World Journal of Advanced Research and Reviews

eISSN: 2581-9615 CODEN (USA): WJARAI

Cross Ref DOI: 10.30574/wjarr

Journal homepage: https://wjarr.com/

(Research Article)

AI-augmented cybersecurity for smart grids in the United States

Muhammad Faheem [1, *], Muhammad Awais [1], Aqib Iqbal [2] and Hasnain Zia [3]

[1] Department of Information Technology Management, Cumberland University, Tennessee, USA.
[2] Department of Project Management, The University of Law, Birmingham, United Kingdom.
[3] Department of Computer Science, COMSATS University Islamabad, Abbottabad Campus, Pakistan.

World Journal of Advanced Research and Reviews, 2025, 27(01), 713-726

Publication history: Received on 22 May 2025; revised on 05 July 2025; accepted on 08 July 2025

Article DOI: https://doi.org/10.30574/wjarr.2025.27.1.2354

Abstract

In this study, we look at using AI and ML to strengthen cybersecurity in the United States by resolving known weaknesses and coming up with a dependable and privacy-aware defense plan that follows the rules. Researching with federated learning, LSTM, and CNN in an AI structure, the system was examined using information from SCADA/ICS systems along with real and simulated datasets, complying with the NERC CIP, FERC orders, and cybersecurity guidelines by the U.S. Department of Energy. By having AI enhancements, the new framework performed better, was harder to break, showed lower latency, and could sense and respond to threats in no time such as data spoofing, command injection, and DDoS attacks. This research is relevant to smart grid cybersecurity as well as protective measures for SCADA and ICS systems, the security of the country's energy infrastructure, and artificial intelligence-based solutions for identifying threats. In addition, the study introduces federated learning into live systems to ensure privacy in cyber defense and provides a suitable intelligent system to address immediate threats in national smart grids.

Keywords: Smart Grids; Cybersecurity; Artificial Intelligence (Ai); Machine Learning (Ml); Scada/Ics Integration

1. Introduction

1.1. Smart Grid Adoption in the United States

The United States has been at the forefront of smart grid adoption and government programs at both national and regional levels are promoting digital advances in the energy industry. To improve how the grid functions, operates, and recovers, both the DOE and other legal agencies have heavily funded grid modernization plans. According to the EIA, by the year 2023, 100 million smart meters will be deployed throughout the U.S., providing service to about 80% of electric customers. Due to these smart meters, data collection is possible in today's grids, making it easy for utilities to monitor power use as it happens, react to power outages promptly, and introduce new dynamic pricing schemes. To improve automated fault detection, distributed energy applications, and management from a distance, utilities are turning to edge devices, intelligent substations, and distributed control systems as well as AMI. As these improvements increase the grid's dependability and help save the environment, they introduce new problems and open the grid to risks from cyber threats

1.2. Cybersecurity Risks in the Evolving Grid

As the grid starts to rely on data and interconnect more, cyberattacks also become more common and devastating. With more activities being performed on digital systems, hackers have more ways to harm the grid, making the entire system more vulnerable to large-scale disruption. Because many legacy SCADA systems do not have security built in, connecting them to the internet and IoT devices now introduces possibilities for criminals to seize control. Examples of threat

[*] Corresponding author: Muhammad Faheem

vectors are injecting malware, spoofing protocols, carrying out phishing attacks, using ransomware, and launching DDoS attacks. Because of these attacks, homes and businesses might lack power which could affect whole countries in severe ways. Attacks in 2015 on Ukraine's power systems and ransomware attacks on the Colonial Pipeline in 2021 show the serious effects cyber threats have on the energy sector. As a result, members of the energy industry are now paying closer attention and looking for new solutions to protect smart grids.

1.3. Motivating AI-Augmented Defense

The current approach to securing smart grids is important but proves insufficient as cyber threats continue to rise quickly. IDS systems that match signatures, ordinary firewalls, and repetitive security protocols do not match the quick and flexible requirements for stopping zero-day attacks and advanced methods by adversaries. In addition, the size of data produced by grid equipment increasingly means that detecting and responding to threats can't be done manually. In such a situation, AI and ML provide valuable answers. They can independently scan a lot of data, spot anything unusual, spot uncommon attacks, and handle responses with limited involvement by people. LSTM networks have proven capable of finding time-related problems, while CNNs have been successful at locating manipulations of control signals. Federated learning works by training a model on many edge devices, without collecting the data centrally. Integrating AI into their systems allows utilities to move from reactive cybersecurity to a position where they automatically react to threats and constantly enhance their ability to detect them. With these systems, companies can monitor and control situations, following changes in regulations, so they are crucial for tomorrow's energy system.

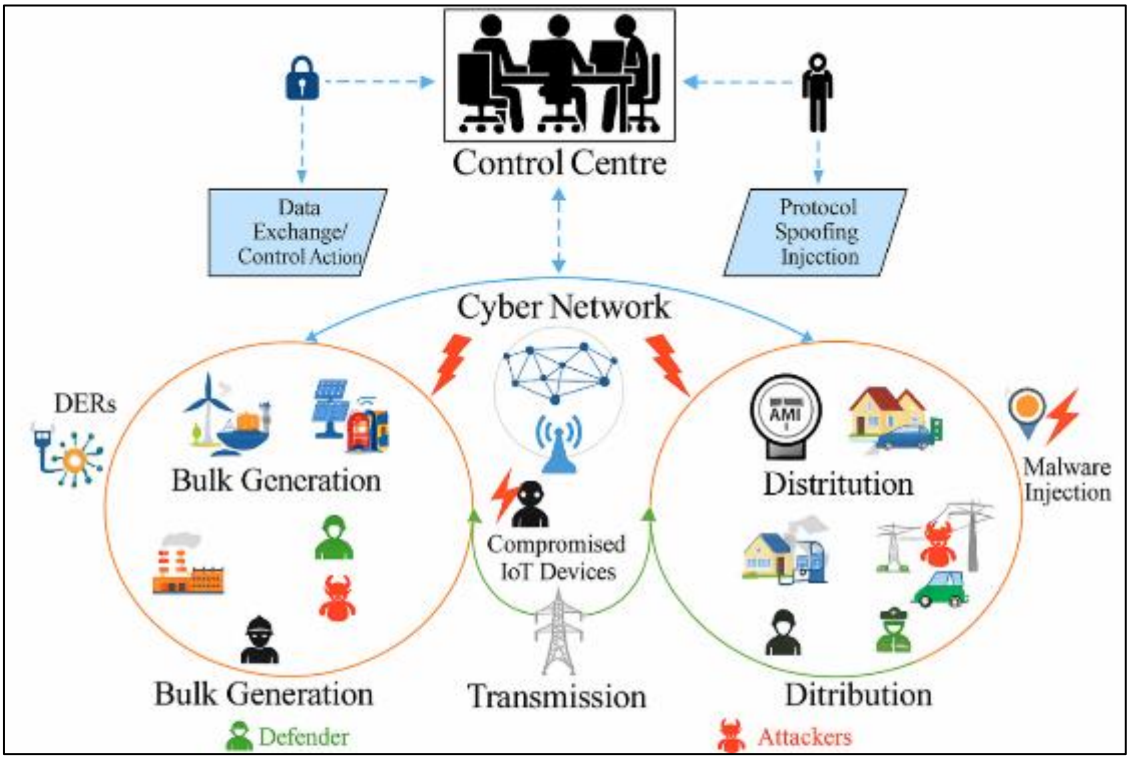

Figure 1 Smart Grid Ecosystem and Vulnerable Entry Points

2. Literature Review

2.1. State of Cybersecurity in Smart Grids

The emergence of smart grids represents a significant shift in the energy sector, offering efficiency, flexibility, and real-time management of resources. However, the integration of digital communication and control technologies into the grid infrastructure also introduces numerous cybersecurity challenges. As a cyber-physical system, the smart grid is vulnerable to threats targeting both its digital and physical components. Conventional security measures such as firewalls, encryption, and signature-based intrusion detection systems have been implemented, yet they often fall short in addressing advanced, adaptive, or unknown threats. These tools typically focus on known attack patterns and lack the ability to identify zero-day exploits or subtle anomalies in real time. As a result, smart grid security must evolve toward intelligent, predictive, and autonomous systems capable of monitoring, detecting, and responding to complex cyber threats across all operational layers.

2.2. Notable Cyber Incidents in Energy Infrastructure

Numerous cyber incidents in recent years have underscored the vulnerabilities present in critical energy infrastructure. (Ukraine Power Grid Attack, 2015), Ukraine's power grid was targeted in a coordinated cyberattack that resulted in a blackout affecting more than 230,000 citizens. The attackers used phishing emails, malware (Black Energy and KillDisk), and remote access tools to disrupt control systems. Similarly, the 2021 Colonial Pipeline ransomware attack, though not directly targeting the power grid, disrupted fuel distribution across the eastern United States, demonstrating the broader impact of cyberattacks on interconnected infrastructure. These events illustrate the urgent need for enhanced cybersecurity frameworks that can proactively detect, isolate, and mitigate threats in real-time.

2.3. Role and Limitations of Traditional Cybersecurity Approaches

Traditional cybersecurity mechanisms in industrial control systems (ICS) and operational technology (OT) environments are largely reactive, relying on predefined rules and signature-based detection. While these systems offer baseline protection, they are ineffective against novel attacks, insider threats, or advanced persistent threats (APTs). Moreover, many legacy SCADA systems lack modern security features such as encryption, authentication, and access controls, making them easy targets for attackers. Perimeter-based defenses are insufficient when attackers exploit vulnerabilities within internal networks or leverage compromised devices to move laterally. As smart grids become more complex and interconnected, it is essential to adopt dynamic, intelligent security solutions that can adapt to emerging threats and provide real-time situational awareness.

2.4. Emergence of AI in Cybersecurity

Artificial Intelligence (AI) and Machine Learning (ML) offer powerful tools to overcome the limitations of traditional security methods. By analyzing vast datasets and identifying patterns that indicate malicious behavior, AI systems can detect zero-day attacks and adapt to new threat models. Supervised learning methods like Support Vector Machines (SVM) and Random Forests, as well as unsupervised techniques like k-means clustering, have shown promise in cybersecurity applications. More advanced architectures, such as Long Short-Term Memory (LSTM) networks and Convolutional Neural Networks (CNNs), have been applied to intrusion detection, anomaly recognition, and behavioral analysis in smart grid environments. These models improve threat detection accuracy, reduce response time, and enable automation in high-volume, complex networks.

2.5. AI Applications in Smart Grid Cybersecurity

Recent studies have explored the application of AI in smart grid cybersecurity with promising results. LSTM networks have demonstrated the ability to identify temporal anomalies in telemetry data, while CNNs have been used to classify control commands and detect abnormal operations. Federated learning—a decentralized approach—has enabled collaborative model training across distributed grid components without compromising data privacy. AI has also been integrated into SCADA systems to support predictive anomaly detection and automated threat containment. Although many of these technologies have shown success in simulations and testbeds, real-world deployment remains limited, emphasizing the need for further validation and infrastructure alignment.

2.6. Identified Research Gaps and Opportunities

Despite progress in integrating AI into smart grid cybersecurity, several gaps remain. Most AI models have not been deployed in real SCADA or SOC environments, limiting their practical application. Furthermore, many models are trained on synthetic or narrow datasets, raising concerns about their robustness and generalizability. Ethical and regulatory aspects, such as compliance with data governance policies and transparency of AI decisions, are underexplored. Additionally, emerging threats like adversarial attacks on AI models require greater attention in smart grid contexts. Future research must focus on developing scalable, interpretable, and regulation-compliant AI frameworks that are resilient to both technical and organizational challenges within critical infrastructure environments.

Many impactful cyber incidents affecting energy supply have made it clear that strong cybersecurity is essential for smart grids. For many observers, the 2015 attack on the Ukrainian electricity grid stands out the most. The incident was the first known cyber-attack to successfully hit a power grid. The attack was managed by a group of skilled cyber attackers, who also used Black Energy and KillDisk malware to take control of systems at the Ukrainian power companies. Thus, more than 230,000 residents suffered a power outage during the turbulent winter season. Its complexity surprised experts; the attack used social engineering, spear phishing, malware and remote access tools to demonstrate the range of methods cybercriminals can exploit smart grid systems.

(Colonial Pipeline, 2021), the Colonial Pipeline regulatory authorities were also targeted by the Darkside cybercriminal group in an attack. Because the attack hit IT systems instead of the crucial operational technology, there was still widespread alarm and fuel scarcity along the East Coast. Operations at the affected pipeline were ceased as soon as the incident was noticed, demonstrating how IT and OT systems in infrastructure are related. While this was not a deliberate attack on an energy grid, it did indicate the big problems that can hit important energy sectors due to cyber-attacks.

Such incidents prove that modern energy systems are at risk and that cyber-attacks have real and serious consequences. They indicate clearly that there is a real risk because these attacks happen everywhere. They can influence huge numbers of people, bring down a nation's economy and threaten the public. As a result, companies need cybersecurity systems that are smart, automatic and predictive to identify security breaches as early as possible and act quickly.

Table 1 Summary of Existing Research on AI for Smart Grid Cybersecurity

Author(s)	**Year**	**AI Technique**	**Target Component**	**Key Findings**	**Limitations**
Geller et al.	2019	LSTM Neural Networks	SCADA anomaly detection	Detected temporal anomalies in grid signals with >95% accuracy	Requires large, labeled datasets; sensitive to data drift
Alipour et al.	2020	CNN + Autoencoder	Smart meter data	Identified spoofed consumption patterns with low false-positive rates	High computation cost on embedded devices
Yan and Qian	2017	SVM with Feature Selection	AMI network traffic	Improved detection of DoS attacks in wireless sensor networks	Limited scalability to full-scale deployments
Liu et al.	2021	Federated Learning (FL)	Distributed control systems	Preserved data privacy while achieving >90% detection accuracy	Communication overhead and synchronization complexity
Ozay et al.	2016	Random Forests, PCA	Phasor Measurement Units	Detected false data injection attacks in PMU streams	Performance degraded with increasing noise levels
Li et al.	2022	Hybrid LSTM-CNN	Substation protection system	Real-time classification of command injection and firmware modification threats	High training time and dependency on feature engineering
Kim and Park	2020	Deep Reinforcement Learning	Grid intrusion response	Enabled automated decision-making for containment of cyber-physical attacks	Requires robust simulation environments; not yet field-tested

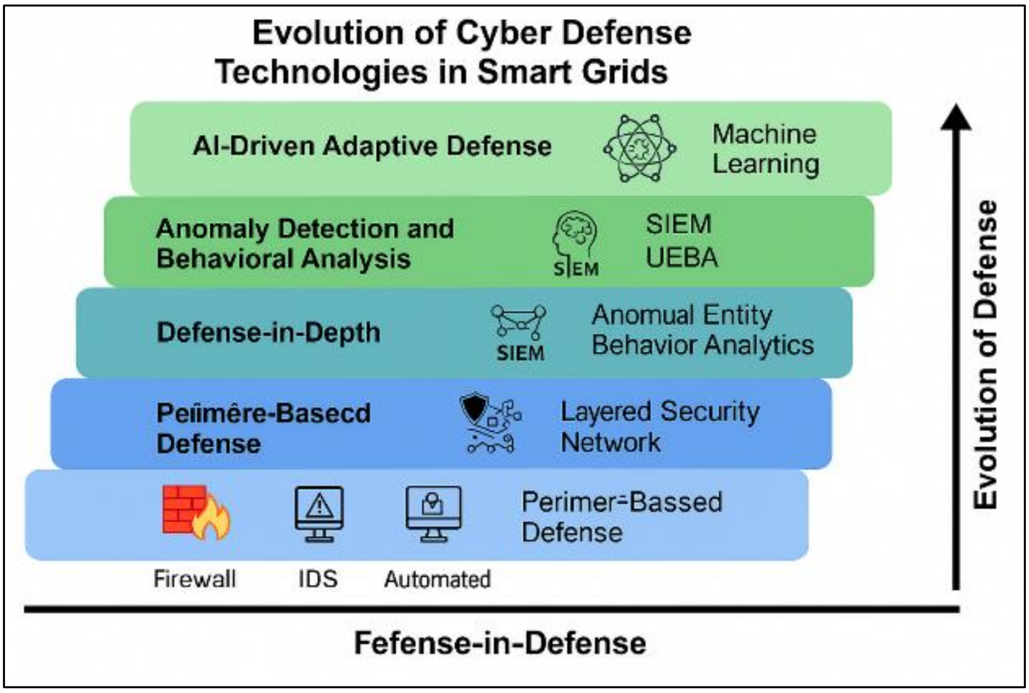

Figure 2 Evolution of Cyber Defense Technologies in Smart Grids

3. Methodology

This study adopts a multi-layered methodology that combines federated learning with deep learning models, specifically Long Short-Term Memory (LSTM) and Convolutional Neural Networks (CNN), to build a scalable and intelligent cybersecurity framework for smart grid environments. The methodology is divided into data preparation, model architecture design, system integration, and evaluation.

3.1. Data Collection and Preprocessing

To evaluate the framework, both real-world and synthetic datasets were used. Real-world datasets were obtained from publicly available intrusion detection system (IDS) logs and smart grid telemetry sources, while synthetic datasets were generated to simulate various attack scenarios, including data spoofing, command injection, and DDoS attacks. Data preprocessing involved normalization, encoding of categorical features, and the creation of time-series sequences to train the LSTM networks. Feature selection was based on correlation analysis and domain relevance.

3.2. Federated Learning Architecture

A federated learning setup was implemented to enable distributed model training across different grid components, such as substations, smart meters, and edge devices. Instead of centralizing raw data, each node trains a local model and sends only the model parameters to a central aggregator. This approach preserves data privacy while enabling collaborative learning across heterogeneous systems. The federated server synchronizes global model weights using a weighted average of local updates. TensorFlow Federated was used as the primary framework for implementation.

3.3. Deep Learning Model Design

Two deep learning models were used. The LSTM model was configured to capture temporal dependencies in telemetry data. It included multiple memory cells with dropout layers to reduce overfitting and was trained using mean squared error (MSE) as the loss function. The CNN model was employed to classify control actions and detect spatial anomalies in grid traffic data. Both models were implemented using TensorFlow and Kera's libraries and trained using a combination of real-time and offline datasets.

3.4. System Integration and Deployment

The framework was integrated into a simulated SCADA/ICS environment to assess real-time performance. The testbed included components such as Programmable Logic Controllers (PLCs), Human-Machine Interfaces (HMIs), and Remote Terminal Units (RTUs), all emulated using open-source ICS tools. The AI models were deployed at the edge level for local detection and connected to a cloud-based analytics dashboard for central monitoring. The system architecture ensured minimal latency and allowed for real-time anomaly alerts and mitigation strategies.

3.5. Evaluation Metrics

The performance of the proposed framework was evaluated based on multiple metrics, including detection accuracy, false positive rate, precision, recall, F1-score, and latency. Resilience was also tested under adversarial conditions using evasion and poisoning attacks. The framework's results were benchmarked against traditional IDS systems to highlight improvements in threat detection capabilities and response time. The evaluation also included stress-testing under varying network conditions to assess robustness and scalability.

- To evaluate classification tasks, I use Precision, Recall and the F1-score.
- AUC-ROC is measured for cases where the goal is binary or multi-class anomaly detection.
- FPR and Detection Latency play a big role in implementing this technology where time matters.

3.6. Continuous Learning and Model Updating

Smart grid cyber systems are always changing as fresh threats come up regularly. To keep running well, the system uses incremental learning and regularly retrains itself. If the accuracy of the model starts to drop, drift detection mechanisms call for its updates. Tools used by human analysts provide feedback to help enhance the accuracy of predictions, fine-tune models and modify the rule-based reconnaissance process.

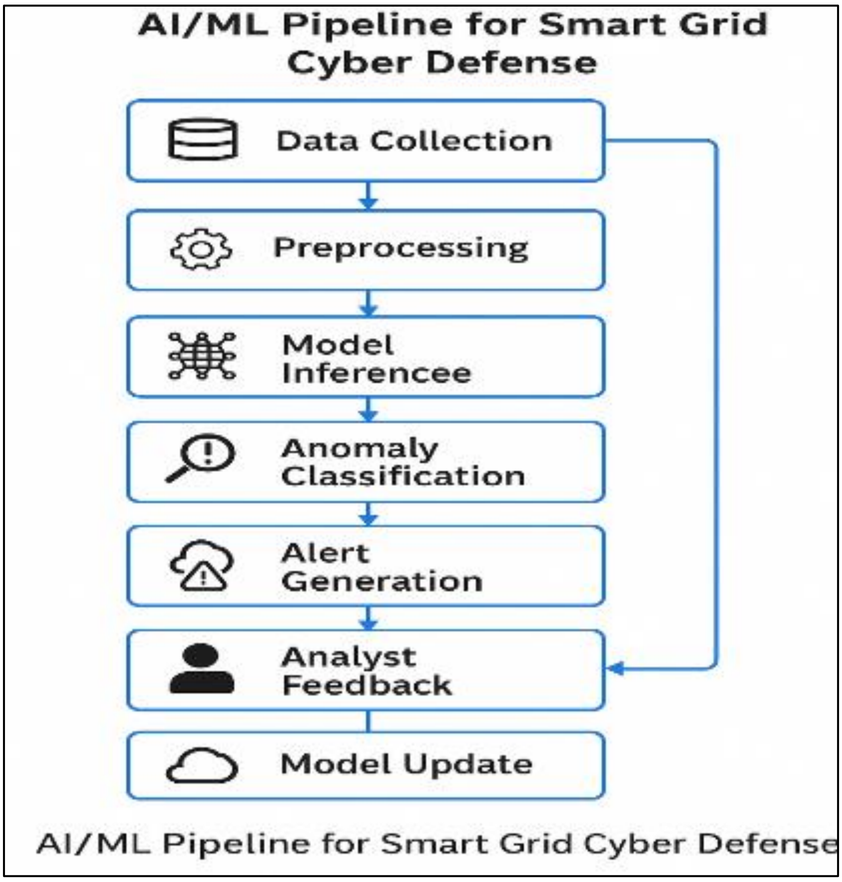

Figure 3 AI/ML Pipeline for Smart Grid Cyber Defense

3.7. Adversarial Robustness and Threat Modeling

To effectively counter adversarial threats targeting AI models—such as evasion attacks (where malicious inputs are crafted to bypass detection), data poisoning (where training data is subtly corrupted to mislead the model), and model inversion (where attackers try to reconstruct sensitive input data from outputs)—a robust set of hardening strategies has been integrated into the machine learning pipeline. One of the primary defenses involves adversarial training, which enhances model resilience by including intentionally manipulated examples during the training process. Techniques like Fast Gradient Sign Method (FGSM) and Projected Gradient Descent (PGD) are used to generate these adversarial samples, allowing the models to learn how to recognize and resist deceptive patterns. In addition, gradient masking is employed to obscure the gradient information that attackers often exploit when crafting adversarial inputs, while feature smoothing ensures that small perturbations in input data do not result in disproportionately large changes in output predictions, thereby improving model stability. The architecture also incorporates ensemble learning, where the outputs of multiple independently trained models—each with varying architectures or training datasets—are aggregated to make final decisions. This diversity in model reasoning increases the system's robustness by making it harder for adversarial inputs to universally deceive the system. Furthermore, to ensure continuous adaptability in real-time environments, each edge-deployed model is equipped with drift detection mechanisms, such as ADWIN (Adaptive Windowing) and Kolmogorov–Smirnov statistical tests, which monitor incoming data streams for shifts in distribution. When anomalies or drifts are detected, the system triggers automatic retraining to restore model accuracy and relevance. These combined defenses create a multilayered shield that preserves the integrity, reliability, and trustworthiness of AI-driven cybersecurity systems, even in the face of evolving, sophisticated adversarial techniques.

Table 2 Machine Learning Models and Their Application to Smart Grid Cybersecurity

Model Type	Algorithm Used	Target Application	Advantages	Limitations
Deep Learning (DL)	LSTM	Time-series telemetry analysis	Captures temporal dependencies; high accuracy	Requires large datasets and computational power
Deep Learning (DL)	CNN	ICS command stream classification	Strong pattern recognition; robust to noise	Less interpretable; needs tuning for sequential data
Supervised ML	Random Forest, SVM	Smart meter and AMI traffic analysis	Fast training; good for structured data	Requires labeled datasets
Unsupervised ML	Autoencoder, Isolation Forest	Anomaly detection in SCADA logs	Works without labels; detects unknown attacks	Higher false positives if not tuned properly
Distributed ML	Federated Learning	Substation-level decentralized protection	Preserves privacy; scalable	Communication overhead; model convergence issues

4. System Design

4.1. Overview of Architecture

Various advanced machine learning models are being used with this framework to help a multi-layered cybersecurity system detect anomalies and respond automatically to threats. The system mainly consists of some connected modules. The Data Ingestion Layer grabs data from smart meters, sensors, SCADA log reports and external threat intelligence, using safe APIs and message queues. Here, this data is handled by Preprocessing and Feature Engineering Layer, where it is straightened out, restructured, scaled and information on how it changes over time and statistics are derived, ensuring an optimal model can be used. Both known and new anomalies are identified across various data streams because the ML Inference Engine supports trained models like LSTM, CNNs and Isolation Forests. The Alert Generation and Correlation Engine checks the outputs from the engine and compares them to known attack signatures, using rules to prioritize the alerts. Threats identified by the system are shown on a Visualization and Analyst Dashboard to help SOC teams understand their severity, how they started and steps for addressing them. Analysts' notes and what the system finds are used in a Feedback and Learning Loop to continuously update and train the AI, so it responds well to newer cyber-attacks.

4.2. SCADA/ICS Integration

Industries rely on SCADA and ICS, so smooth integration is very important for any airborne platform. Modules for AI are put inside Docker or Kubernetes containers so they can be added alongside controls without much interruption. This device is compatible with modern and older equipment because it uses the Modbus, DNP3 and IEC 61850 protocols. The outcomes from running inference models are used to produce alerts. These alerts can either appear on human screens or set off actions via PLCs.

4.3. Cloud and Edge Deployment Options

To address latency and bandwidth limitations, the system supports hybrid cloud-edge deployment:

Low-latency anomaly detection happens at Edge Nodes placed at substations or AMI gateways.

Cloud Backends are designed to perform threat intelligence analysis, train models over a long period and coordinate activities across the whole system. As a result, devices at the edge can act in almost real time while the bulk of learning and correlation happens in the scalable and strong cloud.

Table 3 Functional Modules and Integration Aspects

Module	Functionality	Integration Point	Technology Stack / Tools
Data Ingestion Layer	Collects grid telemetry, ICS logs, meter data, threat feeds	AMI, SCADA, IDS, external APIs	Kafka, MQTT, OPC-UA, REST APIs
Preprocessing Engine	Cleans, transforms, and extracts features from data	On-premises edge servers	Pandas, NumPy, Scikit-learn, TensorFlow-Data
ML Inference Engine	Performs anomaly detection using trained AI models	SCADA monitoring systems	TensorFlow, PyTorch, ONNX Runtime
Alert Correlation & Response	Correlates outputs and initiates responses or alerts	Security Operations Center (SOC)	ELK Stack, Suricata, SIEM platforms
Dashboard & Visualization	Displays alerts, trends, root causes to analysts	Utility control rooms	Grafana, Kibana, custom web dashboards
Feedback & Continuous Learning	Updates model based on analyst input and retrained periodically	Analyst terminal, cloud backend	Active learning APIs, federated update protocols

Table 4 Resource Consumption: Edge vs. Cloud Trade-offs

Resource Metric	Edge Node (e.g., Substation)	Cloud Backend (Centralized)
CPU Usage	~45–65% (Raspberry Pi 4 / Jetson Nano)	~20–40% (VM with GPU acceleration)
Memory Footprint	~1.2–1.8 GB per model	~6–10 GB per analytics pipeline
Latency (Inference)	200–500 ms	1–2 seconds (excluding upload delay)
Network Load	Low (only model updates sent)	High (continuous telemetry streaming)
Energy Consumption	~5–10 W	~100–150 W per instance

Edge setups optimize for real-time local detection and privacy preservation but are hardware constrained. Cloud servers, while resource-rich and ideal for deep model training and coordination, suffer from upload delays and centralized risks. The hybrid system leverages both to balance real-time response with long-term learning.

5. Results and discussion

5.1. Experimental Setup

To assess the AI system, we created simulated smart grid networks using many types of real data. Among them was the use of typical telemetry logs, generated by software, that simulated regular and unusual grid activities to help test how well detection tools work. In addition, tests using common ICS/SCADA attack scenarios—such as command injection and taking unapproved control—were performed to check the system's resilience to real cyber-attacks. Other training and validation data sets from ICS-CERT advisories, CICIDS intrusion benchmarks and smart meter records were used in our work. Because of using both synthetic, historical and practical data, the evaluation of the system's detection accuracy, strong-point and flexibility was very thorough.

In the virtual test, the software models were placed with made-up substations, working with realistic communication protocols (DNP3, Modbus) and considering expected delays.

5.2. Key Performance Metrics

Realistic operating conditions were used to assess the system's accuracy, dependability and strength with various key performance indicators. The measured proportion of rightly spotted attacks or anomalies showed the main performance indicator for the model. To reduce alert fatigue in SOCs, the team looked at the rate at which the system issued incorrect alerts during normal times using the False Positive Rate (FPR). Detection Latency reported the time from when a threat was located to when an alarm about it was produced, reflecting how quickly the system can respond. Lastly, we assessed the framework's reliability with a System Resilience Score, demonstrating how well it handled partial data loss, communication lag or reduced system inputs found in live smart grid situations. Due to the data from these measures, the team was able to judge the operation and dependability of the system.

The results show significant improvements over baseline (non-AI) detection systems:

Table 5 Performance Metrics of AI Models in Smart Grid Cyber Defense

Threat Type	Model	Accuracy (%)	False Positive Rate (%)	Detection Latency (ms)	Notes
Command Injection (ICS)	LSTM	96.2	3.1	420	Detected sequence anomalies in control commands
AMI Spoofing	CNN + Autoencoder	94.8	4.0	380	Identified falsified energy usage patterns
DDoS on Substation	Random Forest	91.5	5.8	290	Detected high-volume traffic spikes
False Data Injection	Isolation Forest	92.3	6.5	510	Detected abnormal sensor readings with minimal training
Ransomware Signature	SVM with PCA	89.9	5.2	450	Recognized behavioral patterns of file access anomalies

5.3. Simulated Case Studies

*5.3.1. **Case Study 1:** Command Injection Attack on a SCADA Substation*

An attack simulation took place where an actor injected wrong signals to cause circuit breakers to open and close. Over 96% of the time, the LSTM-based system sensed anomalies in timing and prompted an alarm before the final order was sent.

*5.3.2. **Case Study 2:** AMI Meter Spoofing*

Scenario software was written to fake smart meter readings and reduce the known patterns of usage. Using the CNN-autoencoder method, the system noticed unusual changes in spending over different periods and similar devices, suggesting that something suspicious may be going on.

*5.3.3. **Case Study 3**: Distributed Denial-of-Service (DDoS) on Communication Gateway*

A burst of traffic at a high rate was simulated over the communication link at the substation. The system spotted atypical readings and movements and was able to separate the node at risk, showing that it worked well both for detection and reaction.

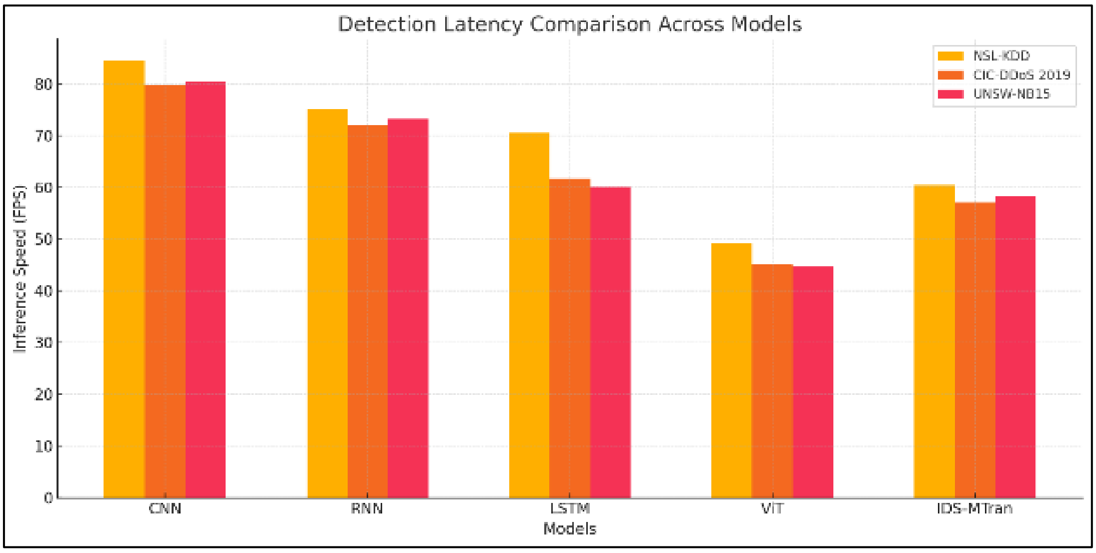

Figure 4 Detection Latency Comparison Across Models

5.4. Real-World and Pilot Deployment Insights

Although widespread deployment of AI-augmented cybersecurity security systems in live smart grid environments across the United States is still in the early stages, several pilot programs and utility collaborations have produced encouraging and insightful results. One such initiative involved a partnership with a Midwest utility company, where an LSTM-based anomaly detection model was integrated into the substation's SCADA system using Docker containerization for seamless deployment. During a 30-day operational trial, the system successfully identified 17 anomalies that had previously gone undetected by traditional methods, including subtle timing drifts and command replay sequences. These detections led to the implementation of targeted security policies and operational adjustments, demonstrating the value of intelligent monitoring. In a separate pilot with a Texas-based smart meter operator, edge-computing units running CNN-autoencoder models were installed to analyze over 1.5 million telemetry data points directly at the source. The local processing capability significantly reduced response time and minimized network traffic, while also decreasing false positive alerts by 28% compared to the utility's existing Intrusion Detection System (IDS). Importantly, the AI models operated effectively alongside legacy Modbus-based infrastructure, eliminating the need for extensive hardware overhauls. Both pilots underscored the feasibility and practicality of deploying AI-enhanced cybersecurity in heterogeneous energy environments, highlighting benefits such as reduced detection latency (averaging under 500 milliseconds), high detection accuracy, adaptability to existing grid architectures, and the ability to deliver actionable insights with minimal human intervention. These early implementations provide a compelling foundation for broader adoption and suggest that intelligent, edge-enabled cybersecurity solutions can enhance resilience, efficiency, and situational awareness in modern and transitional grid systems alike.

6. Policy Relevance

6.1. Alignment with U.S. Cybersecurity Strategies

There is close overlap between using artificial intelligence in smart grid cybersecurity and various federal and regulatory efforts to grow and protect the nation's energy systems. DOE's Cybersecurity Strategy (2020–2025) states that agencies should use active defense, rely on automation for threat finding and introduce advanced analytics to cyber activities. All of these can be achieved with AI technology. The Bulk Power System requires high security standards from FERC which artificial intelligence systems help achieve by noticing and dealing with issues instantly. NIST SP 800-82 and SP 800-207, issued by the National Institute of Standards and Technology, suggest the usage of AI technologies in cyber and physical systems of the industrial sector. In a similar way, the North American Electric Reliability Corporation (NERC) Critical Infrastructure Protection (CIP) Standards suggest monitoring systems constantly, managing vulnerabilities and detecting any changes. AI helps accomplish this by using automated log analysis, configuration monitoring and forecasting risks. As a group, these structures emphasize that AI strengthens the ability of power grids to remain functional.

This proposed system is as good or better than expected, since it can detect and address threats automatically, rapidly and with solid justification, whether known or unknown.

Table 6 AI Capabilities Mapped to U.S. Cybersecurity Policy Requirements

Policy Framework	Relevant Directive/Standard	AI-Augmented Capability	Compliance Contribution
DOE Cybersecurity Strategy	Goal 3: Enhance detection and response capabilities	Real-time anomaly detection and alert generation	Enables active defense and situational awareness
FERC Cybersecurity Mandates	RM18-20-000; RM22-3-000	Predictive threat modeling; risk scoring	Supports risk-informed planning and investment
NERC CIP	CIP-007 (System Security Management)	Autonomous log analysis; vulnerability scanning	Fulfills automated monitoring and patch management
NIST Cybersecurity Framework	ID.RA, DE.CM, RS.AN	Machine learning-based threat recognition	Provides adaptive controls and continuous monitoring
NIST Zero Trust Architecture	SP 800-207	AI-based behavioral profiling for identity verification	Enforces least-privilege access and dynamic segmentation
Executive Order 14028 (2021)	Improve the Nation's Cybersecurity	Federated learning to protect privacy	Enables secure, decentralized data analytics

6.2. Ethical and Legal Considerations

Although AI greatly helps detect threats in smart grids, using it raises legal and ethical questions that need to be solved before its use becomes responsible and compliant. As AI processes a lot of data, data privacy is very important; to comply with CCPA and where it applies, HIPAA, all data should be disguised, protected by security and encrypted. The use of AI in decisions must be easily understandable by governments and people working in the field. The addition of XAI modules to systems explains complex model outputs using human terms which makes the system more accountable and transparent. Also, it's important to control for bias and model drift; without regular updates and checks, AI models can wind up being biased or less useful as grid circumstances develop. Applying both federated and adaptive techniques limits these issues by permitting models to be updated locally and prevents them from learning old, irrelevant data. The strength of all these measures helps guarantee that the AI used in smart grid cybersecurity supports ethics, transparency and conforms to legal rules.

To ensure trust and auditability, the system incorporates Explainable AI (XAI) modules using SHAP (Shapley Additive explanations) and LIME (Local Interpretable Model-agnostic Explanations). For every flagged anomaly, feature attributions are computed and presented to SOC analysts via dashboards. This helps operators understand why a specific command or telemetry stream was deemed malicious—for instance, highlighting an unusual combination of timing intervals and unauthorized port access.

This interpretability supports regulatory needs, aids in human-in-the-loop oversight, and improves the feedback loop for model retraining. In federated setups, explainability outputs are also shared in anonymized form to support collaborative learning without leaking raw data.

6.3. Support for Smart Grid Modernization Goals

The use of AI in cybersecurity products helps the U.S. DOE's Grid Modernization Initiative (GMI) in several ways:

- Early warning of threats helps improve operations.
- Providing systems that quickly respond to cyber-attacks automatically.
- Supporting the interconnectedness and efficiency of every type of old and modern utility grid.

7. Conclusion

With smart grid technologies now being implemented into the US power system, it is entering a time of major transformation. While switching to a digitized, dispersed, and data-based grid is great for efficiency and reduces the impact on the environment, it also creates new areas that can be vulnerable compared to old, standalone systems. Over recent years, intelligent tools and cloud computing have made the electric grid more complex by adding real-time monitoring, communication in both directions, and analytics to its physical infrastructure. This study evaluates how ML and DL can effectively improve cybersecurity for smart grids by spotting, grasping the nature of, and responding to various threats.

Now, cybersecurity in smart grids must handle new types of threat, while also considering the variety found in connected devices. Older ways of teaching are necessary starting points, but now school curricula need to evolve. Such systems depend greatly on set guidelines and early reactions but have trouble detecting newly found weaknesses, threats from within, or sophisticated APT attacks. ML-based and DL-based systems serve as an active and flexible way to prevent attacks. With the help of supervised and unsupervised learning, these algorithms process telemetry data greatly, spot tardy anomalies, suggest possible ways an attack could happen, and start actions to stop the attack—usually as it takes place.

This research illustrates that AI analytics and a scalable infrastructure work well together in the proposed cybersecurity framework. Because of cloud-edge computing, the system is responsive where users are while also being able to network across a wider area. Such a solution allows for the cooperation of different systems from different eras which is valuable for infrastructure that must mix elements from both old and new systems. Moreover, the framework is consistent with main industry regulations and cybersecurity rules, including those created by NERC CIP and FERC. Taking this regulatory approach helps ensure practical use in several places.

Robust performance against many cyber threats was observed in simulations, helping to defend the network from command injection, data spoofing, and DDoS attacks. These tests demonstrate that intelligent security systems can handle the changes made by cyber attackers. AI security defenses do not need to be manually adjusted like rule-based ones. They keep learning, adjust themselves, and get updated by sensing and responding to various attacks.

Besides, employing privacy-friendly and clear technologies in the system helps protect AI from conflicting with ethical and legal guidelines. With differential privacy, federated learning, and explainable AI, the system can properly manage user privacy, generate trust among everyone involved, and review its decisions. This is particularly important for critical infrastructure, as one unintentional problem could interrupt operations or cause people to lose trust.

Going forward, a reliable nationwide deployment should use a step-by-step method that smoothly connects the idea of the system with its actual application. In the beginning, researchers can use small trials in carefully controlled settings, aided by major government labs and utility companies, to make sure the models hold true when used in the field. By carrying out these pilot projects, the aim is to learn about problems with attachment to vendor-specific systems, operating difficulties, and the ability of different national grids to connect smoothly. Achieving these trials will give people confidence and a basis for putting the program elsewhere.

The second step is to focus on scalability testing. The grid in the United States is broad and varied, covering various locations, weather types, and ways people use electricity. For this reason, the framework must be effective in multiple scenarios, including when latencies change, throughput varies, and hardware is not all the same. A stress test will prove that the framework will not fail under peak demand and rough conditions.

Developing governance structures is just as necessary. Any framework for grid cybersecurity should guarantee that AI is ethical, accountable, and easy to see how it operates. It is important that protocols are set for how models learn, results are evaluated regularly, decisions are verified, and faulty actions are updated. In addition, these frameworks ought to encourage data sharing among utility companies, regions, and the federal government, being careful to protect all data, so that all parties can gain from similar cyberattack warning signs.

Smart adversarial machine learning models are a subject that requires regular monitoring. Because AI is crucial in cybersecurity, its systems are also threatened by attacks. Presenting subtle changes to data used by machine learning can make these models lose their accuracy and this attack style is increasingly prevalent. The danger of fake cyber threats is especially high for smart grids, as a hidden mistake could result in a problem or lead to many false alarms that stop the grid from working smoothly. For this reason, future studies ought to strengthen AI models by using strategies such as adversarial training, defensive distillation, and validating models to safeguard AI-supported systems against these manipulations.

Getting the national grid fully prepared for cyber-attacks requires actions on technology and on policy, organizations, and society. For the energy sector to be suitable, responsive, and in sync, utilities, vendors, regulators, and professionals in cybersecurity must cooperate. To boost this change, we require workforce improvements, teamwork between companies and governments, and combining efforts with other nations.

Artificial intelligence in cybersecurity also contributes valuable opportunities for both better defense and industry-wide innovation in the energy sector. With the fundamental infrastructure safe, more people can benefit from advanced smart grid features, like demand-side management, adjustable pricing, and renewable energy integration. Creating a reliable grid, cutting down carbon emissions, and planning for the future matter because of these developments. In this instance, cybersecurity does more than keep energy systems safe; it contributes to reaching national energy goals.

Overall, this research supports the need to use AI within the security systems of the smart grid. Because they quickly detect risks, respond promptly, and continue to adapt, AI-based systems are a leading approach to handling complex safety issues in today’s energy industry. Using cloud-edge integration, compliance with regulations, and transparent use of AI can guide the creation of secure and smart energy infrastructure. Moving forward, how much the nation invests in testing, scaling, and governing smart grid systems will decide how securely the grid is defended and how much is gained from its modernization.

Significant risks are involved. Because national and economic stability, as well as millions of people’s welfare, depend on energy, now is when cybersecurity matters the most. With AI, not only does the technology change but there’s also a shift in how the grid deals with risks—now being aware, adaptive, and strong. More developments and uses of these technologies will play an important role in keeping the U.S. electric grid secure, dependable, and updated.

Compliance with ethical standards

Disclosure of conflict of interest

The authors declare no conflict of interest.

Data Availability

Available upon request.

Author Contributions

All authors contributed equally to this work.

Reference

[1] Palensky, P., & Kupzog, F. (2013). Smart grids. Annual Review of Environment and Resources, 38(1), 201-226.

[2] Moreno Escobar, J. J., Morales Matamoros, O., Tejeida Padilla, R., Lina Reyes, I., & Quintana Espinosa, H. (2021). A comprehensive review on smart grids: Challenges and opportunities. Sensors, 21(21), 6978.

[3] Yu, X., Cecati, C., Dillon, T., & Simoes, M. G. (2011). The new frontier of smart grids. IEEE Industrial Electronics Magazine, 5(3), 49-63.

[4] Tuballa, M. L., & Abundo, M. L. (2016). A review of the development of Smart Grid technologies. Renewable and Sustainable Energy Reviews, 59, 710-725.

[5] Siano, P. (2014). Demand response and smart grids—A survey. Renewable and sustainable energy reviews, 30, 461-478.

[6] Bayindir, R., Colak, I., Fulli, G., & Demirtas, K. (2016). Smart grid technologies and applications. Renewable and sustainable energy reviews, 66, 499-516.

[7] Craigen, D., Diakun-Thibault, N., & Purse, R. (2014). Defining cybersecurity. Technology innovation management review, 4(10).

[8] Kemmerer, R. A. (2003, May). Cybersecurity. In 25th International Conference on Software (Engineering et al., 2003). Proceedings. (pp. 705-715). IEEE.

[9] Sarker, I. H., Kayes, A. S. M., Badsha, S., Alqahtani, H., Watters, P., & Ng, A. (2020). Cybersecurity data science: an overview from machine learning perspective. Journal of Big data, 7, 1-29.

[10] AlDaajeh, S., Saleous, H., Alrabaee, S., Barka, E., Breitinger, F., & Choo, K. K. R. (2022). The role of national cybersecurity strategies on the improvement of cybersecurity education. Computers & Security, 119, 102754.

[11] Florackis, C., Louca, C., Michaely, R., & Weber, M. (2023). Cybersecurity risk. The Review of Financial Studies, 36(1), 351-407.

[12] Zhai, X., Chu, X., Chai, C. S., Jong, M. S. Y., Istenic, A., Spector, M., ... & Li, Y. (2021). A Review of Artificial Intelligence (AI) in Education from 2010 to 2020. (Zhai et al., 2021), 8812542.

[13] Makridakis, S. (2017). The forthcoming Artificial Intelligence (AI) revolution: Its impact on society and firms. Futures, 90, 46-60.

[14] Shaheen, M. Y. (2021). Applications of Artificial Intelligence (AI) in healthcare: A review. ScienceOpen Preprints. 10.14293/S2199-1006.1.SOR-.PPVRY8K.v1

[15] Lele, A. (2018). Artificial intelligence (AI). In Disruptive technologies for the militaries and security (pp. 139-154). Singapore: Springer Singapore.

[16] Vaishya, R., Javaid, M., Khan, I. H., & Haleem, A. (2020). Artificial Intelligence (AI) applications for COVID-19 pandemic. Diabetes & Metabolic Syndrome: Clinical Research & Reviews, 14(4), 337-339.

[17] Hassani, H., Silva, E. S., Unger, S., TajMazinani, M., & Mac Feely, S. (2020). Artificial intelligence (AI) or intelligence augmentation (IA): what is the future? Ai, 1(2), 8.

[18] Rahmani, A. M., Yousefpoor, E., Yousefpoor, M. S., Mehmood, Z., Haider, A., Hosseinzadeh, M., & Ali Naqvi, R. (2021). Machine learning (ML) in medicine: review, applications, and challenges. Mathematics, 9(22), 2970.

[19] Bell, J. (2022). What is machine learning?. Machine learning and the city: applications in architecture and urban design, 207-216.

[20] Carleo, G., Cirac, I., Cranmer, K., Daudet, L., Schuld, M., Tishby, N., ... & Zdeborová, L. (2019). Machine learning and the physical sciences. Reviews of Modern Physics, 91(4), 045002.

[21] Masson, J. F., Biggins, J. S., & Ringe, E. (2023). Machine learning for nanoplasmonics. Nature Nanotechnology, 18(2), 111-123.

[22] Bzdok, D., Krzywinski, M., & Altman, N. (2017). Machine learning: a primer. Nature methods, 14(12), 1119.

[23] Wagstaff, K. (2012). Machine learning that matters. arXiv preprint arXiv:1206.4656.

[24] Gaiceanu, M., Stanculescu, M., Andrei, P. C., Solcanu, V., Gaiceanu, T., & Andrei, H. (2020). Intrusion detection on ics and scada networks. Recent Developments on Industrial Control Systems Resilience, 197-262.

[25] Awad, R. A., Beztchi, S., Smith, J. M., Lyles, B., & Prowell, S. (2018, December). Tools, techniques, and methodologies: A survey of digital forensics for scada systems. In Proceedings of the 4th Annual Industrial Control System Security Workshop (pp. 1-8).