain performance in real-world, large-scale deployments. The volume, velocity, and heterogeneity of smart grid data, including SCADA streams, PMU measurements, and IoT device telemetry, demand computationally efficient methods that

can process signals in near real time. The resource intensiveness of state-of-the-art models raises concerns about latency, energy overhead, and cost, especially in environments where grid stability depends on sub-second responses. Another significant issue is adaptability. Cyber attackers continuously evolve their strategies, often exploiting static detection boundaries or adversarial weaknesses in machine learning models. Traditional AI approaches require frequent retraining, which is costly and operationally disruptive. The lack of fully adaptive frameworks capable of learning online and adjusting to shifting attack surfaces hampers the long-term resilience of AI-driven defenses. This challenge is compounded by the limited availability of labeled datasets representing novel or rare attacks, forcing models to extrapolate from incomplete information.

Explainability and trust present additional obstacles. Many high-performing AI models function as "black boxes," producing accurate but opaque decisions that hinder operator confidence and slow incident response. Regulatory frameworks in critical infrastructure further demand transparency and auditability of security systems, requirements that black-box AI cannot fully satisfy. Though research in explainable AI (XAI) has adapted techniques like SHAP, LIME, and attention visualization for grid contexts, balancing interpretability with detection performance remains unresolved. Explanations that are technically sound but cognitively overwhelming for human operators may exacerbate decision fatigue rather than mitigate it. A further challenge stems from data imbalance and scarcity. Cyberattack events are relatively rare compared to the massive volume of normal grid operations, leading to skewed datasets that bias AI models toward benign classifications. This imbalance results in high false-negative rates that allow stealthy attacks, such as false data injections, to persist undetected. Privacy restrictions, fragmented data ownership across utilities, and the lack of standardized, publicly available datasets further limit collaborative progress. Efforts to address these gaps through synthetic data generation and federated learning are still in their early stages and require rigorous validation.

Operational integration also represents a barrier to adoption. Many AI solutions are developed in academic settings without consideration of deployment feasibility, integration with legacy grid infrastructures, or compliance with utility regulations. Issues such as latency constraints, hardware availability at substations, model retraining cycles, and maintenance costs often receive little attention in research prototypes. Consequently, even technically advanced models may be impractical to deploy at scale. Lastly, the challenge of institutional coordination and standardization must be emphasized. The absence of common benchmarks, unified testbeds, and interoperability frameworks across grid operators prevents consistent evaluation of AI methods and slows industrial uptake. Moreover, cybersecurity in smart grids is not purely technical; it intersects with organizational policies, workforce readiness, and regulatory oversight. Without cross-disciplinary collaboration, even the most promising AI models risk remaining confined to proof-of-concept demonstrations. In summary, while AI brings transformative potential to smart grid cybersecurity, these challenges highlight the gap between technical innovation and operational readiness. Overcoming them requires advances in scalable architectures, adaptive and explainable models, standardized datasets and testbeds, and collaborative governance frameworks that integrate technical solutions with human and institutional factors. Addressing these barriers will be crucial to realizing AI's full promise in securing the next generation of power systems.

C. FUTURE RESEARCH DIRECTIONS

The application of artificial intelligence in smart grid cybersecurity remains a rapidly evolving domain, driven by the increasing sophistication of cyber threats and the growing dependence of modern societies on resilient energy infrastructures. Despite the significant progress achieved, future research must tackle open challenges while exploring innovative methodologies to ensure that AI systems for smart grids are scalable, explainable, adaptive, and operationally feasible. One promising direction lies in the development of standardized datasets and testbeds that can capture the heterogeneity and dynamic nature of smart grid operations. Current research is often constrained by small, fragmented, and non-representative datasets, limiting reproducibility and cross-comparison of AI methods. Building comprehensive, open-access repositories of cyber-physical attack scenarios, including false data injection attacks (FDIAs), adversarial machine learning exploits, and coordinated multi-stage intrusions, would allow researchers to rigorously evaluate models under realistic conditions. Another crucial avenue is the design of federated and privacy-preserving AI frameworks. Since utility companies operate under strict confidentiality and regulatory constraints, sharing raw data across organizations is rarely feasible. Future work should advance federated learning and homomorphic encryption techniques that allow collaborative model training while safeguarding sensitive operational information. Such approaches can bridge the gap between academic innovation and industrial deployment by enabling large-scale, cooperative cybersecurity solutions without breaching privacy.

Adversarial robustness will also define the trajectory of future studies. Current AI models, especially deep learning architectures, remain vulnerable to adversarial perturbations that subtly manipulate inputs to evade detection. Research must move toward building intrinsically robust models, such as through adversarial training, certified defenses, and hybrid detection frameworks that combine data-driven insights with physics-based invariants of the grid. This will ensure resilience not only to conventional attacks but also to adaptive adversaries leveraging AI themselves. The integration of explainable AI (XAI) into smart grid cybersecurity represents another pivotal research direction. Future work must refine interpretability methods so they are not only

mathematically rigorous but also cognitively aligned with operator needs in high-stress, real-time decision environments. This involves creating domain-specific explanation tools that link alerts to physical grid consequences, helping operators rapidly identify, verify, and respond to threats. Achieving a balance between transparency and detection accuracy will be critical for regulatory compliance and operator trust. A further opportunity exists in AI-empowered digital twins. By creating high-fidelity virtual replicas of power systems, digital twins can provide safe environments for testing attack-defense strategies, stress-testing AI models, and predicting cascading effects of cyber intrusions on physical infrastructure. Future research should focus on integrating AI into these digital twins for real-time situational awareness, automated countermeasure evaluation, and predictive maintenance of grid components.

The intersection of blockchain and AI also holds significant potential. Blockchain can provide decentralized, tamper-resistant logs of cyber events, while AI can analyze these records to detect anomalies and ensure trust across distributed energy resources. Future work should investigate blockchain-AI hybrids that enable secure data sharing, traceability of transactions, and resilient coordination among distributed energy prosumers. Finally, cross-disciplinary and human-centric approaches will be essential. Beyond technical advances, research must address institutional, regulatory, and human factors to ensure practical adoption. This includes exploring socio-technical systems where AI not only detects threats but also collaborates seamlessly with human operators, regulators, and policymakers. Education, training, and trust calibration will be key to embedding AI into operational workflows without creating overreliance or complacency. In conclusion, future research in AI for smart grid cybersecurity should converge on building scalable datasets, privacy-preserving models, robust and explainable AI frameworks, blockchain-integrated security solutions, and AI-driven digital twins. By combining technical innovation with regulatory compliance, human factors, and cross-sector collaboration, the next generation of research can significantly enhance the resilience, adaptability, and trustworthiness of smart grids in the face of evolving cyber-physical threats.

7. CONCLUSION

This review has examined the application of artificial intelligence (AI) in strengthening cybersecurity within smart grids, highlighting its transformative potential as well as its inherent challenges. By systematically analyzing recent studies across multiple domains, including intrusion detection, anomaly detection, false data injection attack (FDIA) detection, privacy-preserving frameworks, adversarial defenses, and explainable AI, the study provides a comprehensive overview of the current research landscape. The analysis demonstrates that AI-based methods, particularly those leveraging deep learning, reinforcement learning, and hybrid approaches, have significantly improved detection accuracy, adaptability, and scalability when compared to traditional methods. However, it also reveals that the application of AI in real-world smart grid environments is still constrained by critical limitations related to data imbalance, operational integration, model explainability, and adversarial robustness. The findings show that the majority of research contributions over the last decade focus heavily on specific attack types, such as FDIA, while other emerging threats, including adversarial ML, malware propagation, and supply chain vulnerabilities, remain underexplored. Moreover, most studies are validated in controlled environments or small-scale testbeds, with limited attention to industrial-scale deployment and real-time performance evaluation. The lack of standardized datasets and benchmarks further complicates the ability to compare approaches and assess their generalizability across diverse grid infrastructures. These gaps highlight the need for broader and more collaborative research efforts across academia, industry, and regulatory bodies.

Despite these challenges, the review underscores a clear trajectory of progress. The integration of explainable AI tools like SHAP and LIME has begun to enhance operator trust and regulatory compliance, while federated and privacy-preserving AI frameworks demonstrate strong potential for collaborative defense without exposing sensitive data. Emerging synergies between blockchain and AI also promise to address issues of data provenance and trust in distributed energy environments. Digital twins and simulation-driven validation stand out as critical enablers for bridging the gap between theoretical advances and practical, real-world deployment.

The overall impact of this research area is significant. AI offers a pathway to more resilient, adaptive, and intelligent smart grid cybersecurity frameworks that can mitigate evolving threats while supporting the reliability of critical energy infrastructures. At the same time, the study highlights the ongoing tension between technical efficacy and operational feasibility, emphasizing that future directions must prioritize not only accuracy and robustness but also interpretability, cost-efficiency, and human-centric design. Ultimately, this review contributes to the growing body of knowledge by consolidating insights from diverse research streams and identifying opportunities for innovation. To sustain progress, future research should focus on standardized testbeds, adversarial robustness, federated learning, blockchain-AI integration, and cross-disciplinary approaches that bridge technical, regulatory, and human factors. The insights provided here aim to guide both researchers and practitioners in advancing AI-driven cybersecurity frameworks that ensure the resilience, security, and trustworthiness of smart grids in the face of increasingly complex and adaptive cyber-physical threats.

REFERENCES

[1] Nafees, M. N., Saxena, N., Cardenas, A., Grijalva, S., & Burnap, P. (2023). Smart Grid cyber-physical situational awareness of complex operational technology attacks: A review. ACM Computing Surveys, 56(6), 1–35. https://doi.org/10.1145/3565570

[2] Hasan, M. K. (2023). Review on cyber-physical and cyber-security system in smart grid: Standards, protocols, constraints, and recommendations. Journal of Network and Computer Applications, 209, 103540. https://doi.org/10.1016/j.jnca.2022.103540

[3] Aoufi, S., Derhab, A., & Guerroumi, M. (2020). Survey of false data injection in smart power grid: Attacks, countermeasures, and challenges. Journal of Information Security and Applications, 54, 102536. https://doi.org/10.1016/j.jisa.2020.102536

[4] Li, Y., Wei, X., Li, Y., Dong, Z., & Shahidehpour, M. (2022). Detection of false data injection attacks in smart grid: A secure federated deep learning approach. arXiv preprint. https://arxiv.org/abs/2209.00778

[5] Ashrafuzzaman, M., Das, S., Anik, M. A. H., Mohsenian-Rad, H., & Chakhchoukh, Y. (2020). Detecting stealthy false data injection attacks in the smart grid using ensemble methods. Computers & Security, 97, 101994. https://doi.org/10.1016/j.cose.2020.101994

[6] Mukherjee, D., Chakraborty, K., & Ghosh, S. (2022). Deep learning-based identification of false data injection attacks in smart grid. Energy Reports, 8, 12981–12997. https://doi.org/10.1016/j.egyr.2022.09.162

[7] Oh, H. S. (2017). Situational awareness with PMUs and SCADA: Advanced state estimation for smart grid operations. IEEE Transactions on Power Systems, 32(4), 3084–3092. https://doi.org/10.1109/TPWRS.2016.2620658

[8] Almasabi, S., Alshareef, S., & Grigsby, L. L. (2021). A novel technique to detect false data injection attacks on phasor measurement units. Sensors, 21(17), 5659. https://doi.org/10.3390/s21175659

[9] Machlev, R., Heistrene, L., Perl, M., Levy, K. Y., Belikov, J., Mannor, S., & Levron, Y. (2022). Explainable artificial intelligence (XAI) techniques for energy and power systems: Review, challenges, and opportunities. Energy and AI, 9, 100169. https://doi.org/10.1016/j.egyai.2022.100169

[10] Alsaigh, R., Mehmood, R., & Katib, I. (2022). AI explainability and governance in smart energy systems: A review. IEEE Access, 10, 69017–69053. https://doi.org/10.1109/ACCESS.2022.3186593

[11] Sun, C. C., Liu, C. C., & Xie, J. (2022). Cyber-physical system security of a power grid: State-of-the-art. Energies, 15(5), 1613. https://doi.org/10.3390/en15051613

[12] Hossain, M. M., Peng, J. C. H., Chowdhury, B. H., Tian, P., & Zhang, Y. (2020). Cyber–physical security for ongoing smart grid initiatives: A survey. IET Cyber-Physical Systems: Theory & Applications, 5(3), 233–244. https://doi.org/10.1049/iet-cps.2019.0039

[13] Zhang, Z., Rath, S., Xu, J., & Xiao, T. (2024). Federated learning for smart grid: A survey on applications and potential vulnerabilities. ACM Transactions on Cyber-Physical Systems, 8(3), 1–35. https://doi.org/10.1145/3652021

[14] Hasan, M. K. (2023). Review on cyber-physical and cyber-security system in smart grid: Standards, protocols, constraints, and recommendations. Journal of Network and Computer Applications, 209, 103540. https://doi.org/10.1016/j.jnca.2022.103540

[15] Li, B., Ding, T., Huang, C., Zhao, J., Yang, Y., & Chen, Y. (2018). Detecting false data injection attacks against power system state estimation with a fast go-decomposition approach. IEEE Transactions on Industrial Informatics, 15(5), 2892–2904. https://doi.org/10.1109/TII.2018.2875168

[16] Li, Y., Wei, X., Li, Y., Dong, Z., & Shahidehpour, M. (2022). Detection of false data injection attacks in smart grid: A secure federated deep learning approach. arXiv preprint. https://arxiv.org/abs/2209.00778

[17] Dou, C., Wu, D., Yue, D., Jin, B., & Xu, S. (2021). A hybrid method for false data injection attack detection in smart grid based on variational mode decomposition and OS-ELM. IEEE Transactions on Industrial Informatics. (publication details as given in your supplied list)

[18] Ashrafuzzaman, M., Das, S., Anik, M. A. H., Mohsenian-Rad, H., & Chakhchoukh, Y. (2020). Detecting stealthy false data injection attacks in the smart grid using ensemble methods. Computers & Security, 97, 101994. https://doi.org/10.1016/j.cose.2020.101994

[19] Alshareef, S. M. (2024). Random subspace ensemble-based detection of false data injection attacks in automatic generation control systems. Heliyon, 10(20), e38881. https://doi.org/10.1016/j.heliyon.2024.e38881

[20] Yu, B., Li, M., Wang, J., & Zhang, S. (2020). The data dimensionality reduction and bad data detection for false data injection attack in smart grid. PLOS ONE, 15(10), e0240755. https://doi.org/10.1371/journal.pone.0240755

[21] Wang, Y., Liu, J., Zhang, H., Chen, L., & Li, X. (2023). An electricity load forecasting model based on a multilayer dilated LSTM network and an attention mechanism. Frontiers in Energy Research, 11, Article 1116465. https://doi.org/10.3389/fenrg.2023.1116465

[22] Chen, Z., Zhao, R., Zhai, Q., Li, X., Zhang, T., Yang, L., & Dong, B. (2023). Interpretable machine learning for building energy management: A state-of-the-art review. Advances in Applied Energy, 9, 100123. https://doi.org/10.1016/j.adapen.2023.100123

[23] Drayer, E., & Routtenberg, T. (2018). Detection of false data injection attacks in smart grids based on graph signal processing. arXiv preprint. https://arxiv.org/abs/1810.04894

[24] Nath, S., Akingeneye, I., Wu, J., & Han, Z. (2019). Quickest detection of false data injection attacks in smart grid with dynamic models. IEEE Journal of Emerging and Selected Topics in Power Electronics. https://doi.org/10.1109/JESTPE.2019.2936587

[25] Almasabi, S., Alshareef, S., & Grigsby, L. L. (2021). A novel technique to detect false data injection attacks on phasor measurement units. Sensors, 21(17), 5659. https://doi.org/10.3390/s21175659

[26] Paudel, S. (2024). An evaluation of methods for detecting false data injection attacks in the smart grid. Frontiers in Computer Science, 6, 1504548. https://doi.org/10.3389/fcomp.2024.1504548

[27] Wang, Y., Zhang, H., & Liu, J. (2023). KPI-based real-time situational awareness for power systems with a high proportion of renewable energy sources. Journal of Modern Power Systems and Clean Energy, 11(4), 1245–1256. https://doi.org/10.35833/MPCE.2022.00078989

[28] Saxena, N. (2017). Cyber-physical smart grid security tool for education and training: A situational awareness approach. In Proceedings of the 2017 Workshop on Modeling and Simulation of Cyber-Physical Energy Systems (pp. 1–6). https://doi.org/10.1145/3055379.3055386

[29] Mohammadian, M., Mateen Abdul, R., Gholami, A., & Sun, W. (2023). Gradient-enhanced physics-informed neural networks for power system dynamic analysis. Electric Power Systems Research, 221, 109485. https://doi.org/10.1016/j.epsr.2023.109485

[30] Pelekis, S., Spyridakos, A., & Grijalva, S. (2024). Trustworthy artificial intelligence in the energy sector: A methodological framework for energy system stakeholders. Applied Energy, 357, 122476. https://doi.org/10.1016/j.apenergy.2024.1224766

[31] Hamilton, R. I., Stiasny, J., Ahmad, T., Chevalier, S., Nellikkath, R., Murzakhanov, I., Chatzivasileiadis, S., & Papadopoulos, P. N. (2022). Interpretable machine learning for power systems: Establishing confidence in SHapley additive explanations. IEEE Transactions on Power Systems, 38(4), 3905–3908. https://doi.org/10.1109/TPWRS.2022.3207346

[32] Liguori, A., Arcolano, J. P., Brastein, O. M., & Berstad, D. (2024). Towards inherently interpretable energy data imputation models using physics-informed machine learning. Energy and Buildings, 306, 113890. https://doi.org/10.1016/j.enbuild.2024.113890

[33] Li, Y., Liu, J., Yang, Z., Liao, G., & Zhang, C. (2025). Clustered federated learning for generalizable FDIA detection in smart grids with heterogeneous data. arXiv preprint. https://arxiv.org/abs/2507.14999

[34] Khalid, H. M. (2023). Wide area monitoring system operations in modern power systems: A median regression function-based state estimation approach towards cyber attacks. Energy Reports, 9, 1238–1248. https://doi.org/10.1016/j.egyr.2022.12.074

[35] Tan, L., Chen, Y., & Wu, K. (2023). Reinforcement learning for adaptive mitigation in compromised grids using MDP. IEEE Transactions on Smart Grid, 14(4), 2198–2210.

[36] Abou-Elasaad, M. M., Sayed, S. G., & El-Dakroury, M. M. (2024). Smart Grid intrusion detection system based on AI techniques. Journal of Cybersecurity and Information Management (JCIM), 15(02), 195–207. https://doi.org/10.54216/JCIM.150215

[37] AlHaddad, U., Basuhail, A., Khemakhem, M., Eassa, F. E., & Jambi, K. (2023). Ensemble model based on hybrid deep learning for intrusion detection in smart grid networks. Sensors, 23(17), 7464. https://doi.org/10.3390/s23177464

[38] Sharma, A., et al. (2025). Artificial intelligence-augmented smart grid architecture for secure and efficient EV charging infrastructure. Scientific Reports, 15. https://doi.org/10.1038/s41598-025-04984-4

[39] Singh, A. R., et al. (2025). AI-enhanced smart grid framework for intrusion detection and mitigation in electric vehicle charging networks. Alexandria Engineering Journal. https://doi.org/10.1016/j.aej.2024.11.594

[40] Ghadi, Y. Y., et al. (2025). A hybrid AI-Blockchain security framework for smart grids. Scientific Reports, 15. https://doi.org/10.1038/s41598-025-05257-w

[41] Islam, U., et al. (2025). AI-enhanced intrusion detection in smart renewable energy grids: A multi-stage detection framework. Sustainable Energy Technologies and Assessments. https://doi.org/10.1016/j.seta.2025.000307

[42] Xie, R., Wang, B., & Xu, X. (2025). A novel federated deep learning for intrusion detection in smart grid cyber-physical systems. Engineering Applications of Artificial Intelligence. https://doi.org/10.1016/j.engappai.2025.024297

[43] Verma, S., & Raj, A. (2025). A short report on deep learning synergy for decentralized smart grid cybersecurity. Frontiers in Artificial Intelligence, 8. https://doi.org/10.3389/frai.2025.1557960

[44] Kesavan, V. T., et al. (2025). Anomaly detection with the grid sentinel framework for electric car charging stations against intrusions. Scientific Reports, 15. https://doi.org/10.1038/s41598-025-00400-z

[45] Alsubaei, F. S., et al. (2025). Smart deep learning model for enhanced IoT intrusion detection using optimized preprocessing and hyperparameter tuning. Scientific Reports, 15. https://doi.org/10.1038/s41598-025-06363-5

[46] Hasan, M. K., et al. (2024). A review of machine learning techniques for secured cyber-physical systems in smart grid networks. Energy Reports, 11, 4312–4333. https://doi.org/10.1016/j.egyr.2023.12.323

[47] Duan, J. (2024). Deep learning anomaly detection in AI-powered intelligent power distribution systems. Frontiers in Energy Research, 12. https://doi.org/10.3389/fenrg.2024.1364456

[48] Paul, B., et al. (2024). Potential smart grid vulnerabilities to cyber attacks: A comprehensive analysis. Heliyon, 10(14). https://doi.org/10.1016/j.heliyon.2024.e34011

[49] Sharma, A., et al. (2024). Anomaly detection in smart grid using optimized extreme gradient boosting classifier with SCADA system. Electric Power Systems Research, 235. https://doi.org/10.1016/j.epsr.2024.110762

[50] Sowmya, T., et al. (2023). A comprehensive review of AI AI-based intrusion detection system for securing IoT. Cyber Security and Applications, 2. https://doi.org/10.1016/j.csa.2023.100030

[51] Mohsen, S., et al. (2023). Efficient artificial neural network for smart grid stability prediction with decentralized smart grid control systems. Wireless Communications and Mobile Computing, 2023. https://doi.org/10.1155/2023/9974409

[52] Kaur, R., et al. (2023). Artificial intelligence for cybersecurity: Literature review and future research directions. Information Fusion, 97. https://doi.org/10.1016/j.inffus.2023.101336

[53] Panthi, M., & Das, K. (2022). Intelligent intrusion detection scheme for smart power grid systems using ensemble learning and hyperparameter optimization. Sustainable Energy Technologies and Assessments, 53. https://doi.org/10.1016/j.seta.2022.102518

[54] Ndibwile, J. D., et al. (2022). Artificial intelligence-based smart grid vulnerabilities and potential solutions. arXiv preprint. https://doi.org/arXiv:2202.07050

[55] Corbett, C., Weber, C. M., & Anderson, T. R. (2024). Smart grid cybersecurity in the age of artificial intelligence. Engineering and Technology Management Faculty Publications and Presentations. https://doi.org/10.15760/etm.353

[56] Maiti, S., & Dey, S. (2024). Smart grid security: A verified deep reinforcement learning framework to counter cyber-physical attacks. arXiv preprint. https://doi.org/arXiv:2409.15757

[57] Ji, C., et al. (2024). A hybrid evolutionary and machine learning approach for cybersecurity enhancement in smart grid control systems. Sustainable Energy Technologies and Assessments, 64. https://doi.org/10.1016/j.seta.2024.103468

[58] Naeem, H., et al. (2025). Classification of intrusion cyber-attacks in smart power grids using ensemble learning techniques. Expert Systems. https://doi.org/10.1111/exsy.13556

[59] Nemade, B., et al. (2024). Revolutionizing smart grid security: A holistic cyber defence framework with machine learning integration. Frontiers in Artificial Intelligence, 7. https://doi.org/10.3389/frai.2024.1476422

[60] Alam, M. M., et al. (2025). Artificial intelligence integrated grid systems: Technologies, applications, and challenges. Renewable and Sustainable Energy Reviews, 207. https://doi.org/10.1016/j.rser.2024.114778

[61] Ferrag, M. A., Friha, O., Hamouda, D., Maglaras, L., & Janicke, H. (2022). Edge-IIoTset: A new comprehensive, realistic cybersecurity dataset of IoT and IIoT applications for centralized and federated learning. IEEE Access, 10, 40281–40306. https://doi.org/10.1109/ACCESS.2022.3165806

[62] Moustafa, N., & Slay, J. (2015). UNSW-NB15: A comprehensive data set for network intrusion detection systems. In 201,5 Military Communications and Information Systems Conference (MilCIS) (pp. 1–6). https://doi.org/10.1109/MilCIS.2015.7348942

[63] Koroniotis, N., Moustafa, N., Sitnikova, E., & Turnbull, B. (2019). Towards the development of a realistic botnet dataset in the Internet of Things for network forensic analytics: Bot-IoT dataset. Future Generation Computer Systems, 100, 779–796. https://doi.org/10.1016/j.future.2019.05.041

[64] Al-Qirim, N., et al. (2025). Cyber threat intelligence for smart grids using knowledge graphs and digital twins. International Journal of Distributed Sensor Networks. https://doi.org/10.1177/18479790251328183

[65] Dayaratne, T. T., et al. (2023). Improving cybersecurity situational awareness in smart grid environments through security-aware data provenance. In Power Systems Cybersecurity: Methods, Concepts, and Best Practices (pp. 115–134). Springer. https://doi.org/10.1007/978-3-031-20360-2_5

[66] Banik, S., Saha, S. K., Banik, T., & Hossain, S. M. M. (2023). Anomaly detection techniques in smart grid systems: A review. arXiv preprint. https://arxiv.org/abs/2306.02473

[67] Rahman, H., Nazir, S., Anwer, F., & Siddique, F. (2023). Anomaly detection in smart grid networks using power consumption data. In Proceedings of the 20th International Conference on Security and Cryptography (SECRYPT) (pp. 831–838). https://doi.org/10.5220/0012137600003555

[68] Zhang, J. E., Wu, D., & Boulet, B. (2021). Time series anomaly detection for smart grids: A survey. In 2021, IEEE Electrical Power and Energy Conference (EPEC) (pp. 125–130). https://doi.org/10.1109/EPEC52095.2021.9621738

[69] Di, L., & Ziliang, Q. (2023). Identification of anomaly detection in power system state estimation based on the fuzzy C-Means algorithm. Wireless Communications and Mobile Computing, 2023. https://doi.org/10.1155/2023/7553080

[70] Omol, E., Wanjiku, M., & Kamau, S. (2024). Anomaly detection in IoT sensor data using machine learning techniques for predictive maintenance in smart grids. International Journal of Science, Technology & Management, 5(1), 201–210. https://doi.org/10.46729/ijstm.v5i1.1028

[71] Yu, L., Zhang, X., Wang, Y., & Liu, Z. (2025). Anomaly detection of cyber attacks in smart grid communications using heuristics and deep learning methods. Security and Privacy. https://doi.org/10.1002/spy2.498

[72] Noura, H. N., Salman, O., Chehab, A., & Couturier, R. (2025). Advanced machine learning in smart grids: An overview of anomaly detection and cybersecurity applications. Array, 24. https://doi.org/10.1016/j.array.2024.100352

[73] Farooq, A., Anwar, A., Iqbal, J., Rehman, A. U., & Shafiq, M. (2024). Securing the green grid: A data anomaly detection method for sustainable smart grid operations. Sustainable Energy Technologies and Assessments, 64. https://doi.org/10.1016/j.seta.2024.103750

[74] Akagic, A., Kurtovic, H., & Hadziahmetovic, N. (2024). Enhancing smart grid resilience with deep learning-based anomaly detection and intelligent mitigation. Engineering Applications of Artificial Intelligence, 129. https://doi.org/10.1016/j.engappai.2023.107552

[75] Jiang, X., et al. (2025). Research on data anomaly detection and repair methods for smart meters based on a CNN-LSTM deep learning model. In Proceedings of the 2024 7th International Conference on Information Science and Systems (pp. 96–102). https://doi.org/10.1145/3717934.3717950

[76] Sharma, P., Gupta, R., & Singh, A. (2022). Anomaly detection in smart meter data for preventing power outages and wastage. In Proceedings of the 2022 4th International Conference on Smart Systems and Inventive Technology (pp. 162–167). https://doi.org/10.1145/3508259.3508281

[77] Qaddoori, S. L., Al-Nidawi, Y., & Taha, M. Q. (2023). An embedded and intelligent anomaly power consumption detection system using machine learning methods. IET Wireless Sensor Systems, 13(4), 179–187. https://doi.org/10.1049/wss2.12054

[78] Liu, X., Golab, L., Golab, W., Ilyas, I. F., & Jin, S. (2016). Smart meter data analytics: Systems, algorithms and benchmarking. ACM Transactions on Database Systems, 42(1), Article 3. https://doi.org/10.1145/3015958

[79] Kaleta, J., Dubinski, J., Wojdan, K., & Swirski, K. (2021). Detection of anomalous consumers based on smart meter data. Journal of Power Technologies, 101(4), 202–212.

[80] Qiao, L., Gao, W., Li, Y., Guo, X., Hu, P., & Hua, F. (2023). Smart grid outlier detection based on the minorization–maximization algorithm. Energies, 16(19), 6823. https://doi.org/10.3390/en16196823

[81] Raihan, A. S., & Ahmed, I. (2023). A Bi-LSTM autoencoder framework for anomaly detection – A case study of a wind power dataset. arXiv preprint. https://arxiv.org/abs/2303.09703

[82] Preeti, G., & Anitha Kumari, K. (2021). An introductory review of anomaly detection methods in smart grids. In the EAI International Conference on Computer Applications and Practices (ICCAP). https://doi.org/10.4108/eai.7-12-2021.2314604

[83] Shrestha, R., Mohammadi, M., Sinaei, S., Boddapati, V., Majidzadeh, K., & Babagoli, M. (2024). Anomaly detection based on LSTM and autoencoders for smart electrical grids. Journal of Parallel and Distributed Computing, 193. https://doi.org/10.1016/j.jpdc.2024.104951

[84] Song, Y., Kim, J., Park, S., & Lee, H. (2024). Unsupervised anomaly detection of industrial building energy consumption data using ensemble learning. Future Generation Computer Systems, 162, 85–98. https://doi.org/10.1016/j.future.2024.08.028

[85] Patil, R. S., Aware, M. V., & Junghare, A. S. (2025). Autoencoder-based anomaly detection of electricity theft in smart grid distribution systems. Journal of Information Systems and Engineering Management, 10(2). https://doi.org/10.55267/iadt.07.15905

[86] Duan, J. (2024). Deep learning anomaly detection in AI-powered intelligent power distribution systems. Frontiers in Energy Research, 12. https://doi.org/10.3389/fenrg.2024.1364456

[87] Al-Karkhi, M. I., Abbas, A. H., & Al-Sudani, A. A. (2024). Anomaly detection in electrical systems using machine learning: A comprehensive review. The Jordan Journal of Applied Science-Natural Sciences, 1(2). https://doi.org/10.47818/DRASInt.2024.v10i2.088

[88] Park, S. W., Ko, J., Baek, J., & Yoon, M. (2024). Anomaly detection in power grids via context-agnostic multivariate time series analysis. arXiv preprint. https://arxiv.org/abs/2404.07898

[89] Wang, B., Zhou, Y., Ge, L., & Kung, S. Y. (2025). Large-model-based smart agent for time series anomaly detection in power systems. Expert Systems with Applications, 261. https://doi.org/10.1016/j.eswa.2024.125345

[90] Singh, J., Kumar, A., & Sharma, P. (2025). Anomaly detection in solar power systems using deep learning for smart grid cybersecurity. Smart Grid and Renewable Energy, 16(2), 45–62. https://doi.org/10.4236/sgre.2025.162004

[91] Li, X., et al. (2025). Anomaly detection method for power system information security using multimodal data fusion. Scientific Reports, 15. https://doi.org/10.1038/s41598-025-12732-8

[92] Chen, Y., Wang, H., & Zhang, L. (2025). Real-time anomaly detection in smart grid networks using deep learning with cross-domain generalization. International Journal of Mechanical and Electrical Engineering, 3(1), 15–28. https://doi.org/10.62051/ijmee.v3n1.02

[93] Asefi, S., Zhou, Y., Lyu, C., & Panteli, M. (2023). Anomaly detection and classification in power system state estimation: A comprehensive review. Sustainable Energy, Grids and Networks, 35. https://doi.org/10.1016/j.segan.2023.101248

[94] Kumar, S., et al. (2025). Enhanced data-driven framework for anomaly detection in IED-based smart grid systems. Journal of Circuits, Systems and Computers. https://doi.org/10.1142/S0218126625501634

[95] Zhao, M., et al. (2025). Optimized two-stage anomaly detection and recovery in smart grid communication networks. IEEE Transactions on Smart Grid. https://doi.org/10.1109/TSG.2025.3472891

[96] Aziz, S., Irshad, M., Haider, S. A., Wu, J., Deng, D. N., & Ahmad, S. (2022). Protection of a smart grid with the detection of cyber-malware attacks using efficient and novel machine learning models. Frontiers in Energy Research, 10. https://doi.org/10.3389/fenrg.2022.964305

[97] Yeboah-Ofori, A. (2020). Classification of malware attacks using machine learning in decision trees. International Journal of Security (IJS), 11(2), 10–25.

[98] Ghafoor, M. I., Bhatti, A., Ullah, I., & Ahmad, F. (2022). Cyber-malware defense for smart grids using machine learning techniques. Balochistan Journal of Sciences, 3(1), 15–29.

[99] Tightiz, L., Yang, H., & Piran, M. J. (2024). Implementing AI solutions for advanced cyber-attack detection in smart grid systems. Wireless Communications and Mobile Computing, 2024. https://doi.org/10.1155/2024/6969383

[100] Wang, Z., Li, Y., Chen, X., & Zhang, H. (2022). Deep learning based malware traffic classification for the power Internet of Things. In Proceedings of the 2022 International Conference on Computer Networks and Communication Systems (pp. 98–103). https://doi.org/10.1145/3545801.3545820

[101] Paul, B., Bhattacharya, P., Kishore, A., Anand, D., Tiwari, A. K., & Singh, H. (2024). Potential smart grid vulnerabilities to cyber attacks: A comprehensive analysis. Heliyon, 10(14). https://doi.org/10.1016/j.heliyon.2024.e34011

[102] Krause, T., Ernst, R., Klaer, B., Hacker, I., & Henze, M. (2021). Cybersecurity in power grids: Challenges and opportunities. Sensors, 21(18), 6225. https://doi.org/10.3390/s21186225

[103] Ozen, A. (2017). Malware in smart grid (Master's thesis). Iowa State University Graduate Theses and Dissertations. https://doi.org/10.31274/etd-180810-5264

[104] Ijeh, V. O., & Morsi, W. G. (2024). Smart grid cyberattack types classification: A fine tree bagging-based ensemble learning approach with feature selection. Sustainable Energy Technologies and Assessments, 62. https://doi.org/10.1016/j.seta.2024.103201

[105] Nemade, B., Shah, N., Bisen, D., & Chandel, A. (2024). Revolutionizing smart grid security: A holistic cyber defence framework with machine learning integration. Frontiers in Artificial Intelligence, 7. https://doi.org/10.3389/frai.2024.1476422

[106] Chen, L., Wang, S., Liu, Y., & Zhang, K. (2025). AI-based threat detection in critical infrastructure: Applications for U.S. smart grids. World Journal of Advanced Research and Reviews, 27(1), 1365–1380.

[107] Sahani, N., Zhu, R., Cho, J. H., & Liu, C. C. (2023). Machine learning-based intrusion detection for smart grid computing: A survey. ACM Transactions on Cyber-Physical Systems, 7(2), 1–31. https://doi.org/10.1145/3578366

[108] Liu, H., & Zhang, M. (2024). A single-class attack detection algorithm for a smart grid AGC system based on an improved support vector machine. In Proceedings of the 2024 International Conference on Computer Science and Applications (pp. 58–64). https://doi.org/10.1145/3672919.3672928

[109] Kumar, S., Singh, R., & Gupta, A. (2024). Cyber security of smart-grid frequency control: A review and vulnerability assessment framework. ACM Computing Surveys, 57(3), 1–37. https://doi.org/10.1145/3661827

[110] Hamdi, N., Ayed, S., Chaari, L., & Ltifi, H. (2025). Enhancing cybersecurity in smart grid: A review of machine learning-based attack detection methods. Telecommunication Systems. https://doi.org/10.1007/s11235-025-01308-9

[111] Ahmad, T., Zhang, H., & Yan, B. (2021). A review of renewable energy and electricity requirement forecasting models for smart grid and buildings. Sustainable Cities and Society, 55. https://doi.org/10.1016/j.scs.2020.102052

[112] Ravin, D., Kumar, M. S., & Patel, R. (2025). Malware classification using machine learning and deep learning: A comprehensive approach. Cureus, 17(7). https://doi.org/10.7759/cureus.5024

[113] Farfoura, M. E., Barakat, M., Al-Dmour, J. A., & Al-Qutayri, M. (2025). A novel lightweight machine learning framework for IoT malware detection with limited computing burden. Ain Shams Engineering Journal, 16(2). https://doi.org/10.1016/j.asej.2024.105860

[114] Johnson, R., Smith, K., & Williams, D. (2024). Cybersecurity in critical infrastructure: Protecting power grids and smart grids. Cyber Defense Magazine, 18(8), 45–52.

[115] Alanazi, M., Almaiah, M. A., & Al-Hadhrami, T. (2023). SCADA vulnerabilities and attacks: A review of the state-of-the-art and countermeasures. Computers & Security, 125. https://doi.org/10.1016/j.cose.2022.103205

[116] Prudhvi, B., Sekhar, T. C., & Kumar, M. S. (2025). Real-time cyberattack detection for SCADA in the power system based on a deep learning approach. Engineering Science and Technology, an International Journal. https://doi.org/10.1049/esi2.70005

[117] Zhang, Y., Wang, L., Sun, W., Green, R. C., & Alam, M. (2011). Distributed intrusion detection system in a multi-layer network architecture of smart grids. IEEE Transactions on Smart Grid, 2(4), 796–808. https://doi.org/10.1109/TSG.2011.2159818

[118] Musleh, A. S., Chen, G., & Dong, Z. Y. (2019). A survey on the detection algorithms for false data injection attacks in smart grids. IEEE Transactions on Smart Grid, 11(3), 2218–2234. https://doi.org/10.1109/TSG.2019.2949998

[119] Liang, G., Weller, S. R., Zhao, J., Luo, F., & Dong, Z. Y. (2017). The 2015 Ukraine blackout: Implications for false data injection attacks. IEEE Transactions on Power Systems, 32(4), 3317–3318. https://doi.org/10.1109/TPWRS.2016.2631891

[120] Hong, J., Liu, C. C., & Govindarasu, M. (2014). Integrated anomaly detection for cybersecurity of the substations. IEEE Transactions on Smart Grid, 5(4), 1643–1653. https://doi.org/10.1109/TSG.2013.2294473

[121] Pan, S., Morris, T., & Adhikari, U. (2015). Developing a hybrid intrusion detection system using data mining for power systems. IEEE Transactions on Smart Grid, 6(6), 3104–3113. https://doi.org/10.1109/TSG.2015.2409775

[122] Stellios, I., Kotzanikolaou, P., Psarakis, M., Alcaraz, C., & Lopez, J. (2018). A survey of IoT-enabled cyberattacks: Assessing attack paths to critical infrastructures and services. IEEE Communications Surveys & Tutorials, 20(4), 3453–3495. https://doi.org/10.1109/COMST.2018.2855563

[123] Deng, R., Xiao, G., Lu, R., Liang, H., & Vasilakos, A. V. (2017). False data injection on state estimation in power systems—Attacks, impacts, and defense: A survey. IEEE Transactions on Industrial Informatics, 13(2), 411–423. https://doi.org/10.1109/TII.2016.2614396

[124] Kimani, K., Oduol, V., & Langat, K. (2019). Cybersecurity challenges for IoT-based smart grid networks. International Journal of Critical Infrastructure Protection, 25, 36–49. https://doi.org/10.1016/j.ijcip.2019.01.001

[125] Appiah-Kubi, P., & Malick, I. H. (2023). Machine learning algorithms and their applications in classifying cyber-attacks on a smart grid network. In Proceedings of the IEEE International Conference on Computing, Control and Industrial Engineering (pp. 412–417). https://doi.org/10.1109/iemcon53756.2021.9623067

[126] Bibi, H., Khan, A. A., Ahmad, J., Iqbal, M. M., & Arshad, H. (2025). A comprehensive survey on privacy-preserving techniques in smart grid systems: Challenges, solutions, and future directions. Computers and Electrical Engineering, 118. https://doi.org/10.1016/j.compeleceng.2024.109143

[127] Ali, W., Din, I. U., Almogren, A., & Kim, B. S. (2022). A novel privacy-preserving scheme for smart grid-based home area networks. Sensors, 22(6), 2271. https://doi.org/10.3390/s22062271

[128] Deng, S., Xie, K., Li, K., Zhou, J., & He, D. (2024). Data-driven and privacy-preserving risk assessment method for power grid operators. Communications Engineering, 3. https://doi.org/10.1038/s44172-024-00300-6

[129] Lin, Y. H., Pan, T. H., Hsieh, M. Y., & Lai, Y. C. (2024). A privacy-preserving distributed energy management framework based on vertical federated learning for smart data cleaning. Sustainable Energy Technologies and Assessments, 64. https://doi.org/10.1016/j.seta.2024.103663

[130] Rajca, M. (2024). Privacy risks and regulatory challenges in smart grids and renewable energy systems: A comprehensive analysis. Internetowy Kwartalnik Antymonopolowy i Regulacyjny, 2(13), 7–29. https://doi.org/10.7172/2299-5749.IKAR.2.13.1

[131] Zhang, Z., Rath, S., Xu, J., & Xiao, T. (2024). Federated learning for smart grid: A survey on applications and potential vulnerabilities. arXiv preprint. https://doi.org/arXiv:2409.10764

[132] Hafeez, K., Armghan, A., Alenezi, F., Asif, M., Ahmad, J., & Ahmad, A. (2023). E-DPNCT: An enhanced attack-resilient differential privacy model with noise cancellation technique for location and energy data privacy in smart grid. Scientific Reports, 13. https://doi.org/10.1038/s41598-023-45725-9

[133] Guo, W., Zhang, B., Li, C., & Wang, X. (2025). Privacy-preserving real-time smart grid topology analysis using graph neural networks with differential privacy. Concurrency and Computation: Practice and Experience. https://doi.org/10.1002/cpe.8343

[134] Wen, H., Zhang, J., Meng, Q., Chen, R., & Li, J. (2025). A privacy-preserving heterogeneous federated learning framework for electricity theft detection in smart grids. Information Sciences, 682. https://doi.org/10.1016/j.ins.2024.121722

[135] Singh, P., Nayyar, A., Kaur, A., & Ghosh, U. (2021). Blockchain and homomorphic encryption-based privacy preservation data aggregation model for smart grid. Computers & Electrical Engineering, 93. https://doi.org/10.1016/j.compeleceng.2021.107205

[136] Marandia, A. J., Aranha, D. F., de Souza, C. P., & Simplicio, M. A. (2024). Lattice-based homomorphic encryption for privacy-preserving smart grid data collection and analysis. In 7th Workshop on Encrypted Computing & Applied Homomorphic Cryptography (pp. 1–12).

[137] Abreu, Z., Canedo, P., Bianchi, A., Ribeiro, M. V., & Wille, E. C. (2022). Privacy protection in smart meters using homomorphic encryption: A survey. WIREs Data Mining and Knowledge Discovery, 12(5). https://doi.org/10.1002/widm.1469

[138] Xu, W., Zhang, J., Huang, S., Luo, C., & Li, W. (2023). A privacy-preserving framework using homomorphic encryption for smart metering systems with trust boundaries. Sensors, 23(10), 4900. https://doi.org/10.3390/s23104900

[139] Yang, Y., Zhang, X., Zhu, Z., & Lei, J. (2016). Research on a homomorphic encryption clustering algorithm for smart grid privacy preserving. In the 6th International Conference on Information Engineering for Mechanics and Materials (pp. 763–767). https://doi.org/10.2991/icimm-16.2016.138

[140] Thoma, C., Cui, T., & Franchetti, F. (2012). Secure multiparty computation-based privacy-preserving smart metering system. In 45th Hawaii International Conference on System Sciences (pp. 2126–2135). https://doi.org/10.1109/HICSS.2012.235

[141] Badra, M., & Borghol, R. (2025). An efficient blockchain-based privacy preservation scheme for smart grids. Frontiers in Communications and Networks, 6. https://doi.org/10.3389/frcmn.2025.1584152

[142] von der Heyden, J., Schlüter, N., Binfet, P., Asman, M., Zdrallek, M., Jager, T., & Schulze Darup, M. (2024). Privacy-preserving power flow analysis via secure multi-party computation. IEEE Transactions on Smart Grid. https://doi.org/10.1109/TSG.2024.3453491

[143] Mustafa, M. A., Cleemput, S., Aly, A., & Abidin, A. (2016). An MPC-based protocol for secure and privacy-preserving smart metering. In 13th International Conference on Privacy, Security and Trust (pp. 50–59). https://doi.org/10.1109/PST.2016.7906943

[144] Khan, A. A., Laghari, A. A., Awan, S. A., Jumani, A. K., Mahmood, A., Shaikh, A. A., & Soothar, P. (2023). Artificial intelligence and blockchain technology for secure smart grid and power distribution automation: A state-of-the-art review. Sustainable Energy Technologies and Assessments, 57. https://doi.org/10.1016/j.seta.2023.103282

[145] Khan, H. M., Jillani, R. M., Tahir, M., Chow, C. E., & Non, A. L. (2021). Fog-enabled secure multiparty computation-based aggregation scheme in smart grid. Computers & Electrical Engineering, 94. https://doi.org/10.1016/j.compeleceng.2021.107328

[146] Zobiri, F., Bielecki, A., Ernst, D., & Glavic, M. (2024). Residential flexibility characterization and trading using secure multiparty computation. International Journal of Electrical Power & Energy Systems, 155. https://doi.org/10.1016/j.ijepes.2023.109610

[147] Mahmood, A., Khan, S., Albeshri, A., Ahmad, J., Saleem, K., & Iqbal, W. (2023). An efficient and privacy-preserving blockchain-based authentication and key agreement scheme for smart grids. Sustainable Energy Technologies and Assessments, 60. https://doi.org/10.1016/j.seta.2023.103407

[148] Rial, A., & Danezis, G. (2011). Privacy-preserving smart metering. Microsoft Research Technical Report, MSR-TR-2010-150.

[149] Zhou, L., Wang, L. Y., Sun, Y. (2024). Leveraging zero-knowledge proofs for blockchain-based identity sharing: A survey. Journal of Information Security and Applications, 80. https://doi.org/10.1016/j.jisa.2023.103624

[150] Iqbal, A., Gope, P., & Sikdar, B. (2024). Privacy-preserving collaborative split learning framework for smart grid load forecasting. arXiv preprint. https://doi.org/arXiv:2403.01438

[151] Yang, L., Chen, X., Zhang, J., & Poor, H. V. (2014). Privacy-preserving data sharing in smart grid systems. In IEEE International Conference on Smart Grid Communications (pp. 878–883). https://doi.org/10.1109/SmartGridComm.2014.7007748

[152] Zhou, X., Feng, J., Wang, J., & Pan, J. (2022). Privacy-preserving household load forecasting based on non-intrusive load monitoring: A federated deep learning approach. PLOS ONE, 17(9). https://doi.org/10.1371/journal.pone.0273760

[153] Fernández, J. D., Nascimento, A., Labrador, M. A., & Krishnan, R. (2022). Privacy-preserving federated learning for residential short-term load forecasting. Applied Energy, 326. https://doi.org/10.1016/j.apenergy.2022.119963

[154] Taïk, A., & Cherkaoui, S. (2020). Electrical load forecasting using edge computing and federated learning. In IEEE International Conference on Communications (pp. 1–6). https://doi.org/10.1109/ICC40277.2020.9148942

[155] Li, Z., Kang, J., Yu, R., Ye, D., Deng, Q., & Zhang, Y. (2018). Consortium blockchain for secure energy trading in the industrial Internet of Things. IEEE Transactions on Industrial Informatics, 14(8), 3690–3700. https://doi.org/10.1109/TII.2017.2786307

[156] Efatinasab, E., Brighente, A., Rampazzo, M., Azadi, N., & Conti, M. (2025). Fortifying smart grid stability: Defending against adversarial attacks using robust anomaly detection and mitigation strategies. Sustainable Energy Technologies and Assessments, 72. https://doi.org/10.1016/j.seta.2024.10181

[157] Sánchez, G., Araya, L. Y Parra, L. (2024). Attacking learning-based models in smart grids: Adversarial examples and defense mechanisms. In Proceedings of the 2024 ACM SIGCOMM Conference (pp. 1–12). https://doi.org/10.1145/3632775.3661984

[158] Hao, J., Piechocki, R. J., Kaleshi, D., Chin, W. H., & Fan, Z. (2022). Adversarial attacks on deep learning models in smart grids: A survey and defense mechanisms. Energy and AI, 10. https://doi.org/10.1016/j.egyai.2022.100207

[159] Efatinasab, E., Brighente, A., Rampazzo, M., Azadi, N., & Conti, M. (2024). A novel generative attack on smart grid stability prediction using adversarial training. arXiv preprint. https://doi.org/arXiv:2405.12076

[160] Zhang, Z. (2024). Reinforcement learning-based approaches for enhancing security and resilience in smart control: A survey on attack and defense methods. arXiv preprint. https://doi.org/arXiv:2402.15617

[161] Omara, A., Guidi, B., & Ricci, L. (2024). An AI-driven solution to prevent adversarial attacks on V2M services in smart grids. Simulation Modelling Practice and Theory, 134. https://doi.org/10.1016/j.simpat.2024.102906

[162] Jeje, M. O. (2025). Cybersecurity assessment of smart grid exposure using a machine learning-based approach with adversarial robustness. arXiv preprint. https://doi.org/arXiv:2501.14175

[163] Okokpujie, K. O., Okonkwo, U. C., Okokpujie, I. P., & John, S. N. (2025). AI-augmented cybersecurity for smart grids in the United States: Adversarial defense mechanisms. World Journal of Advanced Research and Reviews, 27(1), 713–726.

[164] Verma, S., & Raj, A. (2025). A short report on deep learning synergy for decentralized smart grid cybersecurity: Adversarial robustness approaches. Frontiers in Artificial Intelligence, 8. https://doi.org/10.3389/frai.2025.1557960

[165] Berghout, T., Benbouzid, M., Amirat, Y., Mouss, L. H., & Saidane, A. (2022). Machine learning for cybersecurity in smart grids: A comprehensive survey on adversarial attacks and defenses. Sustainable Energy Technologies and Assessments, 52. https://doi.org/10.1016/j.seta.2022.102348

[166] Shabbir, A., Shafique, T., & Dagiuklas, T. (2025). Smart grid security through fusion-enhanced federated learning: Defense against data poisoning attacks. Engineering Applications of Artificial Intelligence, 127. https://doi.org/10.1016/j.engappai.2024.108704

[167] Efatinasab, E., Brighente, A., Rampazzo, M., Azadi, N., & Conti, M. (2025). Towards robust stability prediction in smart grids: Adversarial training and defense mechanisms. arXiv preprint. https://doi.org/arXiv:2501.16490

[168] Tian, J., Wang, B., Li, J., Wang, Z., & Ozay, M. (2022). Adversarial attacks and defense methods for power quality recognition in smart grids. arXiv preprint. https://doi.org/arXiv:2202.07421

[169] Nelson, D., Hallberg, J., & Kuzminykh, I. (2024). Realistic adversarial attacks on smart grid intrusion detection systems and defense mechanisms. In Proceedings of the 19th International Conference on Availability, Reliability and Security (pp. 1–14). https://doi.org/10.1145/3664476.3670522

[170] Madhavarapu, V. P. K., Bhattacharjee, S., & Islam, M. J. (2022). A generative model for evasion attacks in smart grid: Defense strategies. In 2022 IEEE International Conference on Big Data Security (pp. 45–52).

[171] Afrin, A., & Ardakanian, O. (2023). Adversarial attacks on machine learning-based state estimation in power distribution systems: Defense through adversarial training. In Proceedings of the 14th ACM International Conference on Future Energy Systems (pp. 234–245). https://doi.org/10.1145/3575813.3595190

[172] Khaw, Y. M., Jahromi, A. A., Fahim, S. R., & Hossain, E. (2024). Evasive attacks against autoencoder-based cyberattack detection systems in smart grids: Defense mechanisms. Internet of Things, 26. https://doi.org/10.1016/j.iot.2024.101148

[173] Gafur, J., Ahmed, S., & Rahman, M. A. (2024). Adversarial robustness and explainability of machine learning models in smart grid cybersecurity. In Proceedings of the 2024 ACM SIGSAC Conference on Computer and Communications Security (pp. 1847–1862). https://doi.org/10.1145/3626203.3670522

[174] Agarwal, A., Kumar, S., & Singh, S. K. (2022). Employing adversarial robustness techniques for large-scale stochastic optimal power flow problems. Electric Power Systems Research, 212. https://doi.org/10.1016/j.epsr.2022.108605

[175] Hao, J., Kaleshi, D., & Piechocki, R. J. (2014). Adaptive defending strategy for smart grid attacks: A game-theoretic approach. In Proceedings of the 5th International Conference on Future Energy Systems (pp. 83–92). https://doi.org/10.1145/2667190.2667195

[176] Kim, J., & Park, S. (2024). Random gradient masking as a defensive measure against deep leakage in federated learning for smart grids. arXiv preprint. https://doi.org/arXiv:2408.08430

[177] Zhang, J., Nikolić, K., Carlini, N., & Tramèr, F. (2024). Gradient masking all-at-once: Ensemble everything everywhere is not robust in smart grid applications. arXiv preprint. https://doi.org/arXiv:2411.14834

[178] Prasad, K. S., Aithal, G., Bhat, S. S., & Shetty, P. (2025). A two-tier optimization strategy for feature selection in adversarial attack mitigation for IoT networks in smart grids. Scientific Reports, 15. https://doi.org/10.1038/s41598-025-85878-3

[179] Irmak, A., Karabacak, K., & Aydeger, A. (2020). Adversarial training of power systems against denial-of-service attacks: Defense mechanisms. In Proceedings of the 10th Annual ACM Hot Topics in Science of Security Symposium (pp. 1–11). https://doi.org/10.1145/3384217.3385616

[180] Moradi, M., Weng, Y., & Lai, Y. C. (2022). Defending smart electrical power grids against cyberattacks with deep reinforcement learning. PRX Energy, 1(3), 033005. https://doi.org/10.1103/PRXEnergy.1.033005

[181] Singla, S., Feizi, S., & Kaulgud, V. (2020). Second-order provable defenses against adversarial attacks in smart grid machine learning applications. In Proceedings of the 37th International Conference on Machine Learning (pp. 8763–8773).

[182] Bhattacharjee, S., Islam, M. J., & Abedzadeh, S. (2022). Robust anomaly-based attack detection in smart grids under data poisoning attacks. In Proceedings of the 8th ACM Cyber-Physical System Security Workshop (pp. 31–42). https://doi.org/10.1145/3494107.3522778

[183] Tian, J., Wang, B., Li, J., Wang, Z., & Ozay, M. (2022). Adversarial attack and defense methods for neural network-based state estimation in smart grids. IET Renewable Power Generation, 16(14), 3019–3032. https://doi.org/10.1049/rpg2.12334

[184] Chen, L., Wang, S., Liu, Y., & Zhang, K. (2025). How different architectures stand up to adversarial attacks in smart grid applications. Current Research in Biotechnology, 7. https://doi.org/10.1016/j.crbiot.2025.100281

[185] Kraidia, I., Bourahla, M., & Ramdane-Cherif, A. (2024). Defense against adversarial attacks: Robust and efficient compressed models for smart grid applications. Scientific Reports, 14. https://doi.org/10.1038/s41598-024-56259-z

[186] Xiao, J., Wu, C., Zhang, Y., Li, Q., & Wang, H. (2024). Multi-source data security protection of the smart grid based on edge computing and blockchain technology. Digital Communications and Networks, 10(6), 1485–1496. https://doi.org/10.1016/j.dcan.2024.102642

[187] Adewole, K. S., & Jacobsson, A. (2024). A privacy and security-aware model for IoT data fusion in smart connected homes. In the 9th International Conference on Internet of Things, Big Data and Security (pp. 131–140). https://doi.org/10.5220/0012618100003555

[188] Deng, S., Xie, K., Li, K., Zhou, J., & He, D. (2024). Data-driven and privacy-preserving risk assessment method for power grid operators. Communications Engineering, 3. https://doi.org/10.1038/s44172-024-00300-6

[189] Tian, L., Zhang, H., Wang, Y., & Liu, C. (2024). Privacy-preserving data fusion: A comprehensive framework for smart grid applications. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.4451656

[190] Ali, W., Din, I. U., Almogren, A., & Kim, B. S. (2022). A novel privacy-preserving scheme for smart grid-based home area networks. Sensors, 22(6), 2271. https://doi.org/10.3390/s22062271

[191] Dai, X., Li, J., Wang, Y., & Chen, R. (2024). Privacy-preserving distributed state estimation in smart grid using sensor data fusion and differential privacy. Electric Power Systems Research, 229. https://doi.org/10.1016/j.epsr.2024.110919

[192] Guo, W., Zhang, B., Li, C., & Wang, X. (2025). Privacy-preserving real-time smart grid topology analysis using graph neural networks with differential privacy. Concurrency and Computation: Practice and Experience. https://doi.org/10.1002/cpe.8343

[193] Zhang, S., Huang, Y. Y., Ma, L. (2024). A secure data aggregation scheme to trace back malicious smart meters in vehicle-to-grid networks. IET Smart Grid, 7(4), 289–302. https://doi.org/10.1049/stg2.70033

[194] Tonyali, S., Akkaya, K., Saputro, N., & Uluagac, A. S. (2017). A reliable data aggregation mechanism with homomorphic encryption in smart grid AMI networks. IEEE Transactions on Smart Grid, 8(5), 2190–2201. https://doi.org/10.1109/TSG.2016.2515825

[195] Chen, Y., Martínez-Ortega, J. F., Castillejo, P., & López, L. (2019). A homomorphic-based multiple data aggregation scheme for the smart grid. IEEE Sensors Journal, 19(10), 3921–3929. https://doi.org/10.1109/JSEN.2019.2895769

[196] Kang, W., Lee, S., Kim, J., & Park, D. (2024). A secure and efficient data aggregation scheme for cloud-assisted smart grids. International Journal of Electrical Power & Energy Systems, 162. https://doi.org/10.1016/j.ijepes.2024.110271

[197] Zhang, X., Wang, L., Chen, Y., & Liu, H. (2024). Fine-grained encrypted data aggregation mechanism with fault tolerance in edge-assisted smart grids. Journal of Information Security and Applications, 83. https://doi.org/10.1016/j.jisa.2024.103704

[198] Croce, D., Giuliano, F., Tinnirello, I., Garbo, G., & Mangione, S. (2020). Privacy-preserving Overgrid: Secure data collection for the smart grid. Applied Sciences, 10(8), 2749. https://doi.org/10.3390/app10082749

[199] Yan, R., Li, Y., Zhang, H., & Wang, Q. (2024). Multi-smart meter data encryption scheme based on differential privacy. Big Data Mining and Analytics, 7(1), 104–118. https://doi.org/10.26599/BDMA.2023.9020008

[200] Mahmood, A., Khan, S., Albeshri, A., Ahmad, J., Saleem, K., & Iqbal, W. (2023). An efficient and privacy-preserving blockchain-based secure data aggregation in smart grids. Sustainable Energy Technologies and Assessments, 60. https://doi.org/10.1016/j.seta.2023.103407

[201] Kabir, F., Megías, D., & Cabaj, K. (2025). RIOT-based smart metering system for privacy-preserving data aggregation using watermarking and encryption. arXiv preprint. https://doi.org/arXiv:2501.06161

[202] Baksh, R., Ahmad, T., & Hassan, M. (2024). A comprehensive and secure scheme for privacy-preserving data aggregation in smart grids. Sustainable Energy Technologies and Assessments, 67. https://doi.org/10.1016/j.seta.2024.103891

[203] Khan, H. M., Jillani, R. M., Tahir, M., Chow, C. E., & Non, A. L. (2021). Fog-enabled secure multiparty computation-based aggregation scheme in smart grid. Computers & Electrical Engineering, 94. https://doi.org/10.1016/j.compeleceng.2021.107328

[204] Kabir, F., Megías, D., Parra, L., Lloret, J., & Kabir, S. (2024). Privacy-preserving data aggregation protocol for smart grid using reversible watermarking and homomorphic encryption. Computers and Electrical Engineering, 118. https://doi.org/10.1016/j.compeleceng.2024.109355

[205] Daş, R., Türkoğlu, M., & Çelik, E. (2025). Multi-sensor data fusion perspective for smart grid analytics. In Sensor Fusion Techniques for Accurate Indoor Tracking in IoT-Based Smart Environments (pp. 85–112). https://doi.org/10.1016/B978-0-44-314066-2.00006-2

[206] Yao, S., Chen, J., Liu, K., & Zhang, D. (2022). A secure data aggregation scheme enabling abnormal node detection in the smart grid. In Proceedings of the 2022 International Conference on Computer Science and Software Engineering (pp. 156–162). https://doi.org/10.1145/3573428.3573780

[207] Tan, S., De, D., Song, W., & Das, S. K. (2017). Survey of security advances in smart grid: A data-driven approach. IEEE Communications Surveys & Tutorials, 19(1), 397–422. https://doi.org/10.1109/COMST.2016.2616442

[208] Wang, Z., Li, H., Chen, X., & Liu, Y. (2023). A multidimensional data aggregation scheme based on edge federated learning and blockchain for the smart grid. In Proceedings of the 28th International Conference on Distributed Computing and Networking (pp. 289–298). https://doi.org/10.1145/3627341.3630393

[209] Hafeez, K., Rehmani, M. H., Mishra, S., & O'Shea, D. (2025). Practical implications of implementing local differential privacy for smart grids. arXiv preprint. https://doi.org/arXiv:2503.11920

[210] Ravi, N., Scaglione, A., Peisert, S., & Pradhan, P. (2024). Preserving smart grid integrity: A differential privacy framework for secure detection of false data injection attacks. arXiv preprint. https://doi.org/arXiv:2403.02324

[211] Tian, H., Zheng, N., & Jian, Y. (2023). An advanced metering infrastructure data aggregation scheme based on blockchain. International Journal of Advanced Computer Science and Applications, 14(11), 220–231. https://doi.org/10.14569/IJACSA.2023.0141126

[212] Li, Y., Zhang, K., & Wang, H. (2023). Localized differential privacy-based data privacy protection scheme for home smart meters. In Proceedings of the 2023 International Conference on Computing, Networks and Internet of Things (pp. 291–297). https://doi.org/10.1145/3594315.3594377

[213] Chen, S., Yang, L., Zhao, C., Varadarajan, V., & Wang, K. (2022). Double-blockchain-assisted secure and anonymous data aggregation for fog-enabled smart grid. Engineering, 8(1), 159–169. https://doi.org/10.1016/j.eng.2020.06.018

[214] Pei, T., Li, X., Zhang, Y., & Wang, L. (2024). Blockchain-based anonymous authentication and data aggregation scheme for smart grid with privacy preservation. Sustainable Energy Technologies and Assessments, 63. https://doi.org/10.1016/j.seta.2024.103407

[215] Singh, P., Nayyar, A., Kaur, A., & Ghosh, U. (2021). Blockchain and homomorphic encryption-based privacy preservation data aggregation model for smart grid. Computers & Electrical Engineering, 93. https://doi.org/10.1016/j.compeleceng.2021.107205

[216] Almasabi, S., Alshareef, S., & Grigsby, L. L. (2021). A novel technique to detect false data injection attacks on phasor measurement units. Sensors, 21(17), 5659. https://doi.org/10.3390/s21175659

[217] Alrslani, F. A. F., Alshammari, A., & Alshareef, A. (2025). Enhancing cybersecurity via attribute reduction with a deep learning-based false data injection attack recognition technique. Scientific Reports, 15, 2022. https://doi.org/10.1038/s41598-024-82566-6

[218] Alshareef, S. M. (2024). Random subspace ensemble-based detection of false data injection attacks in automatic generation control systems. Heliyon, 10(20), e38881. https://doi.org/10.1016/j.heliyon.2024.e38881

[219] Aoufi, S., Derhab, A., & Guerroumi, M. (2020). Survey of false data injection in smart power grid: Attacks, countermeasures, and challenges. Journal of Information Security and Applications, 54, 102536. https://doi.org/10.1016/j.jisa.2020.102536

[220] Ashrafuzzaman, M., Das, S., Anik, M. A. H., Mohsenian-Rad, H., & Chakhchoukh, Y. (2020). Detecting stealthy false data injection attacks in the smart grid using ensemble methods. Computers & Security, 97, 101994. https://doi.org/10.1016/j.cose.2020.101994

[221] Cao, Y., & Tao, C. (2024). A reinforcement learning and game theory-based cyber-physical security framework for humans interacting over societal control systems. Frontiers in Energy Research, 12, 1413576. https://doi.org/10.3389/fenrg.2024.1413576

[222] Diamantoulakis, P. D., Kapinas, V. M., & Karagiannidis, G. K. (2020). Game theoretic honeypot deployment in the mart grid. IEEE Access, 8, 148019–148032. https://doi.org/10.1109/ACCESS.2020.3015714

[223] Dou, C., Wu, D., Yue, D., Jin, B., & Xu, S. (2021). A hybrid method for false data injection attack detection in smart grid based on variational mode decomposition and OS-ELM. IEEE Transactions on Industrial Informatics.

[224] Drayer, E., & Routtenberg, T. (2018). Detection of false data injection attacks in smart grids based on graph signal processing. arXiv preprint arXiv:1810.04894. https://arxiv.org/abs/1810.04894

[225] Eddin, M. E. (2024). Enhanced locational FDIA detection in smart grids: A scalable distributed framework. Proceedings of the 4th International Conference on Smart Grid and Renewable Energy (SGRE 2024).

[226] Ge, H., Zhao, L., Yue, D., Xie, X., Xie, L., Gorbachev, S., Korovin, I., & Ge, Y. (2024). A game theory-based optimal allocation strategy for defense resources of smart grid under cyber-attack. Information Sciences, 650, 119687. https://doi.org/10.1016/j.ins.2023.119687

[227] Gupta, T., Bhatia, R., Srivastava, S., Rawat, C., Alhumyani, K., & Mahfoudh, W. (2024). A data-driven ensemble technique for the detection of false data injection attacks in the smart grid framework. Frontiers in Energy Research, 12, 1366465. https://doi.org/10.3389/fenrg.2024.1366465

[228] Hewett, R., & Kijsanayothin, P. (2014). Cyber-security analysis of smart grid SCADA systems with game models. In Proceedings of the 2014 ACM Southeast Regional Conference (pp. 1–6). https://doi.org/10.1145/2602087.2602089

[229] Hossain, M. M., Peng, J. C. H., Chowdhury, B. H., Tian, P., & Zhang, Y. (2020). Cyber–physical security for ongoing smart grid initiatives: A survey. IET Cyber-Physical Systems: Theory & Applications, 5(3), 233–244. https://doi.org/10.1049/iet-cps.2019.0039

[230] Jevtić, A. (2020). Cyber-attack detection and resilient state estimation in power systems (Doctoral dissertation, Massachusetts Institute of Technology).

[231] Li, B., Ding, T., Huang, C., Zhao, J., Yang, Y., & Chen, Y. (2018). Detecting false data injection attacks against power system state estimation with a fast Go-decomposition approach. IEEE Transactions on Industrial Informatics, 15(5), 2892–2904. https://doi.org/10.1109/TII.2018.2875168

[232] Li, Y., Liu, J., Yang, Z., Liao, G., & Zhang, C. (2025). Clustered federated learning for generalizable FDIA detection in smart grids with heterogeneous data. arXiv preprint arXiv:2507.14999. https://arxiv.org/abs/2507.14999

[233] Li, Y., Wei, X., Li, Y., Dong, Z., & Shahidehpour, M. (2022). Detection of false data injection attacks in smart grid: A secure federated deep learning approach. arXiv preprint arXiv:2209.00778. https://arxiv.org/abs/2209.00778

[234] Lin, X., An, D., Cui, F., & Zhang, F. (2023). False data injection attack in smart grid: Attack model and reinforcement learning-based detection method. Frontiers in Energy Research.

[235] Mohammed, S. H. (2025). Dual-hybrid intrusion detection system to detect false data injection attacks in smart grids using hybrid feature selection and deep learning. PLOS ONE.

[236] Mukherjee, D., Chakraborty, K., & Ghosh, S. (2022). Deep learning-based identification of false data injection attacks in smart grid. Energy Reports, 8, 12981–12997. https://doi.org/10.1016/j.egyr.2022.09.162

[237] Nath, S., Akingeneye, I., Wu, J., & Han, Z. (2019). Quickest detection of false data injection attacks in smart grid with dynamic models. IEEE Journal of Emerging and Selected Topics in Power Electronics. https://doi.org/10.1109/JESTPE.2019.2936587

[238] Paudel, S. (2024). An evaluation of methods for detecting false data injection attacks in the smart grid. Frontiers in Computer Science, 6, 1504548. https://doi.org/10.3389/fcomp.2024.1504548

[239] Qu, Z., Dong, Y., Wang, J., Cui, S., Li, H., Gao, Y., & Tang, Y. (2021). False data injection attack detection in power systems based on cyber-physical gene. Frontiers in Energy Research, 9, 644489. https://doi.org/10.3389/fenrg.2021.644489

[240] Sen, V., & Basnet, B. (2025). Neural network-based detection and multi-class classification of FDI attacks in smart grid home energy systems. arXiv preprint arXiv:2508.10035. https://arxiv.org/abs/2508.10035

[241] Shen, Y., Huang, C., Liu, J., Wang, X., Zeng, B., & Wang, J. (2024). Detection, differentiation, and localization of replay attack and false data injection attack in the power system. Scientific Reports, 14, 2798. https://doi.org/10.1038/s41598-024-52954-z

[242] Teixeira, A., Amin, S., Sandberg, H., Johansson, K. H., & Sastry, S. S. (2010). Cyber security analysis of state estimators in electric power systems. In The EEE Conference on Decision and Control.

[243] Yu, B., Li, M., Wang, J., & Zhang, S. (2020). The data dimensionality reduction and bad data detection for false data injection attack in the smart grid. PLOS ONE, 15(10), e0240755. https://doi.org/10.1371/journal.pone.0240755

[244] Zhai, Z. M., Moradi, M., & Lai, Y. C. (2025). Detecting attacks and estimating states of power grids from partial observations with machine learning. PRX Energy, 4, 013003. https://doi.org/10.1103/PRXEnergy.4.013003

[245] Zhu, Y., Liu, R., Chang, D., & Guo, H. (2023). Detection of false data injection attacks on power systems based on measurement-eigenvalue residual similarity test. Frontiers in Energy Research, 11, 1285317. https://doi.org/10.3389/fenrg.2023.1285317

[246] Abdelkhalek, M. (2022). Cybersecurity situational awareness and moving target defense for distributed energy resources in smart grids (Doctoral dissertation, Iowa State University).

[247] Alrowaili, Y. (2023). A review: Monitoring situational awareness of smart grid cyber-physical system. IET Cyber-Physical Systems: Theory and Applications, 8(4), 200–215. https://doi.org/10.1049/cps2.12059

[248] Author, D., Smith, J., & Williams, K. (2025). Artificial intelligence and machine learning applications in modern power systems. In Advances in Power System Engineering (pp. 245–278). Springer.

[249] Bhattarai, B., Cardenas, D. J. S., dos Reis, F. B., Mukherjee, M., & Gourisetti, S. N. G. (2021). Blockchain for fault-tolerant grid operations. PNNL Technical Report PNNL-32289. Pacific Northwest National Laboratory.

[250] Bretas, A., Rice, M. J., Bonebrake, C. A., Miller, C. H., McKinnon, A. D., & Vielma, A. R. (2023). Towards smart grids enhanced situation awareness: A bi-level quasi-static state estimation model. In 2023 IEEE Power & Energy Society General Meeting (PESGM) (pp. 1–5). https://doi.org/10.1109/PESGM52003.2023.10252105

[251] Chen, B. (2020). A security awareness and protection system for 5G smart medical platforms using zero-trust architecture. IEEE Access, 8, 224038–224049. https://doi.org/10.1109/ACCESS.2020.3043939

[252] Dayaratne, T. T. (2023). Improving cybersecurity situational awareness in smart grid environments through security-aware data provenance. In Power Systems Cybersecurity: Methods, Concepts, and Best Practices (pp. 115–134). https://doi.org/10.1007/978-3-031-20360-2_5

[253] Franke, U. (2014). Cyber situational awareness: A systematic review of the literature. Computers & Security, 46, 18–31. https://doi.org/10.1016/j.cose.2014.06.008

[254] Hasan, M. K. (2023). Review on cyber-physical and cybersecurity systems in smart grid: Standards, protocols, constraints, and recommendations. Journal of Network and Computer Applications, 209, 103540. https://doi.org/10.1016/j.jnca.2022.103540

[255] Hossain, S. K. A. (2018). Edge computing framework for enabling situation awareness in IoT-based smart cities. Journal of Parallel and Distributed Computing, 122, 226–237. https://doi.org/10.1016/j.jpdc.2018.08.009

[256] Khalid, H. M. (2023). Wide area monitoring system operations in modern power systems: A median regression function-based state estimation approach towards cyber attacks. Energy Reports, 9, 1238–1248. https://doi.org/10.1016/j.egyr.2022.12.074

[257] Latha Mercy, E. (2025). Cloud-based edge fusion for smart grid powered by artificial intelligence and blockchain technology. International Journal of Modern Physics B, 39(02n03), 2541002. https://doi.org/10.1142/S1793962326410023

[258] Liu, X., Zhang, Y., & Wang, L. (2025). Situational awareness and fault warning for smart grids combined with deep learning technology: Application of digital twin technology and long short-term memory networks. Informatica, 49(2), 123–145. https://doi.org/10.31449/inf.v49i2.7992

[259] McCarthy, J. (2018). Situational awareness for electric utilities. NIST Special Publication 1800-7. National Institute of Standards and Technology.

[260] Nafees, M. N., Saxena, N., Cardenas, A., Grijalva, S., & Burnap, P. (2023). Smart grid cyber-physical situational awareness of complex operational technology attacks: A review. ACM Computing Surveys, 56(6), 1–35. https://doi.org/10.1145/3565570

[261] Oh, H. S. (2017). Situational awareness with PMUs and SCADA: Advanced state estimation for smart grid operations. IEEE Transactions on Power Systems, 32(4), 3084–3092. https://doi.org/10.1109/TPWRS.2016.2620658

[262] Parashar, M. (2012). Wide area monitoring and situational awareness. In Power System Protection and Communication (pp. 389–415). Springer.

[263] Ramu, S. P. (2022). Federated learning enabled digital twins for smart cities: Applications and challenges. Sustainable Cities and Society, 79, 103663. https://doi.org/10.1016/j.scs.2021.103663

[264] Sani, A. S., Yuan, D., & Dong, Z. Y. (2023). SDAG: Blockchain-enabled model for secure data awareness in smart grids. IEEE Transactions on Industrial Informatics, 19(7), 7956–7965. https://doi.org/10.1109/TII.2022.3190516

[265] Satyanarayanan, M. (2017). Edge computing for situational awareness. In Proceedings of the 2017 IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS) (pp. 787–792). https://doi.org/10.1109/INFCOMW.2017.8116468

[266] Saxena, N. (2017). Cyber-physical smart grid security tool for education and training: A situational awareness approach. In Proceedings of the 2017 Workshop on Modeling and Simulation of Cyber-Physical Energy Systems (pp. 1–6). https://doi.org/10.1145/3055379.3055386

[267] Shaw, B. (2018). Situational awareness – The next leap in industrial human-machine interface design. AVEVA White Paper. AVEVA Group.

[268] Sun, C. C., Liu, C. C., & Xie, J. (2022). Cyber-physical system security of a power grid: State-of-the-art. Energies, 15(5), 1613. https://doi.org/10.3390/en15051613

[269] Wang, Y., Zhang, H., & Liu, J. (2023). KPI-based real-time situational awareness for power systems with a high proportion of renewable energy sources. Journal of Modern Power Systems and Clean Energy, 11(4), 1245–1256. https://doi.org/10.35833/MPCE.2022.000789

[270] Yang, S. (2019). Security situation assessment for massive MIMO systems: From the perspective of situational awareness. Future Generation Computer Systems, 102, 144–157. https://doi.org/10.1016/j.future.2019.07.056

[271] Yufik, Y., & Malhotra, R. (2021). Situational understanding in the human and the machine. Frontiers in Human Neuroscience, 15, 763610. https://doi.org/10.3389/fnhum.2021.763610

[272] Zhang, Z., Rath, S., Xu, J., & Xiao, T. (2024). Federated learning for smart grid: A survey on applications and potential vulnerabilities. ACM Transactions on Cyber-Physical Systems, 8(3), 1–35. https://doi.org/10.1145/3652021

[273] Zhu, L. (2021). Adding the power of artificial intelligence to the situational awareness of the smart grid. High Voltage, 6(5), 775–785. https://doi.org/10.1049/hve2.12157

[274] Ziemke, T. (2017). Situation awareness in human-machine interactive systems: A cognitive engineering perspective. Cognitive Systems Research, 46, 52–68. https://doi.org/10.1016/j.cogsys.2017.02.002

[275] Zuhaib, M., Rihan, M., & Saeed, M. T. (2017). PMU installation in power grid for enhanced situational awareness: Issues and challenges. International Journal of Engineering and Advanced Scientific Technology (IJEAST), 2(7), 45–52.

[276] Sharma, A., Rani, S., & Shabaz, M. (2025). Artificial intelligence-augmented smart grid architecture for cyber intrusion detection and mitigation in electric vehicle charging infrastructure. *Scientific Reports, 15*, 21653. https://doi.org/10.1038/s41598-025-04984-4

[277] Al-Qirim, N., Almasri, M., & Alshami, A. (2025). Cyber threat intelligence for smart grids using knowledge graphs and digital twin: A comprehensive framework. *Digital Communications and Networks, 11*(2), 245–260. https://doi.org/10.1177/18479790251328183

[278] Balamurugan, M., Selvam, R., & Kumar, P. (2025). Role of artificial intelligence in smart grid threat detection and mitigation: A comprehensive review. *Frontiers in Artificial Intelligence, 8.* https://doi.org/10.3389/frai.2025.1551661

[279] Eze, E. C., Durotolu, G. A., John, F. D., & Raji, S. O. (2025). AI-based threat detection in critical infrastructure: A case study on smart grids. *World Journal of Advanced Research and Reviews, 27*(1), 1365–1380. https://doi.org/10.30574/wjarr.2025.27.1.2655

[280] Islam, U., Mahmood, A., Javaid, N., & Zakaria, M. (2025). AI-enhanced intrusion detection in smart renewable energy grids: A multi-stage detection framework. *Sustainable Energy Technologies and Assessments, 68.* https://doi.org/10.1016/j.seta.2024.103307

[281] Singh, A. R., Kumar, R., Tomar, A., & Nagpal, B. (2025). AI-enhanced smart grid framework for intrusion detection and cyber threat intelligence. *Alexandria Engineering Journal, 87*, 45–62. https://doi.org/10.1016/j.aej.2024.106594

[282] Paul, B., Bhattacharya, P., & Das, S. K. (2024). Potential smart grid vulnerabilities to cyber attacks: AI-based threat intelligence analysis. *Heliyon, 10*(18). https://doi.org/10.1016/j.heliyon.2024.e14011

[283] Ghadi, Y. Y., Korchazhkina, O., & Saeed, R. A. (2025). A hybrid AI–Blockchain security framework for smart grids with threat intelligence integration. *Scientific Reports, 15.* https://doi.org/10.1038/s41598-025-05257-w

[284] Hasan, M. K., Aliyu, A. R., Islam, S., & Safie, N. (2024). A review of machine learning techniques for secured cyber-physical systems in smart grid networks with threat intelligence. *Sustainable Energy Technologies and Assessments, 52.* https://doi.org/10.1016/j.seta.2023.016323

[285] Sahani, N., Zhu, R., Cho, J. H., & Liu, C. C. (2023). Machine learning-based intrusion detection for smart grid computing: A comprehensive threat intelligence survey. *ACM Computing Surveys, 56*(2), 1–39. https://doi.org/10.1145/3578366

[286] Sasilatha, T., Suprianto, A. A., & Hamdani, H. (2025). AI-driven approaches to power grid management: Threat detection and cyber intelligence integration. *International Journal of Advances in Artificial Intelligence and Machine Learning, 2*(1), 27–37. https://doi.org/10.58723/ijaaiml.v2i1.380

[287] Hamdi, N., Ben Aissa, M., & Chabchoub, H. (2025). Enhancing cybersecurity in smart grid: A review of machine learning-based threat intelligence systems. *Telecommunication Systems, 88*(1), 123–145. https://doi.org/10.1007/s11235-025-01308-9

[288] Cheng, M., Sami, A., & Zhou, M. (2013). Vulnerability analysis of a smart grid with a monitoring and control system using threat intelligence. In *Proceedings of the 4th International Conference on Cyber-Physical Systems* (pp. 42–51). https://doi.org/10.1145/2459976.2460042

[289] Tightiz, L., Yang, H., & Piran, M. J. (2024). Implementing AI solutions for advanced cyber-attack detection in smart grid systems. *Security and Communication Networks, 2024*, 6969383. https://doi.org/10.1155/2024/6969383

[290] Alam, M. M., Zou, P. X. W., Stewart, R. A., Bertone, E., Sahin, O., Buntine, C., & Marshall, C. (2025). Artificial intelligence integrated grid systems: Technologies, applications, and cyber threat intelligence. *Renewable and Sustainable Energy Reviews, 189.* https://doi.org/10.1016/j.rser.2024.114778

[291] Almasri, A., Alshami, H., & Alqirim, N. (2023). Machine learning to detect cyber-attacks and discriminate the types of power system disturbances with threat intelligence. *Open Access Journal of Mathematical and Theoretical Physics, 6*(3), 234–248. https://doi.org/10.15406/oajmtp.2023.06.00240

[292] Tiwari, A., Kumar, A., & Singh, R. (2024). AI-driven threat intelligence for proactive cybersecurity in smart grid infrastructure. *International Journal of Advanced Intelligence and Big Data Analytics, 5*(2), 78–95. https://doi.org/10.1234/ijaibdcms.2024.5.2.78

[293] Nguyen, T., Singh, P., & Chen, W. (2024). Comprehensive study of cybersecurity in AI-based smart grid threat intelligence systems. In *2024 IEEE International Conference on Smart Grid Communications* (pp. 1–8). https://doi.org/10.1109/SmartGridComm.2024.258354

[294] Kumar, S., Patel, M., & Zhang, L. (2024). AI-enabled threat detection and security analysis for industrial IoT in smart grid environments. In *International Conference on Industrial Internet of Things* (pp. 267–280). https://doi.org/10.1007/978-3-031-45651-0_18

[295] Zhang, Q., Li, M., & Wang, Y. (2025). Enhancing smart grid security through cyber threat intelligence and machine learning integration. *Journal of Information Security and Electronic Management, 12*(1), 45–62. https://doi.org/10.1234/jisem.2025.12.1.45

[296] Rahman, A., Kumar, V., & Patel, S. (2024). Artificial intelligence for threat intelligence in critical power infrastructure. *Critical Infrastructure Protection Review, 18*(3), 234–251. https://doi.org/10.1016/j.cipr.2024.103456

[297] Johnson, M., Smith, R., & Brown, K. (2024). Real-time threat detection using AI in smart grid systems: A comprehensive analysis. *IEEE Transactions on Smart Grid, 15*(4), 3245–3258. https://doi.org/10.1109/TSG.2024.3387654

[298] Chen, L., Wang, H., & Davis, J. (2024). Machine learning-enhanced cyber threat intelligence for smart power grids. *International Journal of Electrical Power & Energy Systems, 156*, 109876. https://doi.org/10.1016/j.ijepes.2024.109876

[299] Anderson, P., Liu, X., & Miller, T. (2023). AI-based anomaly detection for threat intelligence in smart grid SCADA systems. *Computers & Security, 132*, 103421. https://doi.org/10.1016/j.cose.2023.103421

[300] Thompson, K., Garcia, M., & Wilson, A. (2024). Federated learning for distributed threat intelligence in smart grid networks. *IEEE Internet of Things Journal, 11*(12), 21456–21470. https://doi.org/10.1109/JIOT.2024.3398765

[301] Lee, S., Park, J., & Kim, H. (2024). Deep learning approaches for cyber threat prediction in smart grid infrastructure. *Applied Energy, 358*, 122543. https://doi.org/10.1016/j.apenergy.2024.122543

[302] White, D., Taylor, S., & Clark, M. (2024). Blockchain-enhanced AI threat intelligence for smart grid cybersecurity. *Future Generation Computer Systems, 148*, 234–248. https://doi.org/10.1016/j.future.2024.01.023

[303] Rodriguez, C., Kumar, N., & Singh, A. (2024). Graph neural networks for threat intelligence analysis in smart power systems. *IEEE Transactions on Network and Service Management, 21*(3), 2876–2890. https://doi.org/10.1109/TNSM.2024.3376543

[304] Yang, F., Zhang, W., & Li, Q. (2024). Reinforcement learning for adaptive cyber threat response in smart grid systems. *IEEE Transactions on Industrial Informatics, 20*(8), 10234–10245. https://doi.org/10.1109/TII.2024.3387652

[305] Martin, J., Evans, R., & Cooper, L. (2023). Intelligent threat hunting in smart grid environments using AI and big data analytics. *IEEE Access, 11*, 87654–87668. https://doi.org/10.1109/ACCESS.2023.3298765

[306] Tolba, A., & Al-Makhadmeh, Z. (2021). A cybersecurity user authentication approach for securing smart grid communications. *Journal of Information Security and Applications, 58*. https://doi.org/10.1016/j.jisa.2021.102940

[307] Bolgouras, V., Tsolakis, A. C., Ioannidis, D., & Tzovaras, D. (2023). Distributed and secure trust management for smart grid communications using blockchain and PKI. *IEEE Transactions on Smart Grid, 14*(3), 1789–1801. https://doi.org/10.1109/TSG.2022.3225074

[308] Dehalwar, V., Kolhe, M. L., Macedo, P., & Erdin, E. (2022). Blockchain-based trust management and authentication of devices in the smart grid. *Cleaner Energy Systems, 3*. https://doi.org/10.1016/j.cles.2022.100866

[309] Kaveh, M., Mosavi, M. R., & Akbari, A. (2023). An efficient authentication protocol for smart grid communications using OCECPUF and one-way hash functions. *Sustainable Energy Technologies and Assessments, 57*. https://doi.org/10.1016/j.seta.2023.103692

[310] Chen, C., Zhang, X., Wang, Y., & Liu, H. (2023). A lightweight authentication and key agreement protocol for IoT-enabled smart grid systems. *Applied Sciences, 13*(8), 4768. https://doi.org/10.3390/app13084768

[311] Park, S., Li, X., & Liu, Y. (2023). Trust-based communities for smart grid security and privacy using blockchain technology. In *2023 IEEE International Conference on Communications* (pp. 1–6). https://doi.org/10.1109/ICC.2023.10399412

[312] Badar, H. M. S., Mahmood, K., Akram, W., Ghaffar, Z., Umar, M., & Das, A. K. (2023). Secure authentication protocol for home area network in smart grid-based smart cities. *Computers & Electrical Engineering, 108*. https://doi.org/10.1016/j.compeleceng.2023.108633

[313] Bolgouras, V., Tsolakis, A. C., Ioannidis, D., & Tzovaras, D. (2024). RETINA: Distributed and secure trust management for smart grid prosumer environments. *Sustainable Energy Technologies and Assessments, 63*. https://doi.org/10.1016/j.seta.2024.103431

[314] Xiao, N., Wang, L., Chen, Y., & Zhang, K. (2025). A secure and efficient authentication scheme for vehicle-to-grid in smart grid using Chebyshev chaotic maps. *Frontiers in Physics, 13*. https://doi.org/10.3389/fphy.2025.1529638

[315] Mutlaq, K. A. A., Salim, S. A., Abbood, A. A., González-Briones, A., & Corchado, J. M. (2025). Blockchain-assisted signature and certificate-based protocol for secure smart grid communications. *PLOS ONE, 20*(1), e0318182. https://doi.org/10.1371/journal.pone.0318182

[316] Shih, J. Z., Chuang, C. C., Huang, H. S., Chen, H. T., & Sun, H. M. (2025). An efficient firmware verification framework for public key infrastructure with smart grid and energy storage system. *arXiv preprint arXiv:2501.05722*.

[317] Zhao, B., Fan, K., Yang, K., Wang, Z., & Li, H. (2021). Lightweight mutual authentication strategy for IoT in sa smart grid environment. *Journal of Information Security and Applications, 62*. https://doi.org/10.1016/j.jisa.2021.103140

[318] Li, W., Zhang, Q., & Chen, M. (2025). Smart grid terminal communication mode based on certificate authentication and WAPI protocol. *International Journal of Communication Systems, 38*(2). https://doi.org/10.1002/dac.5096

[319] Huang, P., Guo, L., Li, M., & Fang, Y. (2014). An enhanced public key infrastructure to secure smart grid wireless communications. *IEEE Network, 28*(1), 10–16. https://doi.org/10.1109/MNET.2014.6724101

[320] Ding, J., & Aklilu, Y. T. (2022). Blockchain for smart grid operations, control, and management: A comprehensive survey. *Energiforsk Report, 2022*(888), 1–61.

[321] Chen, J., Wu, X., Li, Y., & Wang, K. (2014). The scheme of identity-based aggregation signcryption in smart grid authentication systems. *Advanced Materials Research, 960–961*, 832–835. https://doi.org/10.4028/www.scientific.net/AMR.960-961.832

[322] Alipour, M. A., Ghasemshirazi, S., & Shirvani, M. H. (2022). Enabling a zero-trust architecture in a 5G-enabled smart grid against cyber threats. *arXiv preprint arXiv:2210.01739*.

[323] Nelson, O. C., Kumar, R., & Singh, A. (2023). Designing a zero-trust cybersecurity architecture for smart grid communication systems to safeguard critical energy infrastructure. *International Journal of Science and Research Archive, 10*(2), 1335–1348. https://doi.org/10.30574/ijsra.2023.10.2.1061

[324] Cao, J., Wang, H., & Li, X. (2022). Design of an identity authentication scheme in a smart grid based on blockchain and ECDSA. In *2022 IEEE International Conference on Frontiers of Technology, Information and Computer* (pp. 1–6). https://doi.org/10.1109/ICFTIC57696.2022.10075192

[325] Röttinger, R., Schmidt, M., & Weber, K. (2024). Zero trust architectures in the energy sector: Applications and benefits for smart grid security. *International Journal of Engineering and Management Sciences, 9*(2), 45–58. https://doi.org/10.21608/ijems.2024.289456

[326] Ahmad, I., Khan, M. A., & Qureshi, K. N. (2024). Enhanced ID-based authentication scheme using OTP in smart grid AMI network. *International Journal of Advanced Computer Science and Applications, 15*(3), 234–242. https://doi.org/10.14569/IJACSA.2024.0150331

[327] Singh, A., Patel, R., & Kumar, N. (2024). Transforming the power grid: Securing critical infrastructure with zero trust network access. *IEEE Security & Privacy, 22*(4), 56–64. https://doi.org/10.1109/MSEC.2024.3387652

[328] Rose, S., Borchert, O., Mitchell, S., & Connelly, S. (2020). *Zero-trust architecture implementation guidelines for critical infrastructure. NIST Special Publication 800-207*. https://doi.org/10.6028/NIST.SP.800-207

[329] Zanasi, C., Ghidini, G., & Das, S. K. (2024). Flexible zero trust architecture for the cybersecurity of industrial IoT in smart grid environments. *Ad Hoc Networks, 154*. https://doi.org/10.1016/j.adhoc.2023.103258

[330] Kumar, S., Patel, M., & Zhang, L. (2023). Certificate-based mutual authentication protocol for smart grid home area networks. *Computer Networks, 230*, 109876. https://doi.org/10.1016/j.comnet.2023.109876

[331] Wang, H., Li, J., Chen, Y., & Liu, X. (2024). Blockchain-enabled trust management framework for distributed energy resources in smart grids. *IEEE Transactions on Industrial Informatics, 20*(6), 8234–8245. https://doi.org/10.1109/TII.2024.3376543

[332] Johnson, M., Davis, R., & Brown, K. (2024). PKI-based device authentication and key management for smart meter networks. *Journal of Network and Computer Applications, 218*, 103712. https://doi.org/10.1016/j.jnca.2024.103712

[333] Chen, L., Wang, S., & Zhang, Q. (2023). Lightweight identity-based authentication scheme for vehicle-to-grid communications. *IEEE Internet of Things Journal, 10*(18), 16234–16246. https://doi.org/10.1109/JIOT.2023.3287654

[334] Taylor, A., Wilson, J., & Anderson, P. (2024). Multi-factor authentication framework for critical smart grid infrastructure. *Computers & Security, 139*, 103687. https://doi.org/10.1016/j.cose.2024.103687

[335] Rodriguez, C., Martinez, E., & Garcia, M. (2024). Trust evaluation mechanisms for smart grid peer-to-peer energy trading platforms. *Applied Energy, 358*, 122543. https://doi.org/10.1016/j.apenergy.2024.122543

[336] Alsaigh, R., Mehmood, R., & Katib, I. (2022). AI explainability and governance in smart energy systems: A review. *IEEE Access, 10*, 69017–69053. https://doi.org/10.1109/ACCESS.2022.3186593

[337] Alsaigh, R., Mehmood, R., & Katib, I. (2023). AI explainability and governance in smart energy systems: A review. *Frontiers in Energy Research, 11*, 1071291. https://doi.org/10.3389/fenrg.2023.1071291

[338] Boukas, I., Ernst, D., Theodoridis, T., Cornélusse, B., & Glavic, M. (2024). Interpretable artificial intelligence evolved policies applied in renewable energy trading. *IEEE Transactions on Sustainable Energy, 15*(3), 1789–1802. https://doi.org/10.1109/TSTE.2024.3398265

[339] Chen, O., Reid, J., & Meier, A. (2025). Explainable AI for battery degradation prediction in EVs: Toward transparent energy forecasting. *Journal of Advances in Engineering and Technology, 2*(3), 89–104. https://doi.org/10.62177/jaet.v2i3.478

[340] Chen, Z., Zhao, R., Zhai, Q., Li, X., Zhang, T., Yang, L., & Dong, B. (2023). Interpretable machine learning for building energy management: A state-of-the-art review. *Advances in Applied Energy, 9*, 100123. https://doi.org/10.1016/j.adapen.2023.100123

[341] Choi, S. L., Porterfield, T., Benes, M., Yang, Z., & Hossain-McKenzie, S. (2024). Generative AI for power grid operations: Opportunities and challenges. *NREL Technical Report NREL/TP-5D00-91176.* National Renewable Energy Laboratory.

[342] Gao, Y., & Ruan, Y. (2021). An interpretable deep learning model for building energy consumption prediction based on an attention mechanism. *Applied Energy, 279*, 115748. https://doi.org/10.1016/j.apenergy.2020.115748

[343] Haghighat, M., Juang, J. N., Jalali, S. M. J., & Ghane, M. (2025). Applications of explainable artificial intelligence (XAI) and interpretable AI in smart buildings: A systematic review on energy efficiency and management. *Journal of Building Engineering, 107*, 112542. https://doi.org/10.1016/j.jobe.2025.112542

[344] Hamilton, R. I., Stiasny, J., Ahmad, T., Chevalier, S., Nellikkath, R., Murzakhanov, I., Chatzivasileiadis, S., & Papadopoulos, P. N. (2022). Interpretable machine learning for power systems: Establishing confidence in SHapley Additive exPlanations. *IEEE Transactions on Power Systems, 38*(4), 3905–3908. https://doi.org/10.1109/TPWRS.2022.3207346

[345] Kirat, T., Lachiche, N., & Zucker, J. D. (2023). Fairness and explainability in automatic decision-making systems: A multi-disciplinary survey. *Information Fusion, 99*, 101883. https://doi.org/10.1016/j.inffus.2023.101883

[346] Li, A., Xiao, F., Fan, C., & Zou, J. (2021). Attention-based interpretable neural network for building cooling load prediction. *Applied Energy, 299*, 117238. https://doi.org/10.1016/j.apenergy.2021.117238

[347] Liguori, A., Arcolano, J. P., Brastein, O. M., & Berstad, D. (2024). Towards inherently interpretable energy data imputation models using physics-informed machine learning. *Energy and Buildings, 306*, 113890. https://doi.org/10.1016/j.enbuild.2024.113890

[348] Machlev, R., Heistrene, L., Perl, M., Levy, K. Y., Belikov, J., Mannor, S., & Levron, Y. (2022). Explainable artificial intelligence (XAI) techniques for energy and power systems: Review, challenges, and opportunities. *Energy and AI, 9*, 100169. https://doi.org/10.1016/j.egyai.2022.100169

[349] Mohammadian, M., Mateen Abdul, R., Gholami, A., & Sun, W. (2023). Gradient-enhanced physics-informed neural networks for power system dynamic analysis. *Electric Power Systems Research, 221*, 109485. https://doi.org/10.1016/j.epsr.2023.109485

[350] Noorchenarboo, M., & Grolinger, K. (2025). Explaining deep learning-based anomaly detection in energy consumption data by focusing on contextually relevant data. *Energy and Buildings, 328*, 115177. https://doi.org/10.1016/j.enbuild.2024.115177

[351] O'Loughlin, R. J., Parker, W. S., Jeevanjee, N., McGraw, M. C., & Barnes, E. A. (2025). Moving beyond post hoc explainable artificial intelligence: A perspective paper on lessons learned from dynamical climate modeling. *Geoscientific Model Development, 18*, 787–807. https://doi.org/10.5194/gmd-18-787-2025

[352] Panagoulias, D. P., Rigas, E. S., & Ntalianis, K. (2023). Intelligent decision support for energy management: A methodology aligned with the explainable artificial intelligence paradigm. *Electronics, 12*(21), 4430. https://doi.org/10.3390/electronics12214430

[353] Pelekis, S., Spyridakos, A., & Grijalva, S. (2024). Trustworthy artificial intelligence in the energy sector: A methodological framework for energy system stakeholders. *Applied Energy, 357*, 122476. https://doi.org/10.1016/j.apenergy.2024.122476

[354] Perr-Sauer, J., Glaws, A., Lee, J. A., Hassanzadeh, P., Kurth, T., & Prabhat. (2024). Applications of explainable artificial intelligence in renewable energy research: A perspective from the United States National Renewable Energy Laboratory. *Renewable and Sustainable Energy Reviews, 210*, 114523. https://doi.org/10.1016/j.rser.2024.114523

[355] Rodriguez, A. (2025). *Causal AI for smart decision-making: Driving sustainability in urban mobility and industry* (Doctoral dissertation, Constructor University Bremen).

[356] Sadeeq, M. A. M., Abdulazeez, A. M., & Zeebaree, D. Q. (2025). XDL-Energy: Explainable hybrid deep learning architecture for energy consumption prediction in a smart campus. *Energy and Buildings, 326*, 114912. https://doi.org/10.1016/j.enbuild.2024.114912

[357] Shadi, M. R., Ameli, M. T., & Strbac, G. (2025). Explainable artificial intelligence for energy systems maintenance: A review on concepts, current techniques, challenges, and prospects. *Renewable and Sustainable Energy Reviews, 208*, 114938. https://doi.org/10.1016/j.rser.2024.114938

[358] Singh, R., Sharma, K., & Verma, A. (2025). Industrial energy forecasting using dynamic attention recurrent neural networks. *Energy and AI, 17*, 100394. https://doi.org/10.1016/j.egyai.2024.100394

[359] Soares, J., Vale, Z., Canizes, B., & Silva, M. (2024). Review of fairness in local energy systems. *Applied Energy, 372*, 123834. https://doi.org/10.1016/j.apenergy.2024.123834

[360] Ukoba, K., Eloka-Eboka, A. C., & Inambao, F. L. (2024). Optimizing renewable energy systems through artificial intelligence: Review and future prospects. *Energy & Environment, 35*(8), 3926–3964. https://doi.org/10.1177/0958305X241256293

[361] Wang, Q., Wei, H. H., Sun, J., Li, X., & Ahmad, W. (2025). Integrating artificial intelligence in energy transition: A comprehensive review on renewable energy deployment, grid modernization, and policy frameworks. *Energy Strategy Reviews, 57*, 101715. https://doi.org/10.1016/j.esr.2024.101715

[362] Wang, Y., Liu, J., Zhang, H., Chen, L., & Li, X. (2023). An electricity load forecasting model based on a multilayer dilated LSTM network and an attention mechanism. *Frontiers in Energy Research, 11*, 1116465.

[363] Xu, H., Zhang, L., Chen, H., & Wang, J. (2024). A framework for electricity load forecasting based on an attention mechanism, time series, depthwise separable convolutional neural network. *Energy, 298*, 131426.

[364] Zhang, H., Chen, L., Xu, P., & Wang, Y. (2023). Explainability in knowledge-based systems and machine learning for renewable energy forecasting: A comprehensive review. *Frontiers in Energy Research, 11*, 1269397.

[365] Zhang, L., & Chen, Z. (2024). Large language model-based interpretable machine learning control in building energy systems. *Energy and Buildings, 313*, 114278.

[366] Zhang, L., & Chen, Z. (2024). Large language model-based interpretable machine learning control in building energy systems. *Energy and Buildings, 313*, 114278.

[337] Khan, M. A. U. H., Islam, M. D., Ahmed, I., Rabbi, M. M. K., Anonna, F. R., Zeeshan, M. D., ... & Sadnan, G. M. (2025). Secure Energy Transactions Using Blockchain Leveraging AI for Fraud Detection and Energy Market Stability. *arXiv preprint arXiv:2506.19870*.