VOL. 08 NO. 9,
MAY, 2025
E-ISSN 3026-958X
P-ISSN 3027-1169

African Scholar Publications & Research International
www.africanscholarpub.com

Journal of Science Innovation & Technology Research (JSITR)

Machine Learning for Cybersecurity in the Utilities and Power Sector: A Practical Approach to Infrastructure Protection

Fatima Rilwan Ododo [1]; & Ridwan Rahmat Sadiq

[1]Computer Science, Montana State University, Bozeman, MT 59717, USA. [2]Computer Science, Nasarawa State University, Keffi, Nigeria.

Correspondence author: fatimaododo@montana.edu
DOI: https://doi.org/10.70382/ajsitr.v8i9.038

Abstract

Cyberattacks on critical infrastructure systems, including power grids and water treatment facilities, pose a growing threat to public safety and national security. The increasing complexity and connectivity of these systems have introduced new vulnerabilities that traditional rule-based cybersecurity solutions struggle to address. This paper presents a practical exploration of how machine learning (ML) techniques can enhance cybersecurity within the utilities and power sector. We examine common cyber threats affecting industrial control systems, explore how ML techniques—such as supervised learning and anomaly detection—can detect and mitigate these threats, and propose a simple, interpretable ML based security framework. Through real-world incidents and hypothetical use cases, we demonstrate the practicality and effectiveness of ML in improving infrastructure resilience.

Keywords: Cybersecurity, machine learning, critical infrastructure, utilities sector, power grid, SCADA, anomaly detection, industrial control systems, supervised learning, reinforcement learning.

Introduction

The increasing digitalization of critical infrastructure has created both unprecedented operational efficiencies and complex cybersecurity vulnerabilities. In sectors such as electricity generation, water purification, and gas distribution, industrial control systems (ICS) and supervisory control and data acquisition (SCADA) platforms have become central to automation and remote management. However, these systems, originally designed with minimal consideration for cybersecurity, are now exposed to a rapidly evolving threat landscape (Alcaraz and Zeadally, 2015; Sridhar et al., 2011). The convergence of operational technology (OT) with information technology (IT), coupled with the integration of legacy hardware and software into modern networks, has significantly increased the attack surface of utility infrastructures (Ododo and Sadiq, 2025a).

Cyberattacks on the utilities and power sector are no longer hypothetical. Real-world incidents illustrate the tangible risks. The 2015 cyberattack on Ukraine's power grid disrupted electricity supply to more than 230,000 people and involved the use of remote access trojans, spear-phishing, and destructive malware (Case, 2016). More recently, in 2021, the Colonial Pipeline ransomware attack temporarily halted fuel distribution across the eastern United States, demonstrating how digital vulnerabilities can translate into physical and economic disruption (Sanger et al., 2021). Such incidents are particularly alarming because utility systems are considered high-value targets by both state-sponsored and financially motivated threat actors.

Traditional cybersecurity approaches, such as rule-based intrusion detection systems (IDS), firewalls, and access control policies, are often insufficient in modern critical infrastructure environments (Ododo and Sadiq, 2025b). These approaches generally rely on known attack signatures or manually crafted rules and struggle to detect novel or zero-day exploits. Moreover, utility environments present constraints such as real-time processing requirements, the need for high availability, and limited tolerance for false positives—factors that complicate the deployment of conventional cybersecurity tools (Morris et al., 2011; Ghafir et al., 2014).

To address these limitations, researchers and practitioners are increasingly turning to machine learning (ML) as a proactive and adaptive defense mechanism (Ododo and Sadiq, 2025a; Kalonde et al., 2024a). ML-based systems are capable of learning patterns in historical data and identifying deviations that may indicate malicious activity (ODODO and ADDOTEY, 2025a). This data-driven approach is particularly suited to environments where manual threat identification is impractical due to scale, complexity, or data variability (Ahmed et al., 2016). Techniques such as supervised learning, anomaly detection, and reinforcement learning have been

applied to intrusion detection, anomaly recognition in network traffic, and dynamic defense strategy formulation in ICS and smart grid settings (Sun et al., 2022; Zhang et al., 2015).

However, applying ML to the utilities sector presents unique challenges. These include the scarcity of labeled cybersecurity datasets, the risk of adversarial manipulation of ML models, integration difficulties with legacy systems, and limited explainability of complex algorithms (Doshi-Velez and Kim, 2017; Koay et al., 2023). Consequently, effective deployment of ML in infrastructure security must balance technical innovation with system compatibility and interpretability.

This paper aims to provide a comprehensive yet accessible exploration of how machine learning can support cybersecurity in the utilities and power sector. We begin by reviewing prevalent cyber threats in this domain and then explain, in intuitive terms, the main ML techniques applicable to infrastructure protection. We further propose a simplified framework for ML-based threat detection and response that can be implemented in resource constrained settings. Drawing from both real-world cases and hypothetical scenarios, this work illustrates the potential of ML to enhance the resilience of critical infrastructure systems against modern cyber threats.

Literature Review

The application of machine learning (ML) in cybersecurity for critical infrastructure systems has garnered growing attention due to the limitations of conventional rule-based approaches in dynamic threat environments. Over the past decade, a significant body of research has explored ML-based techniques for detecting and mitigating cyber threats in industrial control systems (ICS), supervisory control and data acquisition (SCADA) networks, and smart grid environments (ODODO and ADDOTEY, 2025b). This section reviews relevant contributions and highlights persistent gaps that motivate the need for practical and interpretable ML solutions in the utilities and power sector.

ML Applications in Industrial Control Systems: ICS platforms such as programmable logic controllers (PLCs) and distributed control systems (DCS) are foundational to critical infrastructure but are often built on legacy technologies with limited built-in security. Several studies have demonstrated the viability of ML in monitoring ICS behavior for signs of intrusion or sabotage.

Goh et al. (2017) introduced a dataset and anomaly detection model using Long Short-Term Memory (LSTM) neural networks to monitor sensor and actuator signals in a secure water treatment (SWaT) system. Their model successfully detected cyber-physical attacks by identifying deviations in time-series patterns. Similarly, Inoue et al. (2017) applied

unsupervised deep learning to model ICS behavior and detect previously unseen threats. These approaches underscore the effectiveness of sequence modeling and anomaly detection in ICS environments.

However, these models rely on well-structured, clean datasets, which are often unavailable in real-world ICS settings. Ahmed, Mahmood, and Hu (2016) emphasized that most ICS environments lack labeled data due to infrequent and covert attack occurrences. This has led to a growing interest in semi-supervised and unsupervised learning methods that do not rely heavily on historical attack data.

Cybersecurity in SCADA and Smart Grid Systems: SCADA systems, widely used in utilities, integrate remote sensors and control systems over communication networks. Their distributed nature and reliance on legacy communication protocols make them attractive targets for cyber adversaries. Numerous works have applied ML to enhance SCADA system defenses.

Morris, Gao, and Reaves (2010) evaluated the performance of support vector machines (SVMs) in identifying replay and command injection attacks in SCADA traffic. Their findings show that SVMs, when trained with appropriate features, can distinguish between benign and malicious patterns with high accuracy. Similarly, Al-Madani, Shawahna, and Qureshi (2019) developed an anomaly detection approach for SCADA networks that combines support vector machines (SVMs), decision trees, and k-nearest neighbors (KNN) classifiers. Their ensemble method, which utilizes inter-arrival curve features, demonstrated high accuracy in distinguishing normal and malicious traffic while effectively reducing false positives.

In the smart grid domain, Boyaci et al. (2021) introduced a Graph Neural Network (GNN)–based model for detecting stealth false data injection attacks in substations and grid environments, leveraging the physical and topological relationships between grid nodes to identify subtle anomalies that conventional classifiers often overlook.

Predictive Defense and Adaptive Security Policies: Beyond passive monitoring, researchers have explored the use of ML for active defense and cyber risk forecasting. El Mrabet, Kaabouch, and Ghazi (2018) proposed a probabilistic risk assessment model using Bayesian networks to estimate the likelihood of successful cyber intrusions. Such models help prioritize defenses for high-risk assets.

Ni and Paul (2017) proposed a reinforcement learning (RL) framework to model the sequential decision-making process between attackers and defenders in smart grid environments. By formulating the interaction as a Markov decision process (MDP), their approach enables an RL agent to simulate multi-stage attack scenarios and learn optimal

defensive strategies over time. While this method shows promise for adaptive and proactive defense, RL techniques require extensive training and validation to ensure safe deployment, particularly in systems with strict real-time constraints.

Limitations of Current ML-Based Security Solutions: Despite their potential, ML-based security systems face several practical challenges in critical infrastructure contexts: Data Scarcity and Quality: Most ML algorithms require large, representative datasets. However, real-world attack data is often limited, proprietary, or inconsistent (Goh et al., 2017).

- Adversarial ML Risks: Sophisticated attackers may craft inputs designed to fool ML models, a vulnerability known as adversarial evasion (Koay et al., 2023).
- Explainability and Trust: Complex models, especially deep neural networks, are difficult to interpret. Doshi-Velez and Kim (2017) argue that explainability is essential for operator trust and regulatory compliance in safety-critical systems.
- Integration with Legacy Systems: Utility infrastructures frequently operate on outdated platforms with limited support for modern cybersecurity tools, complicating ML deployment (Sridhar et al., 2011).

Research Gaps and Opportunities: The literature establishes a strong foundation for applying machine learning to intrusion detection and risk assessment in utility systems. However, many existing approaches rely on idealized assumptions—such as the availability of labeled datasets, high-performance computing infrastructure, and complete system visibility—that often diverge from the realities encountered in operational environments.

To address these limitations, there is a growing demand for lightweight, interpretable, and modular ML solutions that can function effectively in resource-constrained settings and integrate seamlessly with legacy monitoring tools. In addition, the absence of standardized benchmarks and publicly available datasets continues to impede reproducibility and the generalization of models across different utility domains (Sarker et al., 2020).

This paper responds to these gaps by proposing a streamlined and adaptable machine learning framework tailored for real-time anomaly detection and threat response within utilities and power systems.

Methodology

This paper adopts a narrative review methodology to synthesize current knowledge and propose a practical machine learning (ML) framework for cybersecurity in the utilities and power sector. Relevant academic and technical literature was collected from major digital databases, including IEEE Xplore, ScienceDirect, SpringerLink, and Google Scholar. The

search focused on publications from 2010 to 2025 to ensure coverage of both foundational work and recent developments in machine learning applications for critical infrastructure protection.

Keywords used during the search included: "machine learning in SCADA security," "ICS anomaly detection," "cybersecurity in utilities," "smart grid ML defense," "reinforcement learning for infrastructure protection," and "adversarial ML in industrial control." Articles were included if they provided empirical findings, proposed ML-based frameworks, or addressed security threats in utility systems such as electricity, gas, or water. Preference was given to peer-reviewed studies, technical reports, and government publications.

The review excluded papers unrelated to the utilities sector or those lacking a machine learning component. Additionally, the proposed ML-based framework presented in this paper is based on insights derived from patterns observed in the reviewed literature. It reflects a synthesis of successful defense strategies, practical deployment considerations, and emerging ML techniques tailored to real-time, resource-constrained environments.

This methodological approach ensures transparency, reproducibility, and alignment with current research trends, while also allowing for the design of a simplified, modular, and context-aware framework suitable for operational deployment in critical infrastructure settings.

Cyber Threats in Utilities and Power Systems

Cyber threats targeting utility infrastructure—particularly electricity, water, and gas systems—have increased in both frequency and sophistication. These infrastructures rely on interconnected systems of sensors, controllers, and remote communication networks that present a wide and often poorly secured attack surface. Unlike generic IT systems, cyberattacks in this sector have real-world consequences, including widespread service disruptions, economic losses, environmental damage, and even risks to human life (Sridhar et al., 2011; Alcaraz and Zeadally, 2015). This section provides an overview of common cyber threats faced by utilities and power systems, focusing on their mechanisms, examples, and impact.

Malware and Ransomware: Malware, particularly ransomware, has emerged as a dominant threat to utility organizations. In ransomware attacks, malicious actors encrypt files or lock control systems, demanding payment for restoration. The 2021 Colonial Pipeline attack is a notable example, where the DarkSide ransomware group halted fuel distribution across the U.S. East Coast, causing significant supply chain disruptions (Sanger et al., 2021). Such attacks exploit weak user authentication, outdated software, or unmonitored remote access points to gain control over critical assets.

Phishing and Social Engineering: Phishing remains a primary vector for initial compromise in many cyberattacks. These attacks rely on deceiving employees into revealing login credentials or installing malware. The Ukraine power grid attack in 2015 began with a phishing campaign that delivered BlackEnergy malware via malicious email attachments (Case, 2016). Once attackers gained internal access, they used remote access tools to disrupt operations at multiple substations.

Advanced Persistent Threats (APTs): APTs are prolonged and targeted attacks often linked to state-sponsored actors. Their goal is to silently infiltrate a network, maintain persistence, and gradually gain control over key systems. APTs targeting utilities often aim to disable or sabotage infrastructure. For instance, the Triton malware discovered in 2017 was designed to compromise safety instrumented systems in an industrial facility, posing a severe risk of physical harm (El Mrabet et al., 2018).

SCADA and ICS-Specific Exploits: Utility operations are managed through SCADA and ICS environments, which communicate over standardized but often insecure protocols such as Modbus, DNP3, and IEC 60870. Attacks against these systems may involve: • Replay attacks, where previously recorded control commands are resent to disrupt operations.

- Command injection, where unauthorized or modified control instructions are transmitted.
- Parameter tampering, which alters configuration settings, thresholds, or operational logic.

These exploits can result in flooding, blackouts, or equipment destruction if not detected early (Morris et al., 2011).

Insider Threats: Insider threats—whether malicious or accidental—pose a unique challenge, as they originate from users with legitimate access. Former employees, contractors, or negligent staff may introduce malware, leak credentials, or alter control settings. Detecting such threats often requires behavioral profiling and real-time monitoring that can differentiate between normal and anomalous actions.

Internet of Things (IoT) Vulnerabilities: Modern utility systems increasingly integrate IoT devices, including smart meters, environmental sensors, and remote-control units. While these devices improve operational visibility, they also introduce new vulnerabilities. Many IoT devices lack encryption, regular firmware updates, or intrusion

detection mechanisms, making them attractive entry points for attackers (Yang et al., 2017).

Machine Learning Techniques for Infrastructure Protection

Machine learning (ML) offers a dynamic and adaptive approach to detecting and mitigating cyber threats in critical infrastructure environments (Kalonde et al., 2024b). Unlike traditional security tools, which rely on predefined signatures or static rules, ML models can learn from historical data, detect subtle deviations, and adapt to emerging attack patterns (Garba et al., 2019). This is particularly valuable in utilities and power systems, where the operating environment is complex, the cost of downtime is high, and attack vectors are increasingly diverse. In this section, we review key ML techniques applicable to infrastructure protection, categorizing them based on their learning paradigms and practical use cases.

Supervised Learning: Supervised learning involves training a model on a labeled dataset where input features are mapped to known outputs (e.g., "normal" or "malicious" activity). This paradigm is effective for detecting known attack types and classifying incoming events.

For example, support vector machines (SVM), decision trees, and neural networks have been employed to classify SCADA network traffic and sensor behaviors (Morris et al., 2011). These models can achieve high accuracy, provided the training data is representative and well-balanced. However, their effectiveness diminishes when facing previously unseen or evolving threats, especially in real-time ICS environments where labeled data is scarce.

Unsupervised Learning and Anomaly Detection: Unsupervised learning techniques are well suited to environments where labeled attack data is unavailable. These methods seek to identify patterns or clusters in data and flag points that deviate from learned norms as anomalies.

Clustering algorithms (e.g., k-means, DBSCAN) and dimensionality reduction techniques (e.g., PCA, t-SNE) are often used to establish baselines of normal behavior in ICS and SCADA systems. Inoue et al. (2017) demonstrated the use of autoencoders—unsupervised deep learning models—to detect cyber-physical anomalies in a water treatment system by reconstructing input signals and analyzing reconstruction error. These approaches are particularly effective for identifying novel or stealthy intrusions, but they may also suffer from high false positive rates in noisy environments.

Semi-Supervised Learning: Semi-supervised learning bridges the gap between supervised and unsupervised approaches. It leverages a small amount of labeled data and a larger pool of unlabeled data to improve model performance. Techniques such as label propagation and self-training have been explored for SCADA traffic classification and anomaly ranking (Ahmed et al., 2016)

These methods are attractive for real-world utility applications where labeled cyberattack data is limited, yet vast volumes of unlabeled telemetry are available. However, model accuracy heavily depends on the quality of the labeled subset and the assumptions made about the distribution of the data.

Reinforcement Learning: Reinforcement learning (RL) involves training an agent to interact with an environment and learn optimal strategies through reward and punishment. In cybersecurity, RL has been used to simulate adversarial interactions and develop adaptive defense strategies.

Ni and Paul (2019) proposed a reinforcement learning framework for smart grid security that models the interaction between defenders and attackers as a multistage game. In this setting, the defender learns to adaptively allocate security resources in response to evolving threats using a reinforcement learning agent. While this approach demonstrates the potential for developing proactive and resilient defense strategies, it also highlights the challenges of requiring extensive simulation environments and high computational costs—factors that can hinder real-world deployment.

Ensemble Learning: Ensemble methods—such as random forests, boosting algorithms (e.g., AdaBoost, XGBoost), and stacking—are widely used to enhance anomaly detection accuracy and robustness in industrial environments. For example, Al-Abassi et al. (2020) proposed a deep learning–based ensemble framework combining deep neural networks (DNNs), decision trees, random forests, and AdaBoost to detect cyber-attacks in ICS networks. Their approach, evaluated on two real-world ICS datasets, outperformed traditional classifiers and significantly reduced false positives.

These models are effective in reducing variance and capturing diverse attack patterns, making them suitable for complex ICS datasets. However, they may introduce computational overhead and reduce transparency, depending on the number and nature of the base learners involved.

Deep Learning and Time-Series Modeling: Given the sequential nature of sensor data and network traffic in utilities, deep learning models—particularly recurrent neural

networks (RNNs), long short-term memory (LSTM), and convolutional neural networks (CNNs)—are well-suited for time-series anomaly detection.

Goh et al. (2017) applied LSTM networks to the SWaT dataset to predict sensor values and detect deviations. Their results showed that temporal dependencies in ICS data could be leveraged to improve detection accuracy. Despite their effectiveness, deep learning models require large datasets, significant computational resources, and careful tuning to prevent overfitting.

Proposed Framework for ML-Based Infrastructure Protection

To translate the potential of machine learning (ML) into practical cybersecurity solutions for utilities and power systems, it is essential to design frameworks that are not only effective but also lightweight, interpretable, and compatible with existing operational technology. Many ML-based solutions in research are too complex, data-hungry, or computationally intensive for real-time industrial environments. This section introduces a modular and resource-conscious ML framework tailored for anomaly detection and response in critical infrastructure.

Framework Overview: The proposed framework as shown in Figure 1 consists of four core components:

- Data Acquisition and Preprocessing: Data is collected from various sources such as sensors, actuators, control logs, network traffic, and user authentication systems. Preprocessing includes normalization, timestamp alignment, handling missing values, and feature extraction. Feature engineering may involve computing statistical summaries (e.g., mean, variance) or domain-specific indicators (e.g., pump pressure gradients, power usage patterns).
- Behavior Modeling and Learning: Depending on the availability of labeled data, the system uses:
    1. Supervised learning (e.g., decision trees, SVM) for known attack patterns.
    2. Unsupervised learning (e.g., clustering, autoencoders) for anomaly detection without labeled input.
    3. Hybrid or semi-supervised learning for environments with partial labels. Model choice is guided by operational constraints—e.g., LSTM for time-series, ensemble models for classification robustness.
- Anomaly Detection and Alerting: The trained model is deployed in a streaming or batch mode to analyze incoming data in real time. It flags any deviation from the learned baseline as a potential anomaly. Alerts are classified based on severity

levels (e.g., informational, warning, critical) and can trigger predefined rule-based responses, such as isolating a subnet, disabling a compromised user, or alerting a human analyst.

- Response and Feedback Integration: Human analysts can validate, override, or escalate alerts. This feedback loop is vital for retraining models to reduce false positives and adapt to changing conditions. Over time, the system refines its detection thresholds and retrains periodically or continuously to improve resilience.

System Integration and Deployment Considerations: For adoption in real-world utility environments, the following considerations are essential:

- Interoperability: The system should support integration with existing SCADA and ICS platforms via APIs or protocol adapters (e.g., OPC UA, Modbus).
- Edge vs. Cloud Deployment: Lightweight models (e.g., decision trees, anomaly score thresholds) may be deployed at the edge near field devices, while more complex models (e.g., deep learning) can operate in centralized or cloud-based security operations centers (SOCs).
- Explainability: Models must provide interpretable outputs—e.g., highlighting which features contributed to an alert—to foster operator trust.
- Fail-Safe Mechanisms: In high-risk environments, ML-driven decisions must be accompanied by manual override options and fail-safe triggers.
- Fail-Safe Mechanisms: In high-risk environments, ML-driven decisions must be accompanied by manual override options and fail-safe triggers.

Scalability and Resilience: The modular nature of the framework allows it to scale across different infrastructures—e.g., from a single water treatment plant to a national power grid. Redundancy, data replication, and secure communication protocols are included to maintain resilience during attacks or component failures.

Case Studies and Hypothetical Scenarios

To demonstrate the applicability of the proposed ML-based cybersecurity framework, this section presents real-world incidents and hypothetical scenarios relevant to the utilities and power sector. These examples illustrate how machine learning could have enhanced detection, prevention, and response efforts in each case.

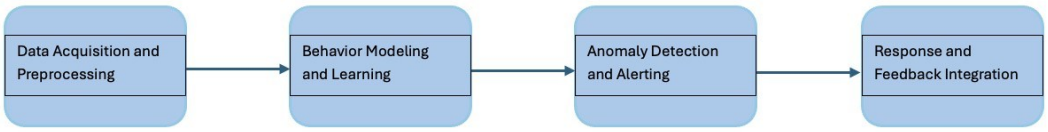


Figure 1: Conceptual framework for ML-based cybersecurity in utility systems. The architecture includes four modular stages—data acquisition, behavior modeling, anomaly detection, and response feedback—designed for real-time threat mitigation and adaptive learning in resource-constrained infrastructure environments.

Real-World Case Study: Ukraine Power Grid Attack (2015) **-** The December 2015 cyberattack on Ukraine's power grid is one of the most prominent incidents involving infrastructure compromise. Attackers used spear-phishing emails to deliver BlackEnergy malware, enabling remote access to industrial control systems (Case, 2016). They manipulated operator interfaces to disconnect substations, overwrote firmware in serial-to-ethernet devices, and launched denial-of-service attacks on call centers to prevent outage reports.

ML Application:

- A supervised learning model trained on operator activity logs and remote login patterns could have identified anomalous behaviors (e.g., login from unusual IP addresses, repeated unauthorized access attempts).
- An unsupervised model could detect changes in system commands and human-machine interface (HMI) patterns inconsistent with normal operations.
- A reinforcement learning-based policy engine could have dynamically adjusted firewall rules or revoked remote access during the anomaly.

Real-World Case Study: *Colonial Pipeline Ransomware Attack (2021)*- In May 2021, Colonial Pipeline, which supplies nearly half of the U.S. East Coast's fuel, was hit by a ransomware attack from the DarkSide group (Sanger et al., 2021). Although the attackers targeted the business IT network, fear of lateral spread led the company to shut down the operational technology (OT) systems.

ML Applications:

- A real-time anomaly detection model could monitor internal network behavior to detect early signs of command-and-control communication.

- Feature extraction on email metadata and attachment content could enhance phishing detection using ensemble classifiers.
- The feedback loop component of the ML framework would allow continuous retraining of models based on analyst-labeled incidents, improving detection of novel ransomware variants.

Hypothetical Scenario: Water Treatment Chemical Overdose - Imagine a water treatment plant where a compromised PLC begins injecting excessive chlorine into the system. Traditional threshold-based alerts might miss this if the increase is gradual or if limits are manually changed by the attacker.
ML Application:

- An LSTM-based predictive model trained on historical sensor data could detect subtle trends in chlorine levels inconsistent with expected fluctuations.
- The anomaly detection module would raise alerts even in the absence of static threshold violations, enabling preemptive intervention.

Hypothetical Scenario: Insider Threat in Substation Automation - An engineer with valid access credentials begins downloading large configuration files and modifying setpoints outside of scheduled maintenance periods. While technically authorized, this behavior is inconsistent with routine operation.
ML Application:

- A behavioral profiling model could compare the engineer's current actions with their historical activity and those of similar users. • Role-based anomaly detection, combined with contextual features (e.g., time of day, device accessed), could flag suspicious activity.
- The response module might initiate session recording, notify security personnel, or restrict access until review.

Challenges and Future Directions

While machine learning holds significant promise for enhancing cybersecurity in utilities and power systems, several challenges limit its practical deployment. A major concern is the scarcity of high-quality, labeled data. Most real-world cyberattacks on critical infrastructure are confidential or rare, making it difficult to train supervised models effectively. Although testbed datasets like SWaT are useful, they often fail to reflect the complexity of operational environments (Goh et al., 2017; Morris et al., 2011). Incomplete or imbalanced data can also lead to false positives, undermining the reliability of ML-based alerts (Ahmed et al., 2016).

Another limitation is the lack of interpretability in complex models, particularly deep learning systems. In safety-critical settings, explainability is essential for building operator trust and meeting regulatory requirements (Doshi-Velez and Kim, 2017). Additionally, ML models are vulnerable to adversarial attacks, such as data poisoning or evasion techniques that exploit model weaknesses (Koay et al., 2023). Ensuring robustness against such threats remains an active area of research.

Integration with legacy infrastructure is another barrier. Many utility systems rely on outdated hardware and software that cannot support resource-intensive ML frameworks. This calls for lightweight, edge-deployable models that can function within constrained environments (Sridhar et al., 2011). Furthermore, anomaly detection models often produce high false alarm rates, contributing to alert fatigue and diminishing effectiveness.

The absence of standardized benchmarks also hinders the evaluation of ML solutions. Without consistent datasets and metrics, it is difficult to assess performance across different systems and contexts (Sarker et al., 2020). Moving forward, research should focus on developing interpretable and resilient models, leveraging techniques like transfer learning and federated learning, and integrating human expertise through human-in-the-loop systems. Embedding ML into the broader cybersecurity lifecycle—from detection to response—will be key to building resilient and intelligent infrastructure defenses.

Conclusion

As cyber threats to critical infrastructure grow in scale and sophistication, traditional rule-based security measures are proving inadequate, particularly in complex and resource constrained environments like utilities and power systems. Machine learning presents a promising avenue for enhancing cybersecurity by enabling systems to learn from operational data, detect anomalies in real time, and adapt to evolving threats. This paper has examined the nature of cyber threats in the utilities sector, reviewed key machine learning techniques relevant to infrastructure protection, and proposed a practical, interpretable ML-based framework suitable for real-world deployment.

Through real-world incidents and hypothetical scenarios, we have illustrated how ML can strengthen threat detection and response capabilities across diverse attack vectors—including malware, insider threats, and SCADA exploits. Despite its potential, the application of ML in this domain is not without challenges, including data limitations, model explainability, adversarial risks, and integration barriers. Addressing these issues will require collaborative efforts across research, industry, and regulatory bodies to develop resilient, transparent, and scalable solutions.

Ultimately, machine learning is not a standalone solution but a critical enabler within a multi-layered defense strategy. When carefully designed and integrated, ML-based

systems can significantly improve the resilience of utility infrastructure, safeguarding essential services against the increasingly dynamic landscape of cyber threats.

References

Abdulrahman Al-Abassi, Hadis Karimipour, Ali Dehghantanha, and Reza M Parizi. An ensemble deep learning-based cyber-attack detection in industrial control system. *Ieee Access*, 8:83965–83973, 2020.

Abigail MY Koay, Ryan K L Ko, Hinne Hettema, and Kenneth Radke. Machine learning in industrial control system (ics) security: current landscape, opportunities and challenges. *Journal of Intelligent Information Systems*, 60(2):377–405, 2023.

Aliyu Garba, Sandip Rakshit, and Fatima Rilwan. Detection and sentiment analysis of hate speech on twitter in nigerian politics. In *Proceedings of: 2nd International Conference of the IEEE Nigeria*, page 285, 2019.

Basem Al-Madani, Ahmad Shawahna, and Mohammad Qureshi. Anomaly detection for industrial control networks using machine learning with the help from the inter-arrival curves. *arXiv preprint arXiv:1911.05692*, 2019.

Cristina Alcaraz and Sherali Zeadally. Critical infrastructure protection: Requirements and challenges for the 21st century. *International journal of critical infrastructure protection*, 8:53–66, 2015.

David E Sanger, Clifford Krauss, and Nicole Perlroth. Cyberattack forces a shutdown of a top us pipeline. *The New York Times*, 8, 2021.

Defense Use Case. Analysis of the cyber attack on the ukrainian power grid. *Electricity information sharing and analysis center (E-ISAC)*, 388(1-29):3, 2016.

FATIMA ODODO and NICHOLAS ADDOTEY. Advancements and challenges in deep learning for cyber threat detection. *International Journal of Science Research and Technology*, 2025a.

FATIMA ODODO and NICHOLAS ADDOTEY. Understanding the influence of outliers on machine learning model interpretability. *International Journal of African Sustainable Development Research*, 2025b.

Fatima Rilwan Ododo and Ridwan Rahmat Sadiq. A review of outlier detection techniques in cybersecurity: A machine learning perspective. *Journal of Science Innovation and Technology Research*, 2025b.

Fatima Rilwan Ododo and Ridwan Rahmat Sadiq. Adversarial attacks in cybersecurity: A machine learning perspective. *Journal of Science Innovation and Technology Research*, 2025a.

Finale Doshi-Velez and Been Kim. Towards a rigorous science of interpretable machine learning. *arXiv preprint arXiv:1702.08608*, 2017.

Gilbert Kalonde, Samuel Boateng, Lateefat Sanni, Silas Chotwe, and Fatima Ododo. Artificial intelligence and special education: The use and the integration. In *Society for Information Technology & Teacher Education International Conference*, pages 1926–1932. Association for the Advancement of Computing in Education (AACE), 2024b.

Gilbert Kalonde, Samuel Boateng, Silas Chotwe, Lateefat Sanni, and Fatima Ododo. Integration of computer science in rural math classroom: A case study. In *Society for Information Technology & Teacher Education International Conference*, pages 1757–1761. Association for the Advancement of Computing in Education (AACE), 2024a.

Ibrahim Ghafir, Martin Husak, and Vaclav Prenosil. A survey on intrusion detection and prevention systems. In *Proceedings of student conference Zvule, IEEE/UREL. Brno University of Technology*, volume 1014, 2014.

Iqbal H Sarker, ASM Kayes, Shahriar Badsha, Hamed Alqahtani, Paul Watters, and Alex Ng. Cybersecurity data science: an overview from machine learning perspective. *Journal of Big data*, 7:1–29, 2020.

Ji Zhang, Hongzhou Li, Qigang Gao, Hai Wang, and Yonglong Luo. Detecting anomalies from big network traffic data using an adaptive detection approach. *Information Sciences*, 318:91–110, 2015.

Jonathan Goh, Sridhar Adepu, Marcus Tan, and Zi Shan Lee. Anomaly detection in cyber physical systems using recurrent neural networks. In *2017 IEEE 18th International Symposium on High Assurance Systems Engineering (HASE)*, pages 140–145. IEEE, 2017.

Jun Inoue, Yoriyuki Yamagata, Yuqi Chen, Christopher M Poskitt, and Jun Sun. Anomaly detection for a water treatment system using unsupervised machine learning. In *2017 IEEE international conference on data mining workshops (ICDMW)*, pages 1058–1065. IEEE, 2017.

Mohiuddin Ahmed, Abdun Naser Mahmood, and Jiankun Hu. A survey of network anomaly detection techniques. *Journal of Network and Computer Applications*, 60:19–31, 2016.

Osman Boyaci, Mohammad Rasoul Narimani, Katherine R Davis, Muhammad Ismail, Thomas J Overbye, and Erchin Serpedin. Joint detection and localization of stealth false data injection attacks in smart grids using graph neural networks. *IEEE Transactions on Smart Grid*, 13(1):807–819, 2021.

Siddharth Sridhar, Adam Hahn, and Manimaran Govindarasu. Cyber–physical system security for the electric power grid. *Proceedings of the IEEE*, 100(1):210–224, 2011.

Thomas Morris, Anurag Srivastava, Bradley Reaves, Wei Gao, Kalyan Pavurapu, and Ram Reddi. A control system testbed to validate critical infrastructure protection concepts. *International Journal of Critical Infrastructure Protection*, 4(2):88–103, 2011.

Wei Gao, Thomas Morris, Bradley Reaves, and Drew Richey. On scada control system command and response injection and intrusion detection. In *2010 eCrime Researchers Summit*, pages 1–9. Ieee, 2010.

Xin Sun, Zhijun Tang, Mengxuan Du, Chaoping Deng, Wenbin Lin, Jinshan Chen, Qi Qi, and Haifeng Zheng. A hierarchical federated learning-based intrusion detection system for 5g smart grids. *Electronics*, 11(16):2627, 2022.

Yuchen Yang, Longfei Wu, Guisheng Yin, Lijie Li, and Hongbin Zhao. A survey on security and privacy issues in internet-of-things. *IEEE Internet of things Journal*, 4(5):1250–1258, 2017.

Zakaria El Mrabet, Naima Kaabouch, Hassan El Ghazi, and Hamid El Ghazi. Cybersecurity in smart grid: Survey and challenges. *Computers & Electrical Engineering*, 67: 469–482, 2018.

Zhen Ni and Shuva Paul. A multistage game in smart grid security: A reinforcement learning solution. *IEEE transactions on neural networks and learning systems*, 30(9): 2684–2695, 2019.

Zhen Ni, Shuva Paul, Xiangnan Zhong, and Qinglai Wei. A reinforcement learning approach for sequential decision-making process of attacks in smart grid. In *2017 IEEE Symposium Series on Computational Intelligence (SSCI)*, pages 1–8. IEEE, 2017.