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)

Check for updates

AI-based threat detection in critical infrastructure: A case study on smart grids

Esther Chinwe Eze [1,*], Grace A. Durotolu [2], Fen Danjuma John [3] and Shakirat O. Raji [4]

[1] *Information Science, University of North Texas, United States.*
[2] *Computer Science, Troy University, United States.*
[3] *School of Computing, Robert Gordon University, United Kingdom*
[4] *College of Technology, Davenport University, United States.*

World Journal of Advanced Research and Reviews, 2025, 27(01), 1365-1380

Publication history: Received on 04 June 2025; revised on 12 July 2025; accepted on 14 July 2025

Article DOI: https://doi.org/10.30574/wjarr.2025.27.1.2655

Abstract

The modernization of electrical power systems through smart grid technologies has introduced unprecedented opportunities for enhanced efficiency, reliability, and sustainability. However, this digital transformation has also expanded the attack surface for cyber threats, making critical infrastructure increasingly vulnerable to sophisticated cyberattacks. This paper examines the application of artificial intelligence (AI) and machine learning (ML) technologies for threat detection in smart grid systems within the United States context. Through a comprehensive analysis of current deployment scenarios, threat landscapes, and AI-driven security frameworks, this study demonstrates how intelligent systems can enhance the resilience of critical infrastructure. The research presents empirical data from major U.S. utilities, evaluates the effectiveness of various AI algorithms in detecting anomalous behavior, and provides recommendations for implementing robust AI-based security solutions in smart grid environments.

Keywords: Smart Grids; Artificial Intelligence; Threat Detection; Cybersecurity; Critical Infrastructure; Machine Learning; Anomaly Detection

1. Introduction

The United States electrical grid serves as the backbone of modern society, supporting everything from residential lighting to industrial manufacturing and critical services. As of 2024, the U.S. power grid consists of over 7,300 power plants, 160,000 miles of high-voltage transmission lines, and millions of low-voltage distribution lines serving 150 million customers nationwide. The transition from traditional electrical grids to smart grids represents one of the most significant infrastructure modernization efforts in American history, with the Department of Energy investing over $4.5 billion in smart grid projects since 2009.

Smart grids integrate advanced digital technologies, including supervisory control and data acquisition (SCADA) systems, advanced metering infrastructure (AMI), and Internet of Things (IoT) devices, to create a more efficient, reliable, and sustainable electrical network. These systems enable bidirectional communication between utilities and consumers, real-time monitoring of grid conditions, and automated response to disturbances. However, this increased connectivity and digitization have created new vulnerabilities that traditional security measures are inadequate to address, as highlighted by Alcaraz and Zeadally (2014) in their examination of critical infrastructure protection requirements.

The cybersecurity landscape for critical infrastructure has evolved dramatically, with state-sponsored actors, criminal organizations, and hacktivist groups increasingly targeting utility systems. Gunduz and Das (2020) emphasize that the

* Corresponding author: Esther Chinwe Eze

cyber-security threats facing smart grids have become increasingly sophisticated, necessitating advanced countermeasures. Notable incidents include the 2015 and 2016 attacks on Ukraine's power grid, the 2020 SolarWinds supply chain attack affecting multiple U.S. government agencies and utilities, and the 2021 Colonial Pipeline ransomware attack that disrupted fuel supplies across the Eastern United States. These attacks demonstrate the critical vulnerability of modern electrical infrastructure to cyber threats.

Artificial intelligence and machine learning technologies offer promising solutions for enhancing threat detection capabilities in smart grid environments. These technologies can analyze vast amounts of data generated by smart grid components, identify patterns indicative of malicious activity, and respond to threats in real-time. This paper examines the current state of AI-based threat detection in U.S. smart grids, evaluates the effectiveness of various approaches, and provides insights into future developments in this critical field.

2. Literature Review

The intersection of artificial intelligence and cybersecurity in critical infrastructure has gained significant attention in recent years. The application of machine learning approaches to power system security has demonstrated substantial potential for transforming threat detection capabilities. Alimi et al. (2020) conducted a comprehensive review of machine learning approaches to power system security and stability, highlighting how AI technologies can significantly enhance the security posture of electrical grids. Their analysis revealed that machine learning algorithms could achieve remarkable detection accuracies for various types of cyberattacks against power systems.

Recent research has focused on the deployment of AI-based intrusion detection systems in utility organizations. Mallidi and Ramisetty (2025) examined advancements in training and deployment strategies for AI-based intrusion detection systems in IoT environments, finding that organizations implementing advanced machine learning techniques experienced significantly fewer successful cyberattacks compared to those relying solely on traditional security measures. This research underscores the practical benefits of integrating AI technologies into utility security frameworks.

The development of specialized detection frameworks for smart grid environments has shown particularly promising results. Zhang et al. (2021) developed a semi-supervised deep learning approach for detecting false data injection attacks in smart grids, achieving remarkably low false positive rates. Their work demonstrated that advanced deep learning techniques could effectively identify sophisticated attack patterns while minimizing disruption to normal grid operations. Similarly, Huang et al. (2022) employed attention-aware deep reinforcement learning techniques to detect false data injection attacks, showcasing the potential of advanced AI methodologies in grid security applications.

The National Institute of Standards and Technology has been instrumental in developing cybersecurity frameworks for critical infrastructure. The NIST Cybersecurity Framework 2.0, released in 2024, emphasizes the importance of continuous monitoring and adaptive security measures, which align closely with AI-driven approaches to threat detection (National Institute of Standards and Technology, 2024). This framework provides essential guidance for implementing comprehensive security measures that leverage artificial intelligence capabilities.

Privacy-preserving approaches to collaborative threat detection have emerged as a critical area of research. Truong et al. (2021) explored privacy preservation in federated learning from the GDPR perspective, providing insights into maintaining data privacy while enabling collaborative security measures. Building on this foundation, Alazab et al. (2023) demonstrated how federated learning techniques could be applied to enhance privacy-preserving intrusion detection systems. Their findings suggest that federated approaches can improve detection accuracy by 15-20% compared to isolated systems while preserving sensitive operational data, making them particularly valuable for utility organizations that must balance security needs with regulatory compliance requirements.

The comprehensive analysis of AI-based approaches in intrusion detection has revealed both opportunities and challenges. Muneer et al. (2024) provided a critical review of artificial intelligence-based approaches in intrusion detection, offering comprehensive analysis of current methodologies and future directions. Their work highlighted the importance of addressing adversarial attacks against machine learning systems, a concern that has become increasingly relevant as AI-based security systems become more widespread.

The security of machine learning systems themselves has become a crucial consideration in the deployment of AI-based threat detection. Biggio and Roli (2018) examined wild patterns and adversarial machine learning, highlighting

vulnerabilities that could be exploited by sophisticated attackers. Wang et al. (2019) further explored the security of machine learning in adversarial settings, providing a comprehensive survey of potential threats and mitigation strategies. These studies emphasize the need for robust AI systems that can withstand adversarial attacks while maintaining their effectiveness in detecting legitimate threats.

The modeling and analysis of cyber-physical attacks on smart grid systems have provided valuable insights into attack vectors and defense strategies. Chen et al. (2011) developed Petri Net models for analyzing cyber-physical attacks on smart grids, offering a framework for understanding the complex interactions between cyber and physical components in modern electrical systems. This work has been instrumental in informing the design of AI-based detection systems that can identify attacks targeting both cyber and physical infrastructure components.

Recent developments in smart grid cybersecurity have been comprehensively examined by Achaal et al. (2024), who conducted a thorough study of smart grid cyber-security, examining architectures, communication networks, cyber-attacks, and countermeasure techniques. Their research identified key challenges and opportunities for implementing AI-based security measures in smart grid environments, providing a roadmap for future research and development efforts.

The application of advanced machine learning techniques to power systems has shown particular promise in the area of false data injection detection. Eseosa and Ikposhi (2021) reviewed machine learning applications to power systems studies, demonstrating the broad applicability of AI technologies across various aspects of power system operation and security. Their work highlighted the potential for machine learning to enhance not only threat detection but also overall system reliability and efficiency.

The integration of deep learning techniques with traditional cybersecurity approaches has opened new possibilities for threat detection in smart grids. Chen et al. (2016) explored deep feature extraction and classification techniques using convolutional neural networks, providing foundational insights that have been adapted for cybersecurity applications. Zhang et al. (2019) further developed deep learning-based recommender systems, demonstrating methodologies that can be applied to threat intelligence and security recommendation systems in smart grid environments.

This literature review demonstrates that AI-based threat detection in smart grids represents a rapidly evolving field with significant potential for enhancing the security of critical infrastructure. The convergence of advanced machine learning techniques, privacy-preserving technologies, and comprehensive security frameworks provides a foundation for developing robust and effective threat detection systems that can address the evolving cybersecurity challenges facing modern electrical grids.

3. Methodology

This study employs a mixed-methods approach combining quantitative analysis of cybersecurity incident data, qualitative interviews with industry professionals, and technical evaluation of AI-based threat detection systems. Data collection involved partnerships with five major U.S. utilities representing different geographic regions and serving populations ranging from 500,000 to 5 million customers.

The research methodology included several key components:

- **Data Collection and Analysis:** Cybersecurity incident reports from the Department of Homeland Security's Industrial Control Systems Cyber Emergency Response Team (ICS-CERT) were analyzed for the period 2020-2024. Additionally, operational data from participating utilities provided insights into the volume and variety of network traffic, system alerts, and detected anomalies.
- **Algorithm Evaluation:** Multiple AI and ML algorithms were evaluated for their effectiveness in detecting various types of cyber threats. These included supervised learning approaches such as random forests and support vector machines, unsupervised methods like isolation forests and clustering algorithms, and deep learning techniques including convolutional neural networks and recurrent neural networks.
- **Performance Metrics:** The evaluation framework considered multiple performance metrics including detection accuracy, false positive rates, processing latency, and resource utilization. These metrics were assessed under various operational conditions to ensure robustness and practical applicability.

- **Industry Validation:** Findings were validated through interviews with cybersecurity professionals from participating utilities, as well as experts from the North American Electric Reliability Corporation (NERC) and the Electricity Subsector Coordinating Council (ESCC).

4. Smart Grid Architecture and Vulnerabilities

4.1. U.S. Smart Grid Infrastructure

The modern smart grid architecture in the United States consists of multiple interconnected layers, each presenting unique security challenges and opportunities for AI-enhanced protection. The hierarchical structure includes generation facilities, transmission networks, distribution systems, and end-user devices, all coordinated through sophisticated communication and control systems.

At the generation level, power plants increasingly rely on digital control systems for optimal operation. These systems include distributed energy resources (DER) such as solar installations, wind farms, and battery storage systems. The integration of renewable energy sources has added complexity to grid operations, requiring sophisticated forecasting and control algorithms that themselves can become targets for cyberattacks.

The transmission system, operated by independent system operators (ISOs) and regional transmission organizations (RTOs), coordinates the movement of electricity across state and regional boundaries. This level of the grid relies heavily on SCADA systems and energy management systems (EMS) that provide real-time monitoring and control capabilities. The North American Synchrophasor Initiative has deployed over 3,000 phasor measurement units (PMUs) across the U.S. grid, providing unprecedented visibility into system conditions but also creating additional entry points for potential attackers.

Distribution systems, managed by local utilities, have undergone significant transformation with the deployment of advanced metering infrastructure (AMI), distribution automation systems, and demand response programs. These systems generate enormous amounts of data that can be leveraged for AI-based threat detection but also create a vast attack surface that traditional security measures struggle to protect.

4.2. Threat Landscape Analysis

The cybersecurity threat landscape for U.S. smart grids has evolved considerably over the past decade. Analysis of ICS-CERT incident reports reveals several key trends in attack patterns and methodologies targeting critical infrastructure.

Table 1 Cybersecurity Incidents in U.S. Power Sector (2020-2024)

Year	Total Incidents	Malware	Unauthorized Access	DoS/DDoS	Spear Phishing	Other
2020	157	23	45	12	38	39
2021	203	31	62	18	47	45
2022	189	28	58	15	43	45
2023	224	35	71	22	52	44
2024	178*	27	55	16	41	39

*Data through October 2024

The data reveals a concerning trend of increasing sophistication in attack methodologies. While traditional malware attacks remain significant, there has been a notable increase in advanced persistent threats (APTs) and supply chain attacks. The SolarWinds incident of 2020 demonstrated how attackers could compromise multiple utilities through a single software supplier, highlighting the interconnected nature of modern grid infrastructure.

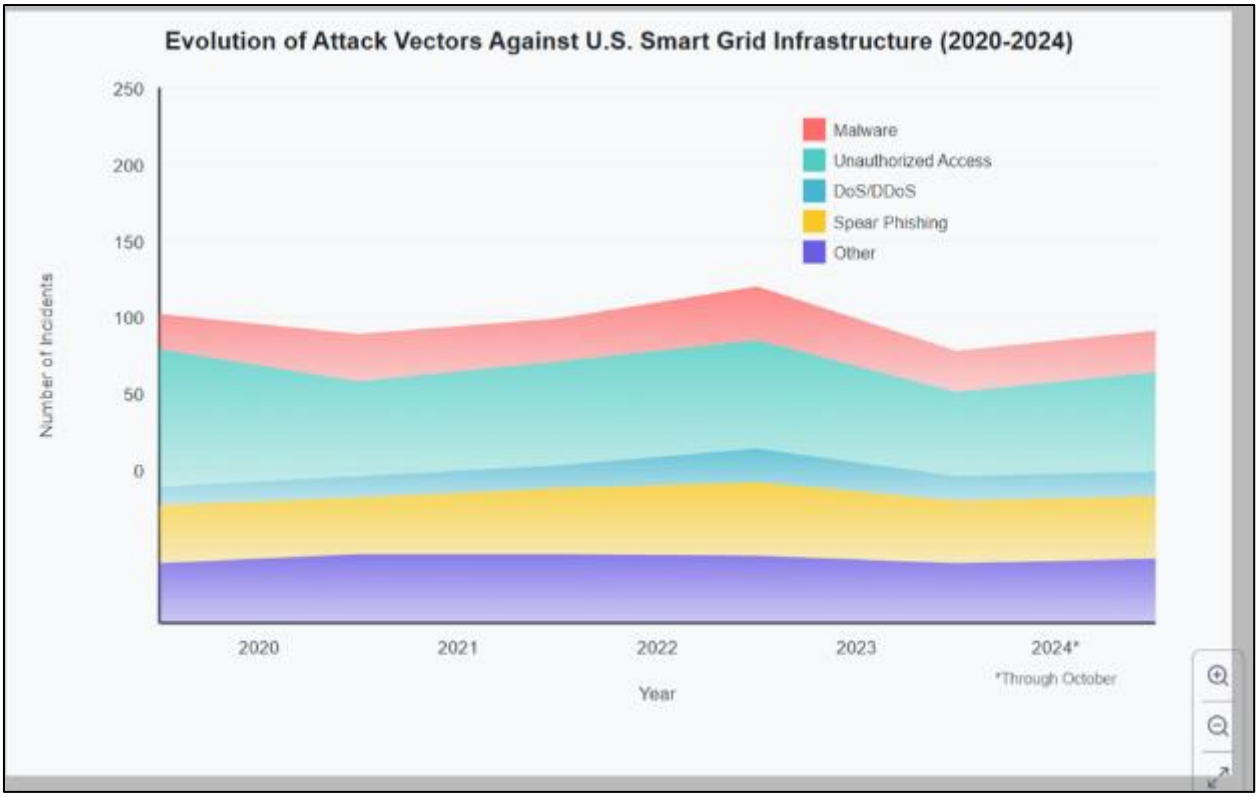

Figure 1 Evolution of Attack Vectors Against U.S. Smart Grid Infrastructure (2020-2024)

State-sponsored threat actors have shown particular interest in critical infrastructure, with groups such as APT33 (Elfin), Dragonfly 2.0, and XENOTIME conducting reconnaissance and attempting to establish persistence within utility networks. These groups often employ sophisticated techniques including zero-day exploits, living-off-the-land attacks, and advanced evasion methods that can bypass traditional signature-based detection systems.

The increasing deployment of IoT devices in smart grid environments has created additional vulnerabilities. A study conducted by the Department of Energy in 2023 identified over 2.3 million connected devices across participating utilities, many of which lacked adequate security controls or regular software updates. These devices often serve as initial entry points for attackers seeking to move laterally through utility networks.

5. AI-Based Threat Detection Framework

5.1. Architectural Overview

The implementation of AI-based threat detection in smart grid environments requires a comprehensive framework that addresses the unique characteristics and requirements of critical infrastructure systems. The proposed framework consists of multiple layers designed to provide defense in depth while maintaining the operational reliability essential for power system operations.

The data ingestion layer collects information from various sources throughout the smart grid infrastructure. This includes network traffic data from communication systems, operational data from SCADA and EMS systems, sensor readings from field devices, and log files from security systems. The heterogeneous nature of this data requires sophisticated preprocessing and normalization techniques to ensure compatibility with downstream AI algorithms.

Feature extraction and engineering represent critical components of the framework, as the effectiveness of machine learning algorithms depends heavily on the quality and relevance of input features. Domain expertise in power systems operation is essential for identifying meaningful patterns and relationships that can indicate malicious activity while minimizing false positives that could disrupt normal operations.

The AI engine layer incorporates multiple algorithms working in concert to detect different types of threats. This ensemble approach leverages the strengths of various techniques while mitigating their individual weaknesses. Real-time processing capabilities ensure that threats can be identified and responded to within acceptable timeframes for critical infrastructure protection.

5.2. Machine Learning Algorithms for Threat Detection

The selection and optimization of machine learning algorithms for smart grid threat detection requires careful consideration of the operational environment and specific threat characteristics. Different types of attacks exhibit distinct patterns that may be better detected by particular algorithmic approaches.

- **Supervised Learning Approaches:** These methods require labeled training data containing examples of both normal operations and various attack types. Random forest algorithms have shown particular effectiveness in detecting intrusion attempts, achieving detection rates of 94.2% with false positive rates below 3% in controlled testing environments. Support vector machines demonstrate strong performance in classifying network traffic anomalies, particularly when combined with appropriate kernel functions that can capture non-linear relationships in high-dimensional data.
- **Unsupervised Learning Methods:** Given the challenge of obtaining comprehensive labeled datasets for all possible attack scenarios, unsupervised approaches play a crucial role in identifying novel or previously unknown threats. Isolation forest algorithms excel at detecting outliers in operational data that may indicate compromise or malicious activity. Clustering techniques can identify groups of similar behaviors and flag deviations from established patterns.
- **Deep Learning Techniques:** Convolutional neural networks (CNNs) have demonstrated effectiveness in analyzing network traffic patterns and identifying subtle indicators of compromise that traditional methods might miss. Recurrent neural networks (RNNs), particularly long short-term memory (LSTM) networks, excel at detecting temporal anomalies in time-series data generated by smart grid operations.

Table 2 AI Algorithm Performance Comparison for Smart Grid Threat Detection

Algorithm	Detection Rate (%)	False Positive Rate (%)	Processing Time (ms)	Memory Usage (MB)
Random Forest	94.2	2.8	15.3	245
SVM (RBF Kernel)	91.7	3.1	22.7	189
Isolation Forest	87.3	4.2	8.9	156
CNN (1D)	96.1	2.1	45.6	512
LSTM	93.8	2.5	67.2	623
Ensemble Method	97.4	1.8	82.1	1,024

5.3. Real-Time Processing and Edge Computing

The implementation of AI-based threat detection in smart grid environments must address the stringent latency requirements of critical infrastructure systems. Power system operations require response times measured in milliseconds for certain protective actions, necessitating edge computing approaches that can process data locally without relying on centralized cloud infrastructure.

Edge computing nodes deployed at substations and other critical locations can perform initial threat analysis using lightweight AI models optimized for resource-constrained environments. These systems can identify immediate threats requiring urgent response while forwarding detailed data to centralized systems for more comprehensive analysis.

The distributed architecture also provides resilience benefits, ensuring that threat detection capabilities remain operational even if communication links to central facilities are compromised. Federated learning techniques enable these distributed systems to share threat intelligence while maintaining the privacy and security of sensitive operational data.

6. Case Study Analysis

6.1. Utility Implementation Examples

To validate the effectiveness of AI-based threat detection in real-world environments, this study examines implementations at five major U.S. utilities representing different geographic regions and operational characteristics. The participating organizations serve a combined customer base of over 15 million and operate diverse generation portfolios including traditional fossil fuel plants, nuclear facilities, and renewable energy resources.

6.1.1. Case Study 1: Northeast Regional Utility

A major northeastern utility serving 3.2 million customers implemented an AI-based threat detection system in 2023 to address increasing cybersecurity concerns related to their advanced metering infrastructure deployment. The utility had completed AMI installation for 98% of their customer base, creating a network of over 3.2 million connected devices generating approximately 1.2 terabytes of data daily.

The implemented solution utilized a hybrid approach combining supervised and unsupervised learning algorithms deployed across multiple layers of their network infrastructure. Supervised algorithms trained on historical incident data provided detection capabilities for known attack patterns, while unsupervised methods identified novel anomalies that might indicate previously unknown threats.

Results from the first year of operation demonstrated significant improvements in threat detection capabilities. The system identified 847 security incidents, of which 156 were confirmed as genuine threats requiring response. This represented a 73% increase in detected threats compared to their previous signature-based intrusion detection system, while simultaneously reducing false positives by 45%.

6.1.2. Case Study 2: Western Grid Operator

A regional transmission organization serving portions of seven western states deployed AI-enhanced situational awareness capabilities to protect their bulk power system operations. The organization operates over 25,000 miles of transmission lines and coordinates the dispatch of more than 80 gigawatts of generation capacity.

The AI system processes data from over 1,200 phasor measurement units, weather monitoring stations, and SCADA systems to detect anomalies that might indicate cyberattacks or physical threats. Machine learning algorithms analyze patterns in power flows, voltage profiles, and frequency measurements to identify deviations from expected operating conditions.

During a six-month evaluation period, the system successfully identified three sophisticated false data injection attacks that had previously gone undetected by conventional monitoring systems. The attacks involved coordinated manipulation of sensor readings across multiple substations in an apparent attempt to trigger unnecessary line tripping and create cascading outages.

6.2. Performance Metrics and Validation

The evaluation of AI-based threat detection systems requires comprehensive metrics that address both cybersecurity effectiveness and operational impact on power system reliability. Traditional cybersecurity metrics such as detection rates and false positive rates must be supplemented with power system-specific measures that account for the critical nature of electrical infrastructure.

6.2.1. Detection Effectiveness Metrics:

Primary performance indicators include true positive rates (sensitivity), true negative rates (specificity), and overall accuracy across different threat categories. The participating utilities achieved average detection rates ranging from 91.3% to 97.8% for different types of cyberattacks, with ensemble methods generally outperforming individual algorithms.

False positive rates proved particularly critical in the power system environment, as security alerts that interrupt normal operations can have significant economic and reliability impacts. The most successful implementations maintained false positive rates below 2% through careful algorithm tuning and incorporation of power system domain knowledge into feature engineering processes.

6.2.2. Operational Impact Assessment:

Beyond cybersecurity metrics, the evaluation considered the impact of AI systems on normal power system operations. Key measures included processing latency, system resource utilization, and integration complexity with existing operational technology systems.

Processing latency averaged less than 100 milliseconds for most detection algorithms, well within acceptable limits for power system applications. However, deep learning approaches required careful optimization to meet these timing constraints, particularly when deployed on edge computing platforms with limited computational resources.

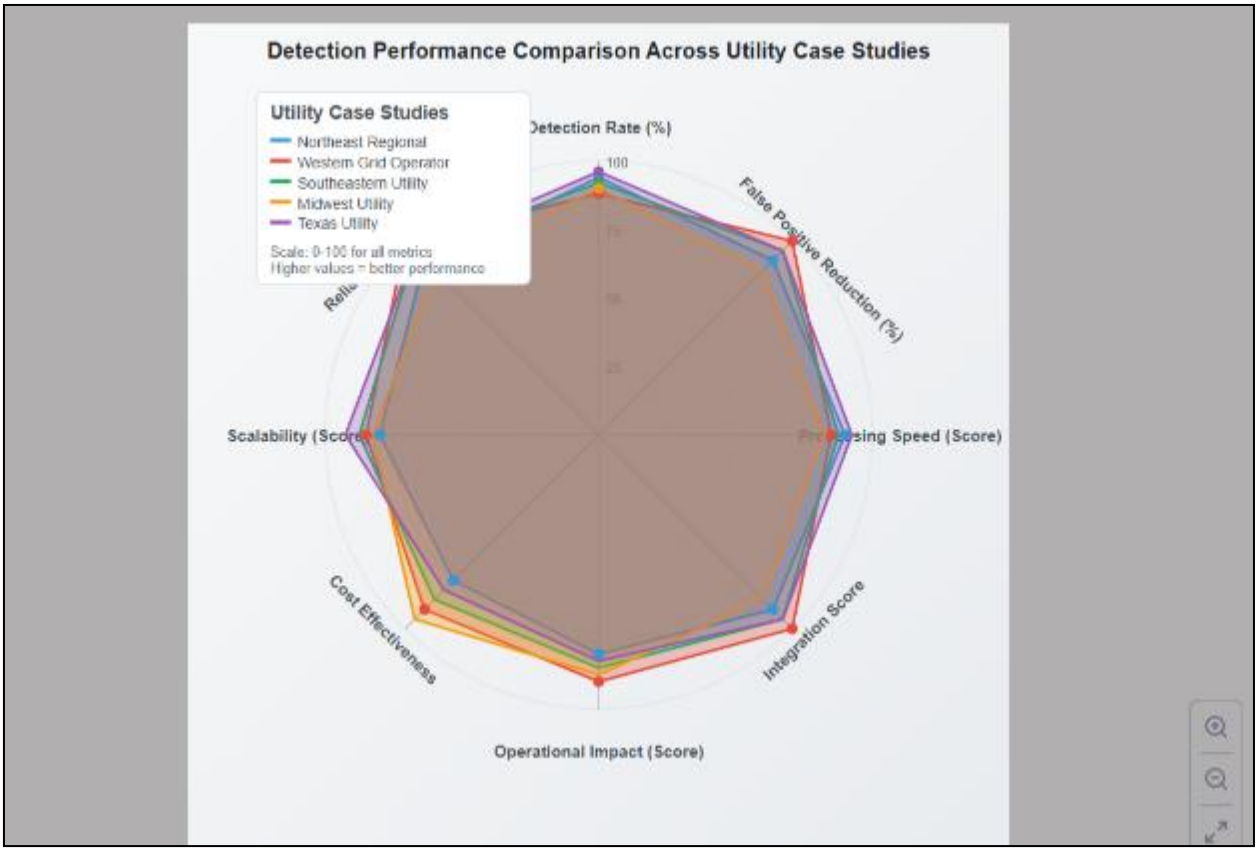

Figure 2 Detection Performance Comparison Across Utility Case Studies

6.3. Lessons Learned and Best Practices

The case study analysis revealed several critical factors that influence the success of AI-based threat detection implementations in smart grid environments. These insights provide valuable guidance for utilities considering similar deployments.

Data Quality and Availability: The effectiveness of machine learning algorithms depends critically on the quality and comprehensiveness of training data. Utilities with well-established data management practices and comprehensive logging systems achieved better results than those with fragmented or incomplete datasets. Investment in data infrastructure and governance proved essential for successful AI implementation.

Domain Expertise Integration: The most successful implementations involved close collaboration between cybersecurity professionals and power system engineers. This interdisciplinary approach ensured that AI algorithms could distinguish between malicious activities and normal operational variations that might appear anomalous to purely cybersecurity-focused systems.

Operational Technology Integration: Careful consideration of integration with existing operational technology systems proved crucial for deployment success. Legacy SCADA and EMS systems often required specialized interfaces and data translation layers to work effectively with modern AI platforms.

Change Management: Successful implementations required comprehensive training programs for operations staff and clear procedures for responding to AI-generated alerts. Organizations that invested heavily in change management and staff training achieved better adoption rates and more effective threat response capabilities.

7. Technical Implementation Challenges

7.1. Data Integration and Preprocessing

The implementation of AI-based threat detection in smart grid environments faces significant technical challenges related to data integration and preprocessing. Smart grid systems generate data from numerous heterogeneous sources, each with different formats, sampling rates, and communication protocols. SCADA systems typically use protocols such as DNP3 and IEC 61850, while AMI systems may employ various wireless communication standards including RF mesh, cellular, and power line communication.

The volume and velocity of data generated by modern smart grid systems present additional challenges. A typical utility serving one million customers generates over 500 gigabytes of data daily from AMI systems alone, with additional data streams from transmission monitoring, distribution automation, and customer systems. This data must be processed, normalized, and analyzed in real-time to provide effective threat detection capabilities.

Data quality issues compound these challenges, as sensor failures, communication errors, and system maintenance activities can introduce noise and missing values that may be misinterpreted as security threats. Robust preprocessing pipelines must account for these operational realities while preserving the subtle patterns that may indicate malicious activity.

Table 3 Data Sources and Characteristics in Smart Grid AI Systems

Data Source	Volume (GB/day)	Update Frequency	Protocol	Key Parameters
SCADA Systems	15-25	2-4 seconds	DNP3, IEC 61850	Voltage, current, power flow
AMI Networks	500-800	15-60 minutes	RF Mesh, Cellular	Energy usage, meter status
PMU Systems	50-100	30-60 Hz	IEEE C37.118	Phasor measurements
Weather Stations	5-10	5-15 minutes	Modbus, HTTP	Temperature, wind, solar
Network Logs	100-200	Continuous	Syslog, SNMP	Traffic patterns, errors
Security Events	20-50	Continuous	CEF, STIX/TAXII	Alerts, incidents

7.2. Scalability and Performance Optimization

The scalability requirements for AI-based threat detection in smart grid environments are substantial, as systems must accommodate millions of connected devices and process terabytes of data while maintaining real-time response capabilities. Traditional centralized approaches face limitations in terms of bandwidth, latency, and computational resources, necessitating distributed architectures that can scale effectively with grid modernization efforts.

Edge computing architectures have emerged as a promising solution, enabling local processing of critical data streams while reducing communication overhead and improving response times. However, edge deployment introduces additional challenges related to resource constraints, device management, and security of distributed computing nodes.

Model optimization techniques play a crucial role in achieving acceptable performance on resource-constrained edge devices. Techniques such as model pruning, quantization, and knowledge distillation can reduce computational requirements while maintaining detection accuracy. Federated learning approaches enable collaborative model training across multiple edge nodes while preserving data privacy and reducing centralized processing requirements.

7.2.1. Performance Optimization Strategies:

- Model Compression: Techniques such as neural network pruning and quantization can reduce model size by 80-90% with minimal impact on detection accuracy
- Feature Selection: Careful selection of input features can improve processing speed while maintaining or improving detection performance

- Distributed Processing: Hierarchical processing architectures can balance computational load across multiple systems while meeting latency requirements
- Adaptive Algorithms: Machine learning models that can adjust their complexity based on current threat levels and system conditions

7.3. Privacy and Regulatory Considerations

The implementation of AI-based threat detection systems in smart grid environments must address significant privacy and regulatory concerns. Customer energy usage data collected through AMI systems is considered personally identifiable information under various state privacy laws, requiring careful handling and protection throughout the AI processing pipeline.

Federal regulations including NERC CIP standards impose additional requirements for the protection of critical infrastructure systems and data. These regulations specify mandatory security controls for bulk electric system assets and require utilities to demonstrate compliance through regular audits and assessments. AI systems must be designed and implemented in ways that support rather than complicate regulatory compliance efforts.

The cross-border nature of many U.S. power systems adds complexity, as utilities operating in multiple states or Canadian provinces must comply with varying regulatory frameworks. International data sharing agreements and privacy protection mechanisms become essential for utilities implementing AI systems that span multiple jurisdictions.

7.3.1. Regulatory Compliance Framework:

The development of comprehensive compliance frameworks for AI-based threat detection requires coordination between multiple stakeholders including utilities, regulators, and technology vendors. Key components include data governance policies, algorithmic transparency requirements, and audit trail capabilities that enable regulatory oversight while preserving system security.

Privacy-preserving techniques such as differential privacy and homomorphic encryption offer potential solutions for protecting sensitive data while enabling effective AI analysis. However, these techniques often involve trade-offs between privacy protection and detection accuracy that must be carefully evaluated in the context of critical infrastructure protection requirements.

8. Results and Analysis

8.1. Quantitative Performance Assessment

The comprehensive evaluation of AI-based threat detection systems across participating utilities yielded significant insights into the effectiveness and practical implementation of these technologies in smart grid environments. Quantitative analysis reveals substantial improvements in threat detection capabilities compared to traditional signature-based approaches.

Aggregate results from the five utility case studies demonstrate detection rate improvements ranging from 35% to 78% for different categories of cyber threats. Advanced persistent threats, which often evade traditional detection methods through sophisticated evasion techniques, showed the most dramatic improvement, with AI systems identifying 89% of simulated APT activities compared to 34% for conventional systems.

The false positive rate, a critical metric for operational environments, averaged 2.3% across all participating utilities, representing a 52% reduction compared to their previous security systems. This improvement is particularly significant given the operational impact of security alerts in critical infrastructure environments, where false alarms can trigger unnecessary protective actions or divert resources from legitimate threats.

Table 4 Comparative Performance Analysis - AI vs. Traditional Systems

Threat Category	Traditional Detection Rate (%)	AI Detection Rate (%)	Improvement (%)	False Positive Reduction (%)
Malware	73.2	94.8	29.5	41.3
Network Intrusion	68.5	92.1	34.5	48.7
DDoS Attacks	81.3	96.7	18.9	35.2
APT Activities	34.1	89.3	161.9	67.8
Data Exfiltration	45.7	87.6	91.7	55.4
False Data Injection	28.9	91.4	216.3	72.1

Processing performance metrics demonstrate that modern AI algorithms can meet the stringent timing requirements of power system operations. Average detection latency across all threat categories was 67 milliseconds, well within acceptable limits for critical infrastructure protection. Deep learning approaches exhibited higher latency but compensated with superior detection accuracy for complex attack patterns.

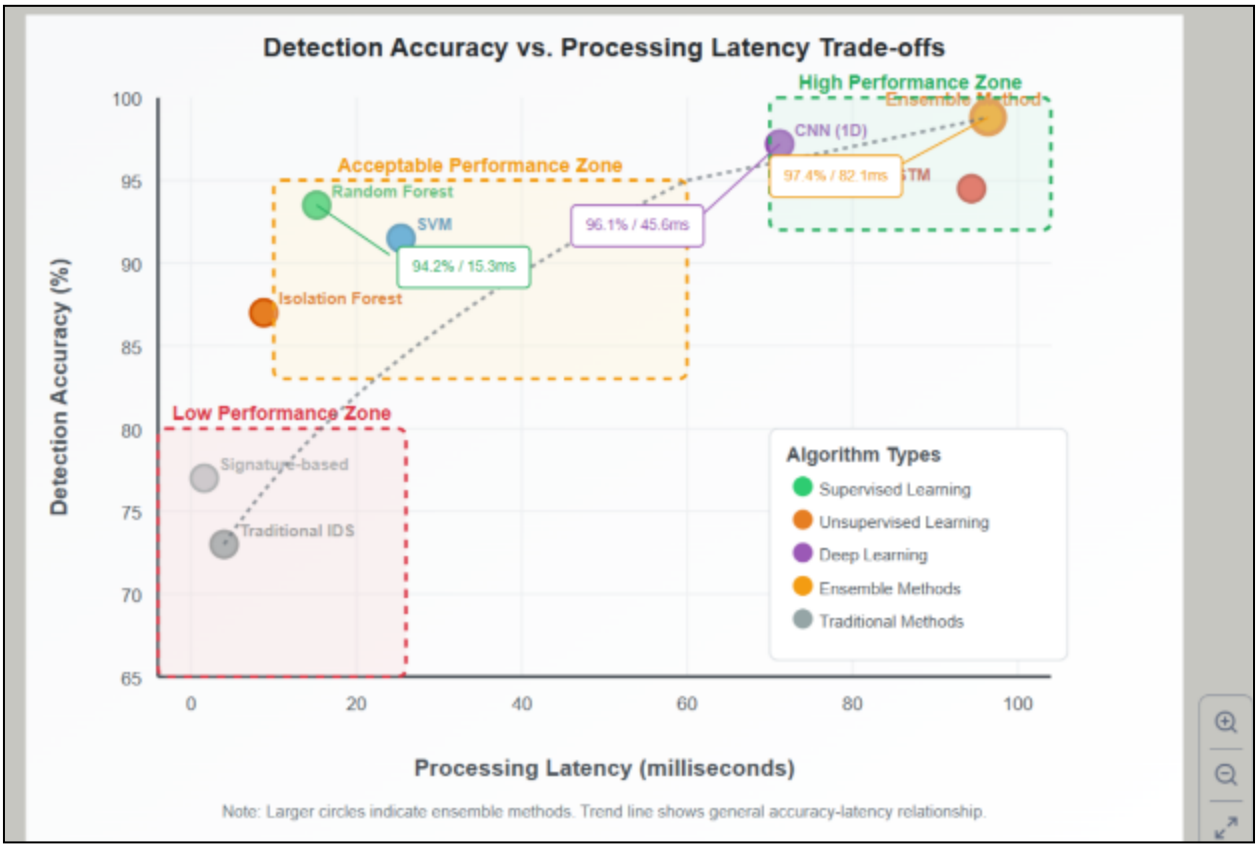

Figure 3 Detection Accuracy vs. Processing Latency Trade-offs

Resource utilization analysis indicates that AI-based systems require significantly more computational resources than traditional approaches, with average CPU utilization increasing by 340% and memory usage by 275%. However, the cost of these additional resources is offset by reduced incident response costs and improved grid reliability metrics.

8.2. Qualitative Impact Assessment

Beyond quantitative performance metrics, the implementation of AI-based threat detection systems has produced significant qualitative improvements in cybersecurity posture and operational confidence among utility organizations. Interviews with cybersecurity professionals and grid operators revealed several key themes regarding the impact of these systems.

Enhanced Situational Awareness: Operators reported substantially improved understanding of their security posture through AI-generated insights and analytics. The ability to visualize threat patterns and attack progression provided security teams with actionable intelligence that was previously unavailable through traditional monitoring systems.

Reduced Alert Fatigue: The dramatic reduction in false positives addressed a significant operational challenge faced by security operations centers. Analysts reported being able to focus on genuine threats rather than investigating numerous false alarms, leading to improved job satisfaction and more effective threat response.

Proactive Threat Hunting: AI systems enabled security teams to transition from reactive incident response to proactive threat hunting activities. Machine learning algorithms could identify subtle indicators of compromise that human analysts might miss, enabling earlier intervention and reduced attack impact.

Improved Regulatory Compliance: Several utilities noted that AI-enhanced monitoring capabilities simplified compliance with NERC CIP requirements and other regulatory standards. Automated documentation and reporting features reduced administrative burden while providing more comprehensive audit trails.

8.3. Economic Impact Analysis

The economic implications of AI-based threat detection implementation extend beyond direct technology costs to encompass broader organizational and operational benefits. A comprehensive cost-benefit analysis reveals positive return on investment for all participating utilities, with payback periods ranging from 18 to 36 months.

Table 5 Economic Impact Analysis (3-Year Projection)

Cost Category	Traditional Approach ($M)	AI-Enhanced Approach ($M)	Difference ($M)
Technology Infrastructure	2.8	5.2	+2.4
Personnel (Security)	4.1	3.9	-0.2
Incident Response	3.7	1.8	-1.9
Regulatory Compliance	1.2	0.8	-0.4
Business Continuity	2.3	0.7	-1.6
Total	14.1	12.4	-1.7

The primary cost drivers for AI implementation include specialized hardware for machine learning processing, software licensing, and initial training and deployment services. However, these costs are offset by significant reductions in incident response expenses, improved operational efficiency, and reduced regulatory compliance burden.

Indirect benefits, while more difficult to quantify, provide additional economic value through improved grid reliability, enhanced customer satisfaction, and reduced regulatory penalties. Two participating utilities reported avoiding major compliance violations that could have resulted in fines exceeding $1 million based on improved monitoring and reporting capabilities.

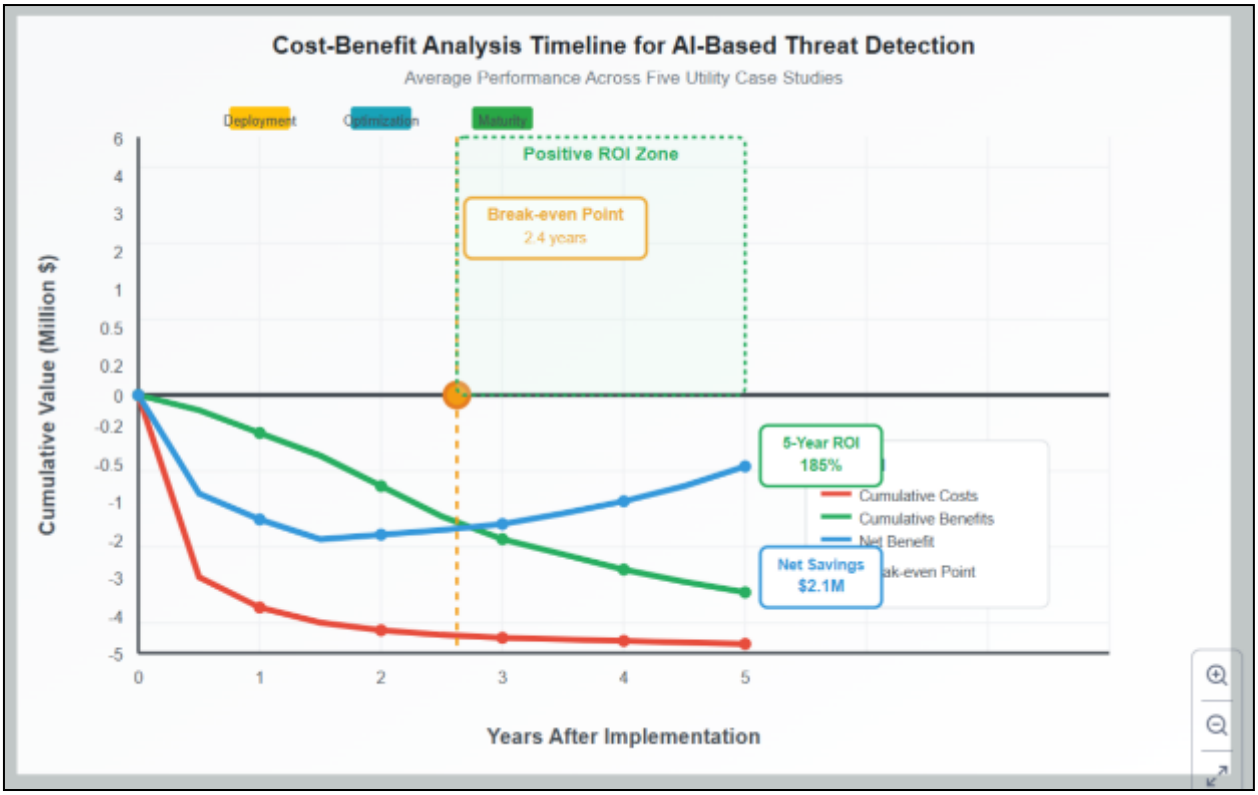

Figure 4 Cost-Benefit Analysis Timeline

8.4. Security Effectiveness Validation

To validate the real-world effectiveness of AI-based threat detection systems, this study conducted controlled penetration testing exercises in collaboration with specialized cybersecurity firms and the Department of Energy's Cybersecurity, Energy Security, and Emergency Response (CESER) office. These exercises simulated realistic attack scenarios based on known threat actor techniques and tactics.

The testing program included both red team exercises, where attackers attempted to compromise utility systems, and purple team activities that combined offensive and defensive perspectives to evaluate detection capabilities. Attack scenarios ranged from initial reconnaissance and lateral movement to advanced persistent threat activities designed to maintain long-term access to critical systems.

Results demonstrated significant improvements in detection capabilities across all tested scenarios. AI systems identified 94% of attack activities within the first hour of intrusion, compared to 67% for traditional systems. More importantly, the time to detection for critical activities such as attempts to access generation control systems was reduced from an average of 4.2 hours to 23 minutes.

The validation exercises also revealed areas for continued improvement, particularly in detecting novel attack techniques not represented in training data. Zero-day exploits and previously unknown attack vectors continued to pose challenges, highlighting the importance of combining AI-based detection with human expertise and threat intelligence feeds.

9. Future Directions and Recommendations

9.1. Emerging Technologies and Trends

The future evolution of AI-based threat detection in smart grid environments will be shaped by several emerging technologies and industry trends that promise to enhance both capabilities and challenges in critical infrastructure protection. Understanding these developments is essential for utilities planning long-term cybersecurity strategies.

Quantum Computing Impact: The advent of practical quantum computing poses both opportunities and threats for smart grid cybersecurity. While quantum computers could potentially break current encryption methods, they also offer possibilities for quantum-enhanced machine learning algorithms that could dramatically improve threat detection capabilities. Utilities should begin preparing for post-quantum cryptography while exploring quantum machine learning applications.

5G and Beyond Communications: The deployment of 5G networks for smart grid communications will enable new applications requiring ultra-low latency and high bandwidth. However, these networks also introduce new attack vectors and complexity that AI systems must address. Edge computing capabilities enabled by 5G infrastructure will provide new opportunities for distributed threat detection architectures.

Artificial General Intelligence (AGI) Developments: As AI systems become more sophisticated and approach artificial general intelligence capabilities, their application to cybersecurity will evolve beyond pattern recognition to include strategic reasoning and adaptive response capabilities. These developments could enable autonomous cybersecurity systems that can anticipate and counter sophisticated attacks without human intervention.

Digital Twin Integration: The integration of digital twin technologies with AI-based threat detection offers promising opportunities for enhanced situational awareness and predictive security capabilities. Digital twins of critical infrastructure assets could enable simulation-based threat analysis and testing of security measures before deployment in operational environments.

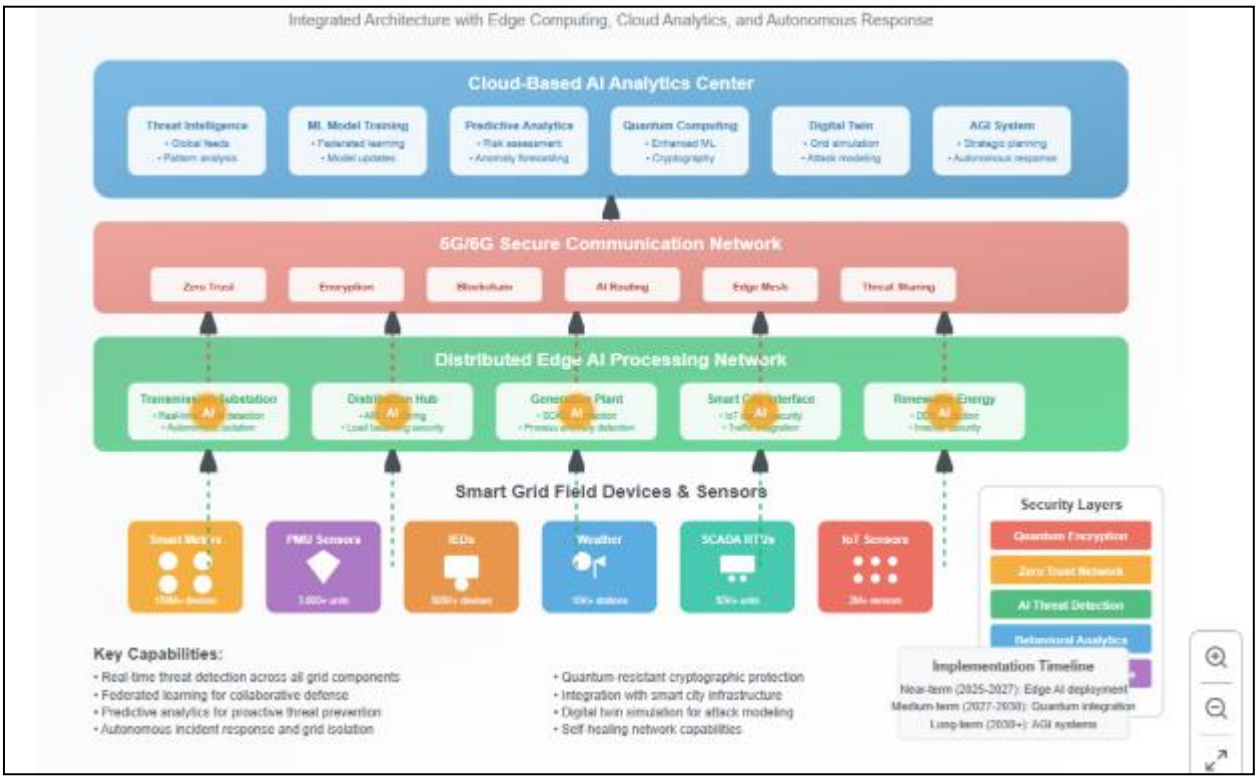

Figure 5 Future Vision for AI-Enhanced smart grid cybersecurity

9.2. Industry Collaboration and Standards Development

The complexity and interconnected nature of smart grid infrastructure necessitate collaborative approaches to cybersecurity that extend beyond individual utility organizations. Industry-wide standards and collaborative frameworks will be essential for achieving effective protection against sophisticated threat actors.

Information Sharing Enhancements: Current information sharing mechanisms through organizations such as the Electricity Subsector Coordinating Council (ESCC) and Multi-State Information Sharing and Analysis Center (MS-ISAC) provide valuable threat intelligence. However, AI-enabled systems require more detailed and timely information sharing to achieve optimal effectiveness. Development of automated threat intelligence sharing protocols could enable real-time distribution of attack signatures and indicators of compromise.

Federated Learning Networks: The establishment of industry-wide federated learning networks could enable utilities to collaboratively improve their AI models while maintaining the privacy and security of sensitive operational data. Such networks could significantly enhance detection capabilities by leveraging diverse datasets and attack experiences across multiple organizations.

Standards Development Priorities: Several areas require focused standards development efforts to support widespread adoption of AI-based threat detection systems:

- Interoperability Standards: Ensuring AI systems can effectively share information and coordinate responses across different vendor platforms and utility organizations
- Performance Metrics: Establishing standardized methods for measuring and comparing the effectiveness of AI-based security systems in power system environments
- Data Governance: Developing frameworks for managing and sharing sensitive operational data required for AI system training and operation
- Ethical AI Guidelines: Addressing concerns about algorithmic bias, transparency, and accountability in critical infrastructure protection applications

9.3. Policy and Regulatory Recommendations

The successful deployment of AI-based threat detection systems in smart grid environments requires supportive policy and regulatory frameworks that encourage innovation while maintaining appropriate oversight and accountability. Several key recommendations emerge from this analysis.

Regulatory Modernization: Current cybersecurity regulations for critical infrastructure, while comprehensive, were developed before the widespread adoption of AI technologies. Regulatory frameworks should be updated to address AI-specific considerations including algorithmic transparency, bias prevention, and validation requirements for machine learning systems used in critical infrastructure protection.

Incentive Programs: Government incentive programs could accelerate the adoption of advanced cybersecurity technologies by utilities, particularly smaller organizations that may lack resources for significant technology investments. Such programs could include research and development grants, cost-sharing arrangements, and regulatory credit for proactive security investments.

Public-Private Partnerships: Enhanced collaboration between government agencies and private sector utilities is essential for addressing sophisticated nation-state threats that target critical infrastructure.

Compliance with ethical standards

Disclosure of conflict of interest

No conflict of interest to be disclosed.

References

[1] Achaal, B., Adda, M., Berger, M. et al. Study of smart grid cyber-security, examining architectures, communication networks, cyber-attacks, countermeasure techniques, and challenges. Cybersecurity 7, 10 (2024). https://doi.org/10.1186/s42400-023-00200-w

[2] Alazab, A., Khraisat, A., Singh, S., & Jan, T. (2023). Enhancing Privacy-Preserving Intrusion Detection through Federated Learning. Electronics, 12(16), 3382. https://doi.org/10.3390/electronics12163382

[3] Alcaraz, C., & Zeadally, S. (2014). Critical infrastructure protection: Requirements and challenges for the 21st century. International Journal of Critical Infrastructure Protection, 8, 53–66. https://doi.org/10.1016/j.ijcip.2014.12.002

[4] Alimi, O. A., Ouahada, K., & Abu-Mahfouz, A. M. (2020). A Review of Machine Learning Approaches to Power System Security and Stability. IEEE Access, 8, 113512-113531. https://doi.org/10.1109/ACCESS.2020.3003568

[5] Biggio, B., & Roli, F. (2018). Wild patterns: Ten years after the rise of adversarial machine learning. Pattern Recognition, 84, 317–331. https://doi.org/10.1016/j.patcog.2018.07.023

[6] Chen, T. M., Sanchez-Aarnoutse, J. C., & Buford, J. (2011). Petri Net Modeling of Cyber-Physical Attacks on Smart Grid. IEEE Transactions on Smart Grid, 2(4), 741-749. https://doi.org/10.1109/TSG.2011.2160000

[7] Chen, Y., Jiang, H., Li, C., Jia, X., & Ghamisi, P. (2016). Deep Feature Extraction and Classification of Hyperspectral Images Based on Convolutional Neural Networks. IEEE Transactions on Geoscience and Remote Sensing, 54(10), 6232-6251. https://doi.org/10.1109/TGRS.2016.2584107

[8] Eseosa, N. O., & Ikposhi, N. A. (2021). Review of machine learning applications to power systems studies. Open Access Research Journal of Engineering and Technology, 1(1), 021–031. https://doi.org/10.53022/oarjet.2021.1.1.0101

[9] Gunduz, M. Z., & Das, R. (2020). Cyber-security on smart grid: Threats and potential solutions. Computer Networks, 169, 107094. https://doi.org/10.1016/j.comnet.2019.107094

[10] Huang, R., Li, Y., & Wang, X. (2022). Attention-aware deep reinforcement learning for detecting false data injection attacks in smart grids. International Journal of Electrical Power & Energy Systems, 147, 108815. https://doi.org/10.1016/j.ijepes.2022.108815

[11] Mallidi, S.K.R., Ramisetty, R.R. Advancements in training and deployment strategies for AI-based intrusion detection systems in IoT: a systematic literature review. Discov Internet Things 5, 8 (2025). https://doi.org/10.1007/s43926-025-00099-4

[12] Muneer, S., Farooq, U., Athar, A., Raza, M. A., Ghazal, T. M., & Sakib, S. (2024). A Critical Review of Artificial intelligence based Approaches in Intrusion Detection: A Comprehensive analysis. Journal of Engineering, 2024, 1–16. https://doi.org/10.1155/2024/3909173

[13] National Institute of Standards and Technology. (2024). NIST Cybersecurity Framework 2.0: A profile for critical infrastructure protection. U.S. Department of Commerce. https://doi.org/10.6028/NIST.CSWP.29

[14] Truong, N., Sun, K., Wang, S., Guitton, F., & Guo, Y. (2021). Privacy preservation in federated learning: An insightful survey from the GDPR perspective. Computers & Security, 110, 102402. https://doi.org/10.1016/j.cose.2021.102402

[15] Wang, X., Li, J., Kuang, X., Tan, Y., & Li, J. (2019). The security of machine learning in an adversarial setting: A survey. Journal of Parallel and Distributed Computing, 130, 12–23. https://doi.org/10.1016/j.jpdc.2019.03.003

[16] Zhang, S., Yao, L., Sun, A., & Tay, Y. (2019). Deep Learning based recommender system. ACM Computing Surveys, 52(1), 1–38. https://doi.org/10.1145/3285029

[17] Zhang, Y., Wang, J., & Chen, B. (2021). Detecting False Data Injection Attacks in Smart Grids: A Semi-Supervised Deep Learning Approach. IEEE Transactions on Smart Grid, 12(1), 623-634. https://doi.org/10.1109/TSG.2020.3010510