Smart Technologies Academic Press

Journal of Cyber Security and Risk Auditing
https://www.jcsra.thestap.com/

A Risk-Based Cybersecurity Auditing Framework for Smart Grid Infrastructure Using Explainable Artificial Intelligence (XAI)

Udit Mamodiya [1], Indra Kishor [2], Hastimal Jangid [3], Rommel AlAli[4*], Ashraf M. Zaher[5] and Shoeb Saleh[6]

[1]*Associate Professor, Faculty of Engineering and Technology, Poornima University, Jaipur 303905, Rajasthan, India*
[2]*Assistant Prof. Dept. of CSE, Poornima Institute of Engineering and Technology, Jaipur 302022, Rajasthan, India*
[3]*Independent Researcher, Coozmoo Digital Solutions, Houston, USA*
[4]*Associate professor, the National Research Center for Giftedness and Creativity, King Faisal University, Al-Ahsa, Saudi Arabia*
[5]*Translation, Authorship and Publication Center, King Faisal University, Al-Ahsa 31982, Saudi Arabia*
[6]*Associate professor, the National Research Center for Giftedness and Creativity, King Faisal University, Al-Ahsa, Saudi Arabia*

ARTICLE INFO

Article History

Received: 20-01-2026
Revised: 25-05-2026
Accepted: 20-06-2026
Published: 27-06-2026

Vol.2026, No.2

DOI:

Corresponding author.
Email: *ralali@kfu.edu.sa*

Orcid:
0000-0001-7375-4856

ABSTRACT

This study suggests a framework for cybersecurity auditing of smart grid infrastructure, which is based on the concept of risk and the use of Explainable Artificial Intelligence (XAI) to produce transparent, prioritized and audit-ready security evidence. The information from public smart grid cybersecurity events was mapped to event labels, asset classes, security-control status, compliance indicators, and cyber-physical impact variables, which were then used to create audit-relevant records. Attack likelihood estimates were made using machine learning models. The attack likelihood, asset criticality, control deficiency score, compliance condition and operational impact were all added together to calculate the final audit risk score. Explainability was used as a technique to identify the most important features that affected each audit decision by applying the SHAP method. The proposed framework achieved 96.38% accuracy, 96.51% precision, 96.38% recall, 96.42% F1-score, and 0.996 ROC-AUC. The results of the ablation showed that the inclusion of the risk component and the XAI component resulted in an improvement in the risk ranking, audit traceability and explanation consistency. The framework translates the cybersecurity detection results into an understandable audit decision, enabling risk-based remediation, compliance review, and an understandable smart grid cybersecurity governance.

Keywords: Smart grid cybersecurity; risk-based auditing; explainable artificial intelligence; cyber-physical security; audit risk scoring; SHAP.

How to cite the article

ISSN: 3079-5354

1. Introduction

Smart grid infrastructure is one of the most important digital infrastructure of modern power system. Smart grids are fundamentally different from traditional electricity networks, which are primarily used for one-way energy transmission, by incorporating sensing devices, intelligent electronic devices, advanced metering infrastructure, phasor measurement units, supervisory control and data acquisition systems, distributed energy resources, and cloud/edge-based analytics into a tightly coupled cyber-physical environment. This change enhances the observability, demand response, integration of renewables, outage management, and operational efficiency. But it is also the same connectivity that increases the attack surface of the power grid. Thus, cybersecurity in smart grid systems should be considered as a risk-governance need in order to ensure reliable electricity delivery, as well as a technical protection [1], [2].

With a combination of regulatory requirements, industrial standards, vendor requirements and operational requirements, the auditability of smart grid cybersecurity has become a critical issue. A typical cybersecurity audit will focus on access control, patching, network segmentation, incident response documentation and compliance documentation. These procedures are helpful, but tend to be checklist oriented. In smart grid, cyber risk is dynamic, however. The criticality of the assets, communication exposure, vulnerability status, attack likelihood, and potential physical impact are all factors that increase the risk of a substation gateway, energy management server or field-level relay. A software vulnerability in a very connected control node could pose a greater operational risk than a high severity vulnerability in a stand-alone administrative system. This mismatch is a good reason to move away from compliance verification to risk-based prioritization for smart grid cybersecurity auditing [3], [4].

This trend is also backed up by recent cyber-threat intelligence. The use of credentials, remote-service attacks, and adversary-in-the-middle attacks, manipulation of firmware, command injection, denial-of-control attacks, and lateral movement across IT-OT boundaries are all methods used to target industrial control systems. In these cases, the security auditor has to not only establish that controls are in place, but also that they are effective in preventing realistic adversarial techniques. This is challenging due to a variety of protocols, legacy devices, vendor-specific configurations and long asset life cycles in the grid infrastructure. It can be time consuming and expensive to replace insecure devices. There are many operational assets that cannot be scanned or patched on a regular basis without disrupting the services. Therefore, the audit process needs to be structured and interpret the incomplete evidence, operational limitations and changing threat patterns [5], [6].

This problem can be better supported with risk-based cybersecurity auditing. It does not audit all controls and assets equally, but prioritizes the audit findings based on their contribution to grid risk. For instance, if a remote engineering workstation is poorly secured, switching-command patterns are unusual, or there is no segmentation around a substation automation network, it should be given more attention if the asset is highly dependent on the operation of the network [7]. Risk-based auditing also facilitates the transformation of the technical findings into information for operators, compliance teams and management that can be used for decision making. But manual risk scoring is subjective, time consuming and can vary between audit teams. Different priorities can be given to the same vulnerability by two auditors when they interpret the asset context, threat exposure or cascading consequence differently. With the large scale deployment of smart grid infrastructure, this subjectivity is a significant constraint when continuous telemetry and security alerts are generated [8], [9].

In the smart grid and industrial system, intrusion detection, anomaly recognition, false data injection detection, malware classification and cyber-risk assessment have been widely researched using artificial intelligence. Machine learning models are able to detect non-obvious correlations between network traffic, control commands, deviations in the sensors, event logs, and historical attack patterns. Under complex operating conditions, deep learning and ensemble methods could be used to enhance the detection accuracy. However, it is not enough to have high accuracy for cybersecurity auditing. Auditors and grid operators should be aware of the rationale behind the identification of a particular asset, event or control as high risk. A black box alert with no explanation might be counterproductive to good governance. In critical infrastructure, the lack of accountability issues can also arise from an unexpected recommendation, particularly if it impacts compliance reporting, incident response or operational continuity [9], [10].

Explainable Artificial Intelligence (XAI) is a solution to this concern by making the decisions made by the model more interpretable. Explanation is not a pretty face in the cybersecurity audit process, it's a part of the evidence chain. The audit report should show if the feeder automation controller was determined to be high risk because of abnormal traffic volume, use of insecure protocols, outdated firmware, indicators of privilege escalation, lack of access logs, or high operational criticality. These explanations enable security teams to verify the results of the model, question assumptions, and correlate technical results with corrective measures. In the context of smart grid, explainability is particularly useful as the operational engineers, cybersecurity analysts, and auditors may have different views on the same incident [11], [12].

Although these developments have been made, the current research typically focuses on cybersecurity of the smart grid in one of three disjointed areas: compliance frameworks, intrusion detection models, or the general cyber-risk assessment. A single auditing framework that integrates asset criticality, control effectiveness, threat exposure, threat scoring with AI, and evidence generation with explanations is not fully developed. There are a number of AI-based intrusion detection systems that claim to have excellent classification accuracy, but do not always produce a risk report that can be used for auditing. Likewise, cybersecurity best practices establish helpful controls, but they don't necessarily prioritize findings according to the context of the real-time operation. This is the motivation for the present study [13], [14].

This paper is novel because it proposes a framework for cybersecurity auditing of smart grid infrastructure based on Explainable Artificial Intelligence (XAI) in a risk-based approach. The proposed framework combines the cyber-physical asset profiling, vulnerability and control assessment, threat-informed risk scoring and interpretation with XAI. The framework is not just a yes/no answer to whether a control is compliant or non-compliant as is the traditional audit model. It provides an estimate of the relative risk contribution of each asset and finding, describes the factors that affect the assigned score and provides support for transparent prioritization of remediation actions. Model explanations are meant to be audit evidence, allowing for a traceability of the data inputs, risk indicators and final recommendations [15] [16]. It is not only a scientific but also a practical motivation. In practice, utilities require scalable, evidence-based and easily understood auditing methods for both the cybersecurity and power-system communities. From a scientific perspective, there is a gap between the accuracy of AI detection and accountability at the governance level. From a scientific point of view, there is a need to close the gap between the accuracy of AI detection and accountability at the governance level. This research will enhance trust, decrease subjectivity, and increase the support of decision making in the smart grid cybersecurity management by incorporating explainability into the risk-based auditing process. The proposed study thus joins the trend in the field of cyber security from reactive security monitoring to intelligent, transparent and risk-aware cyber assurance of critical energy infrastructure [17] and [18].

2. Literature Review

2.1 Explainable AI for Smart Grid Fault and Security Assessment

Many early machine learning techniques provided useful information in identifying faults, unusual load patterns and suspicious communication patterns, but lacked the ability to provide insights into the rationale behind a particular decision. This is a major restriction in critical energy infrastructure. An incorrect or misinterpreted alert could cause an operator to delay action, lead to an audit problem or cause unnecessary maintenance.

The application of explainable AI in microgrid fault detection has recently been demonstrated to be useful for transparency in classification and enhance technical confidence in the automated diagnosis [19]. This direction is very applicable to the audit field of cybersecurity, as the audit teams are not only interested in prediction results. A smart grid-aware cloud computing framework has also highlighted the importance of distributed computation and sustainable energy management in the modern grid operation, demonstrating a close relationship between cybersecurity and energy management, instead of them being separate engineering issues [20]. Likewise, recent studies on the use of AI in smart grid security point to the growing reliance on smart grid systems on intelligent control, energy routing and automated decision making in transmission and distribution systems [21].

2.2 AI-Driven Cybersecurity and Hybrid Protection Models

The smart grid AI cybersecurity models have been the subject of a vast amount of research. The studies are typically based on deep learning, hybrid classification, blockchain, and adaptive security modules for the identification of

cyberattacks or for enhancing the trust of the system. For instance, blockchain-based cybersecurity solutions are suggested to enhance the trust, data integrity, and decentralized validation in smart grid systems [22]. But, blockchain integration is not enough to address the auditing issue. It can help to safeguard records or transactions, but it doesn't necessarily provide an explanation of what asset is risky, what control is broken, or how an auditor should prioritize remediation.

Hybrid graph-based and temporal methods have been employed to detect anomalies and threats in networks by deep learning [23]. These methods can be applied when attacks are network dependent and/or when they are time dependent deviations. The adaptive detection of malware and lightweight encryption are two areas of IoT-based cybersecurity that have been investigated in resource constrained environments [24]. The ideas are relevant to smart grids as many of the field devices are computation limited. However, there is more attention paid to detection accuracy in the literature than audit interpretation. That is, the model can detect an attack, but there is a lack of an audit trail.

The applications of AI in the smart grid for energy efficiency also demonstrate that intelligent algorithms can be used in many other areas beyond security. They can be used for demand forecasting, load balancing, energy optimization and fault response [25]. This implies an important research implication. A cybersecurity audit framework for smart grids needs to be aware of the operational functions of each asset, rather than its cyber configuration. An endpoint that is a critical monitoring point should have a higher audit priority than a less critical endpoint.

2.3 Smart Grid Reliability, Forecasting, and Interpretability

The smart grid management is usually researched in terms of reliability and operational optimization. A holistic grid-ready management system has been suggested to enhance the reliability of the grid by integrating a control, monitoring and decision-support system [26]. The contributions are significant as cybersecurity risk in smart grids is not independent of reliability. The impact of a cyber-incident on voltage stability, feeder coordination, and communication delay or service continuity escalates the severity of the incident. Recent research also has fused federated learning, attention mechanisms and explainable AI for secure and interpretable smart grid load forecasting [27]. This is a helpful direction as it can decrease the amount of data that needs to be centralized and XAI can provide explanations of model behaviour. However, prediction studies typically are concerned with operational prediction and not formal auditing. A secure prediction model is not a risk-based audit model. The latter has to relate technical evidence to the effectiveness of the controls, the vulnerability exposure and the decision making related to compliance. Hybrid machine learning models have also been created for smart grid cybersecurity enhancement [28]. These models usually have multiple classifiers in order to increase the robustness of detection. The real-time cyberattack detection in smart energy grid is also proposed by a related fuzzy–deep learning framework that combines uncertainty handling and deep representation learning [29]. These studies demonstrate progress, but are mostly detection based. They do not usually offer a formal process to translate the results of the models into audit results, risk ratings or remediation priority.

2.4 Anomaly Detection, Fault Diagnosis, and Cyber-Physical Resilience

The anomaly detection has been a hot topic in AI-based intelligent distribution systems. Under complex grid conditions, deep learning models can be used to detect abnormal behavior, and when traditional threshold-based methods are not effective, they can be used to detect abnormal behavior [30]. But anomaly detection results might not be readily applicable to cybersecurity auditing. An audit finding should be understandable, documented and linked to the asset level risk. As such, there needs to be interpretation and contextual scoring to support anomaly detection.

The use of explainable boosting models for fault detection in electrical transmission systems has been recently investigated, where the decisions made by the models are more explainable than the black-box architectures [31]. This is useful for auditing a grid as it can help minimize the difference between automated classification and manual classification. Meanwhile, new security debates have been raised about the use of artificial intelligence and quantum cryptography as a means to more robust protection models [32]. The methods outlined here could prove useful in the future for secure communication, but are still being developed for use in audit procedures.

The enhancement of resilience has also been investigated using deep reinforcement learning for adaptive smart power grid [33]. These techniques can be used to aid dynamic response in varying situations. Machine learning-based security and anomaly identification have been studied in the context of inverter-based cyber-physical microgrids to tackle the new

vulnerabilities that have been created by power electronic interfaces [34]. Another important research direction is false data injection detection, where false measurements can lead to false estimation and control and operational decisions [35].

2.5 Cybersecurity Resilience, Green Security, and Lightweight Detection

A cybersecurity resilience framework for smart grid stability has been proposed that includes a wider range of protection strategies that involve prevention, detection, response, and recovery [36]. This is more in line with the auditing perspective as resilience needs to be more than just an ability to detect attacks. It needs to be measurable, and have a practical recovery plan. AI, machine learning, and large language models are also explored in the context of energy-efficient threat detection and sustainable security operations, with a focus on green cybersecurity. Additionally, the use of AI, machine learning, and large language models in energy-efficient threat detection and sustainable security operations is explored, highlighting the importance of green cybersecurity. While this is a developing field, it does bring up an important issue: the computational and energy costs of security measures should not be too high in the smart grid infrastructure.

The research on energy efficient materials and electromagnetic design, although not explicitly about cyber security, is part of a broader trend to energy efficient and resilient energy technologies [38]. Concurrently, an adaptive intrusion detection system (IDS) and scalable threat mitigation using energy-efficient hybrid deep learning has been explored for IoT applications [39]. These light weight models are applicable in the field level smart grid devices, where low latency and low power consumption is required. More comprehensive research on smart grid security has also been presented to consider the security of sustainable energy systems at the communication, control and data layers [40].

The literature has recently started to reveal privacy and communication-layer threats in dynamic smart grid networks in the context of real-time eavesdropping detection [41]. While created in a different prediction setting, comparisons of quantum-inspired and classical machine learning suggest that cutting-edge computational models are being tested for complex environmental and energy-related systems increasingly. Comparisons between quantum-inspired and classical machine learning, created in another prediction context, suggest that cutting-edge computational models are increasingly being tested for complex environmental and energy-related systems. Last but not least, work at the review level on AI and quantum cryptography further underscores the importance of more robust security models in the future cyber-physical infrastructures [42].

2.6 Research Gaps

While the previous research has contributed to the smart grid cybersecurity, most of the research work has been done from different perspectives like standards, intrusion detection, blockchain security, forecasting and explainable fault diagnosis. Although these works are valuable, they are not sufficient to enable risk-based cybersecurity auditing, which requires linking the criticality of assets, weakness of controls, cyber-physical impact, AI-based risk scoring, and explaining audit evidence. The major gaps are summarized in Table 1 and how the proposed study addresses the gaps is also summarized. Table 1 demonstrates that there are building blocks available in the literature that can be used to build a smart grid cybersecurity auditing mechanism, but not yet a full one.

Table 1. Research gap and gap-to-solution mapping for smart grid cybersecurity auditing

Author(s) / Ref.	Focus Area	Main Contribution	Key Gap	Link with Present Study
[1]	Smart grid cybersecurity guideline	Defines smart grid security domains, actors, interfaces, and control needs.	Static guideline; no AI-based risk scoring or dynamic audit prioritization.	Provides the baseline control structure for the proposed audit model.
[2]	Cyber-risk governance	Presents CSF 2.0 functions for governance, protection, detection, response, and recovery.	Generic framework; not specific to smart grid cyber-physical assets.	Supports risk-governance alignment in the proposed framework.
[3]	IACS component security	Specifies technical security requirements for industrial control components.	Compliance-focused; lacks explainable risk ranking and audit intelligence.	Helps map smart grid devices to control-level audit checks.

[13]	Cyberattack detection in energy infrastructure	Uses machine learning for attack detection and response recommendation.	Detection output is not converted into audit-ready evidence.	Motivates AI-supported audit decision-making.
[18]	Explainable grid fault diagnosis	Provides interpretable fault classification and location in transmission lines.	Focuses on electrical faults, not cybersecurity audit risk.	Supports the need for explainability in grid decision systems.
[19]	XAI for microgrid fault detection	Uses explainable AI for transparent fault classification.	XAI results are not linked with audit scoring or remediation priority.	Supports XAI-based justification of audit findings.
[22]	AI–blockchain smart grid security	Improves trust, integrity, and protection using AI and blockchain.	Secure records are addressed, but audit prioritization remains limited.	Useful for secure evidence handling in audit workflow.
[27]	Federated learning and XAI	Combines federated learning, attention, and XAI for secure forecasting.	Forecasting-centered; does not assess cybersecurity controls.	Supports distributed and explainable grid intelligence.
Identified Research Gap	Risk-based XAI cybersecurity auditing	Existing works improve standards, detection, trust, or explainability separately.	No unified model links asset criticality, control weakness, AI risk score, cyber-physical impact, and explainable audit evidence.	This study proposes a transparent, risk-prioritized XAI auditing framework for smart grid infrastructure.

Therefore, the current study aims to address this gap by proposing a risk-based explainable AI framework that transforms the data from the smart grid into prioritized, explainable, and auditable cybersecurity evidence.

3. Methodology

3.1 Experimental Design and Audit-Oriented Workflow

The proposed methodology was designed as an experimental cybersecurity auditing framework for smart grid infrastructure, not as a purely conceptual model. The basic concept is to transform the raw smart grid security observations into an audit-ready risk score, and then explain that score using XAI, which can then be reviewed by an auditor, grid engineer, and cybersecurity analyst. The workflow integrates public smart grid attack data, standard-based control mapping, standard classification using machine learning, risk scoring, and explanation generation.

The process of the experiment is divided into five steps: Data acquisition, Preprocessing, Asset-risk mapping, Attack classification, Explainable audit scoring. The structure of the proposed methodology is shown in Figure 1. The control baseline is aligned with NIST smart grid cybersecurity guidance, NIST CSF 2.0, and ISA/IEC 62443 component-level security requirements [1]–[3]. AI-based attack detection and cyber-risk modeling are included because recent studies show that machine learning can improve cyberattack detection and decision support in critical energy infrastructure [13], [29], [35].

The Figure 1 illustrates the complete audit pipeline, starting from public smart grid cybersecurity data acquisition and preprocessing to asset-risk mapping, machine learning-based cyberattack detection, explainable AI interpretation, and final audit risk reporting. As shown in Figure 1, the output of the methodology is not limited to attack or normal classification. Each event is converted into a risk-prioritized audit record containing the predicted security state, affected asset type, control weakness, risk score, explanation features, and remediation priority.

3.2 Threat Model and Audit Assumptions

The following assumptions are made when creating the threat model and conducting an audit:

The proposed framework is based on the assumption that the smart grid infrastructure has cyber-physical components that are interconnected, such as PMUs, relays, substation gateways, communication nodes, and control-center assets. The adversary is assumed to be able to alter measurement streams, add false data, alter the state of relays or breakers, compromise weak authentication or use communication channels inappropriately. The study does not assume that

ISSN: 3079-5354

equipment is destroyed, but rather that there is a cyber-risk to the operation, in which the operation is affected by the presence of malicious or abnormal data that can impact monitoring, protection or audit decisions.

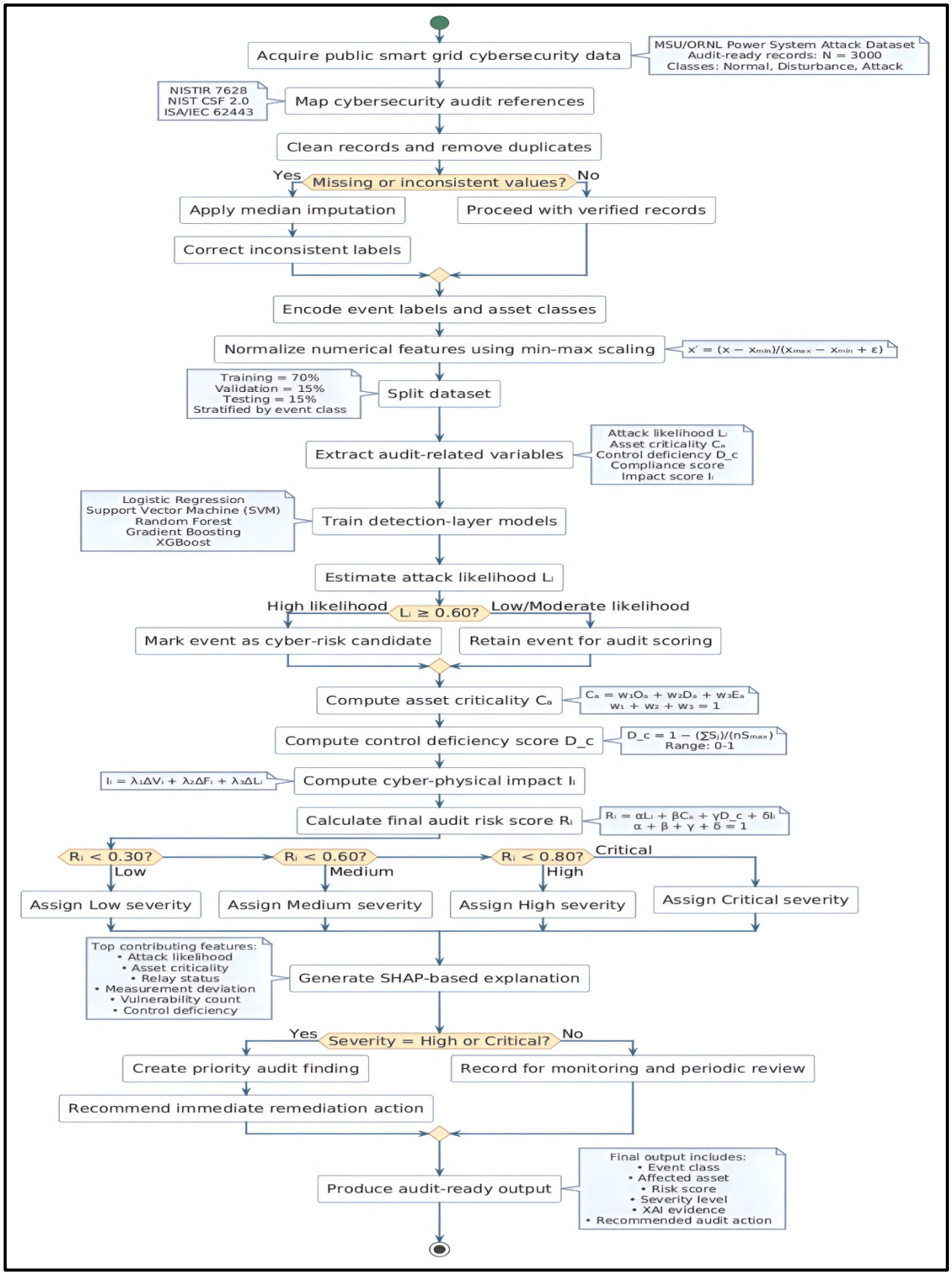


Figure 1. Proposed experimental workflow for risk-based cybersecurity auditing of smart grid infrastructure.

The audit model also assumes that cybersecurity controls can be assessed using indicators like authentication status, logging availability, segmentation, patch condition, vulnerability count, and compliance score that are based on standards. These assumptions are in line with the smart grid cyber security guidance and the security requirements for ICS [1]–[3]. The threat model is used to connect attack likelihood, asset criticality, control deficiency, compliance condition and cyber-physical impact, which is used to formulate the risks later on.

3.3 Public Dataset Selection and Experimental Data Source

The experiment is reproducible by using publicly available smart grid cybersecurity data instead of synthetic data that are only created for this experiment. The Power System Attack Dataset from the Mississippi State University/Oak Ridge National Laboratory (MSU/ORNL) is the primary dataset chosen for the experiment, and is extensively used for smart grid cyberattack and disturbance detection. According to the public description of the data set, it has three versions of the data set, which are based on 15 sets of data, 37 event scenarios, such as normal operation, natural event, attack event, etc. Disturbances of the cyber-physical type are considered in the attack scenarios for power system monitoring and control.

The data set is appropriate for the current study because it has classes of events labeled which can be correlated with cybersecurity audit classes. It also supports evaluation of false data injection, command manipulation, disturbance events, and abnormal grid states, which are directly relevant to smart grid audit evidence. Since false data injection is one of the most serious attack forms in smart grids, the use of such public attack data is consistent with recent intrusion detection studies [35]. “The dataset accession link, selected file version, number of records, number of features, and class distribution will be reported in the experimental setup to ensure reproducibility.

Table 2. Public dataset and audit-role mapping used in the experimental methodology.

Dataset / Source	Event Coverage	Data Type	Audit Use in This Study	Reason for Selection
MSU/ORNL Power System Attack Dataset	Normal events, natural disturbances, cyberattack events	PMU/relay-based smart grid measurements and labels	Attack detection, risk scoring, audit evidence mapping	Publicly available, labelled, and widely used for smart grid security evaluation
NISTIR 7628 control guidance [1]	Smart grid cybersecurity domains	Control and requirement mapping	Audit baseline formation	Provides grid-specific security control structure
NIST CSF 2.0 [2]	Govern, identify, protect, detect, respond, recover	Cyber-risk governance functions	Audit reporting and risk governance alignment	Supports risk-based cybersecurity management
ISA/IEC 62443 [3]	IACS component security	Industrial control security requirements	OT device-level control mapping	Relevant to substations, relays, gateways, and field devices

As summarized in Table 2, the dataset supplies measurable cyber-physical evidence, while the standards supply the audit control structure.

3.3.1 Sample Data Structure and Audit Variable Mapping

To ensure the transparency of the experimental design, the chosen public smart grid cybersecurity dataset was first structured into groups of features relevant to the audit process, and then preprocessed and used to build the models. The goal of this step was to not only consider the dataset as an intrusion detection source, but to also avoid it. Instead, each feature group was mapped to its cybersecurity audit role, operational relevance, and risk-scoring contribution. As shown in Table III, the variables include electrical measurements, relay indicators, PMU observations, event labels, asset class, control deficiency indicators, audit records, and XAI-based evidence. This structure allows the proposed framework to connect attack detection with risk-based audit decision-making, which is consistent with smart grid cybersecurity control guidance and industrial control security requirements [1]–[3].

ISSN: 3079-5354

Table 3. Sample data structure and audit-variable mapping used in the proposed smart grid cybersecurity auditing framework.

S. No.	Feature Category	Example Variables	Sample Range / Example Values	Data Type	Audit Relevance	Role in Proposed Framework
1	Electrical measurements	Voltage, current, frequency, phase angle	Voltage: 0.92–1.08 p.u.; frequency: 49.5–50.5 Hz	Numerical	Indicates abnormal physical behavior or disturbance in grid operation.	Used for cyber-physical impact estimation and attack detection.
2	Relay and breaker status	Relay trip, breaker open/close, protection flag	0 = normal, 1 = triggered/open/tripped	Binary / categorical	Shows protection response, switching activity, or unauthorized control action.	Helps identify high-impact operational events.
3	PMU / sensor observations	Time index, signal deviation, measurement sequence	Deviation score: 0–1; sequence ID: event-based	Numerical / temporal	Captures measurement inconsistency and possible false data injection.	Used as model input for anomaly and attack likelihood estimation.
4	Event category	Normal, disturbance, attack	0 = normal, 1 = disturbance, 2 = attack	Categorical label	Defines the ground-truth state for supervised evaluation.	Used for classifier training, validation, and testing.
5	Asset class	Control center, substation, relay, PMU, communication node	Criticality score: 0.2–1.0	Categorical / numerical	Represents the affected smart grid component and its operational importance.	Used to calculate asset criticality in the audit risk model.
6	Security control status	Authentication, logging, segmentation, monitoring	0 = absent, 1 = partially satisfied, 2 = satisfied	Score-based	Shows whether required security controls are implemented.	Used to estimate control deficiency score using standard-based mapping.
7	Risk attributes	Attack likelihood, asset criticality, control deficiency score, impact score	Normalized score: 0–1	Numerical	Combines cyber and operational indicators into audit-level risk.	Used to compute final risk-based audit score.
8	XAI evidence	SHAP value, top contributing features, feature rank	SHAP value: positive/negative contribution; rank: 1–10	Numerical / ranked	Explains why an event is classified as risky.	Used for audit justification and traceable evidence generation.
9	Compliance and audit records	Patch status, policy compliance score, vulnerability count, audit findings	Patch: 0/1; compliance: 0–100%; vulnerability count: integer	Numerical / categorical	Reflects cybersecurity maturity, policy compliance, and unresolved audit issues.	Supports risk prioritization, remediation ranking, and audit decision-making.

Compliance and audit records add a governance layer by representing patch status, policy adherence, vulnerability exposure, and previous audit findings. These variables are not treated as attack labels; rather, they are used as audit-side metadata derived from standard-based control assessment using NISTIR 7628, NIST CSF 2.0, and ISA/IEC 62443 guidance [1]–[3].

The feature grouping in Table 3 defines how raw smart grid measurements and audit-side metadata are converted into model-ready variables. Electrical, relay, and PMU-related variables support event detection, while asset class and security-control status connect the detected event with audit requirements. Compliance and audit records add a governance layer by representing patch status, policy adherence, vulnerability exposure, and previous audit findings. To further clarify how the processed data were used in the experiment, Table IV presents a sample audit dataset record after feature mapping and normalization.

Table 4. Sample audit dataset record generated after audit-variable mapping.

Record ID	Asset Class	Event Label	Voltage Dev.	Relay Status	Attack Likelihood	Asset Criticality	Control Deficiency Score	Compliance Score	Vulnerability Count	Impact Score	Final Audit Risk	Severity
R-001	PMU	Normal	0.04	0	0.08	0.55	0.2	92	1	0.12	0.21	Low
R-002	Relay	Disturbance	0.31	1	0.42	0.82	0.38	76	3	0.47	0.53	Medium
R-003	Substation gateway	Attack	0.58	1	0.81	0.91	0.64	61	6	0.72	0.77	High
R-004	Control center	Attack	0.67	1	0.89	1	0.78	48	9	0.84	0.88	Critical
R-005	Communication node	Normal	0.09	0	0.15	0.7	0.31	84	2	0.18	0.32	Medium

Each record has technical and audit related variables as seen in Table 4. The event label can be used for supervised classification and attack likelihood is the output of the machine learning. The final audit risk is then calculated based on asset criticality, control deficiency score, compliance score, vulnerability count and impact score. This format allows the dataset to be used in a risk-based cybersecurity auditing process since each prediction is associated with an asset, a control condition, and an audit severity level.

The last audit risk in Table 4 is calculated later with the risk formulation in (5) which takes into account the likelihood of attack, criticality of the asset, the score of the control deficiency and the cyber-physical impact.

3.3.2 Data Preprocessing and Feature Normalization

The raw event data were initially checked for missing data, non-numeric data, duplicate rows, and data label inconsistencies. The values that were missing were filled in with median imputation to minimize the impact of outliers. The labels of the events were coded into the following audit classes: normal, disturbance, attack, and high-risk attack. The final label mapping was kept simple as the more labels are broken up into classes, the weaker the audit interpretation. All the numerical features were normalized using the min–max scaling method as shown in (1):

$$x_i' = \frac{x_i - x_{min}}{x_{max} - x_{min} + \epsilon} \cdots (1)$$

where the i^{th} feature's original value is x_i, the minimum and maximum values of the i^{th} feature in the training set are x_{min} and x_{max}, respectively, and ϵ is a small constant to prevent division by zero. All features are scaled to a similar range prior to classification and risk scoring to ensure they are all contributing within a similar range.

The dataset was split into the training, validation and testing sets with the ratio of 70:15:15 after normalization. Stratified splitting was used to ensure that the samples of normal, disturbance and cyberattack were proportionally represented in each subset. This step is crucial because the smart grid cybersecurity datasets are frequently imbalanced and attack events could be fewer than normal events, but the impact on the operation of the smart grid is much greater.

3.4 Asset Criticality and Control Weakness Mapping

The proposed framework does not classify each cyber event, but rather considers it to be an audit sensitive event. So, before the risk scoring, each record was associated with an asset class that was affected and a control condition. This step is critical as the same attack risk can lead to varying audit severity depending on whether it involves a PMU, relay, substation gateway, control-center workstation or communication node. The mapping of the asset and control categories

was done with the NISTIR 7628, NIST CSF 2.0 and ISA/IEC 62443 [1]–[3] smart grid security requirements. The feature groups and sample audit ready records for this mapping are shown earlier in Table 3 and Table 4.

The asset criticality score was determined based on the operational importance, dependency level and exposure level as outlined in (2):

$$C_a = w_o O_a + w_d D_a + w_e E_a \cdots (2)$$

C_a is the criticality score of asset a, O_a is the operational importance of asset a, D_a is the dependency level of asset a and E_a is the exposure level of asset a. The weighting factors w_o, w_d and we are the factors assigned to the operational importance, dependency and exposure, respectively. In this study, it was assumed that the weights are constrained, i.e. $w_o+w_d+w_e=1$. The higher the C_a is the more cyber-physical implications may be severe in the event of failure or compromise of the asset. Then the weakness of the control was calculated as (3):

$$D_c = 1 - \frac{\sum_{j=1}^{n} S_j}{nS_{max}} \cdots (3)$$

where D_c is the control deficiency score, S_j is the satisfaction score of the j^{th} cybersecurity control, n is the number of applicable controls and S_{max} is the maximum possible score assigned to each control. This version allows absent, partial, and satisfied controls. A value close to 0 indicates stronger control satisfaction, while a value close to 1 indicates a weaker control posture. The control set was derived from NISTIR 7628, NIST CSF 2.0, and ISA/IEC 62443 [1]– [3]. Table 5 is used to convert raw cyber events into audit-sensitive observations. This is where the method becomes risk-based rather than purely prediction-based.

Table 5. Asset criticality and control-mapping criteria for smart grid cybersecurity auditing.

Asset Class	Example Components	Criticality Basis	Control Mapping	Audit Priority
Control center	SCADA server, EMS, operator workstation	System-wide control and visibility	Access control, logging, recovery, monitoring	Very high
Substation automation	RTU, relay, IED, gateway	Protection and switching dependency	Authentication, segmentation, secure communication	High
Field sensing layer	PMU, sensor, smart meter	Measurement integrity	Data validation, encryption, device identity	Medium to high
Communication layer	Router, switch, protocol gateway	Attack propagation path	Network segmentation, anomaly detection, traffic monitoring	High
DER interface	PV inverter, storage controller	Distributed control effect	Firmware integrity, access control, command validation	Medium

3.5 Machine Learning-Based Cyberattack Classification

A supervised machine learning module was used to classify smart grid events into audit-relevant security states. Tree-based ensemble learning was selected as the primary classifier because it performs well on structured tabular cybersecurity data and supports explainability through feature contribution analysis. Random Forest, Gradient Boosting, and XGBoost were evaluated as candidate classifiers. Their performance was compared using accuracy, precision, recall, F1-score, and area under the ROC curve. The predicted class probability for an event i is represented in (4):

$$P_i(y = k \mid X_i) = f_\theta(X_i) \cdots (4)$$

where X_i is the feature vector of event i, y=k denotes the predicted class, f_θ represents the trained machine learning model, and θ contains the learned model parameters. The predicted attack probability P_i is later used as one input to the audit risk score.

The classifier was trained using five-fold cross-validation on the training set. Hyperparameters were selected through grid search. Class imbalance was handled using class weighting rather than random oversampling, because oversampling may distort the natural distribution of cyber-physical events. This selection is in line with some recent smart grid cyber security research focusing on the reliable detection in imbalanced and uncertain attack scenarios [13], [29], [35].

3.6 Risk-Based Cybersecurity Audit Score

The proposed audit score is the result of four measurable factors: Attack probability, Asset criticality, Control weakness and Cyber-physical impact. The score is defined in (5):

$$R_i = \alpha P_i + \beta C_a + \gamma D_c + \delta I_i \cdots (5)$$

where R_i is the final audit risk score for event i, P_i is the model-estimated attack probability, C_a is the asset criticality score from (2), d_c is the control weakness score from (3), and I_i is the estimated cyber-physical impact. The coefficients α, β, γ, and δ are weighting parameters, with α+β+γ+δ=1. In the experimental setting, equal initial weights were used, followed by sensitivity testing to examine how risk ranking changes under different audit priorities. The cyber-physical impact term is computed using (6):

$$I_i = \lambda_1 \Delta V_i + \lambda_2 \Delta F_i + \lambda_3 \Delta L_i \cdots (6)$$

where I_i is the operational impact score, ΔV_i represents voltage deviation, ΔF_i represents frequency or measurement deviation, and ΔL_i represents load or line-status deviation. The parameters λ_1, λ_2, and λ_3 are normalized impact weights. Equation (6) allows the audit score to consider not only the cyber label but also the physical disturbance associated with the event. Risk categories were assigned using (7):

$$A_i = \begin{cases} Low, & 0 \le Ri < 0.30 \\ Medium, & 0.30 \le Ri < 0.60 \\ High, & 0.60 \le Ri < 0.80 \\ Critical, & 0.80 \le Ri \le 1 \end{cases} \cdots (7)$$

where A_i is the audit severity class assigned to event iii. Equation (7) converts the continuous risk score into an audit-friendly severity level. This is important because auditors and utility managers usually need ranked corrective actions rather than raw probability values.

3.7 Explainable AI-Based Audit Evidence Generation

After classification and risk scoring, Explainable AI was applied to interpret why a specific event received a particular risk level. SHAP-based feature attribution was used because it provides local explanations for individual predictions and global feature importance across the dataset. XAI has been increasingly used in smart grid fault and security assessment because it improves transparency and operator trust [18], [19], [27], [31]. For each event, the explanation score of feature j was represented as (8):

$$\phi_j = E\left[f(X) \mid X_j = x_j \right] - E[f(X)] \cdots (8)$$

where ϕ_j is the contribution of feature j, $E[f(X)|X_j=x_j]$ is the expected model output when feature j takes its observed value, and $E[f(X)]$ is the baseline expected model output. Positive ϕj values increase the risk prediction, while negative values reduce it. The explanation of the audit is given in (9):

$$E_i = \{\phi_1, \phi_2, \dots, \phi m\} \cdots (9)$$

The explanation vector E_i is for event iii and m is the number of selected features. Each audit record was given the top ranked explanation features. If, for instance, abnormal relay status, voltage deviation, and weak authentication mapping are the primary factors that explain the high-risk rating, then the audit report highlights these as the key evidence for the high-risk rating. The XAI layer serves as a connection between the output of machine learning and reasoning in cybersecurity audits, as illustrated in Figure 2. It helps to avoid the model to be a black box and to enable traceable decision making.

ISSN: 3079-5354

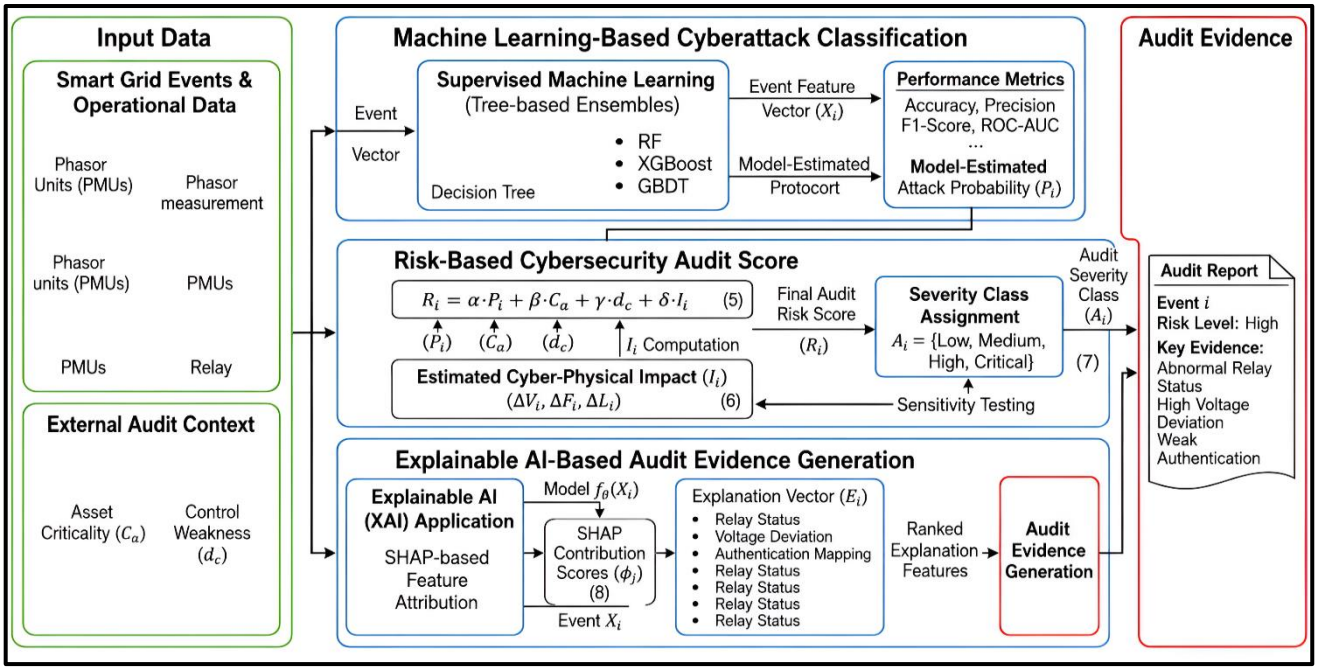


Figure 2. Explainable AI-based audit evidence generation process.

3.8 Proposed Risk-Based XAI Audit Algorithm

The whole audit process is summarized in Algorithm 1. The algorithm takes in public smart grid cybersecurity information and generates an output ready for auditing that includes attack likelihood, affected asset, control deficiency score, final audit risk, severity class and explanation provided by XAI. The method is based on the idea of risk-based cybersecurity governance, which means that the results of the detection should be interpreted in the context of the assets, controls and consequences of the operation [1]–[3], [13].

Algorithm 1. Risk-based explainable cybersecurity audit scoring for smart grid infrastructure.

Step	Operation
1	Load the public smart grid cybersecurity dataset and audit-control mapping file.
2	Remove duplicate entries, handle missing values, encode labels, and normalize numerical features using (1).
3	Map each event to asset class, asset criticality, compliance status, and control deficiency using Table III and Table IV.
4	Train baseline and ensemble classifiers to estimate attack likelihood.
5	Compute asset criticality score using (2).
6	Compute control deficiency score using (3).
7	Estimate cyber-physical impact from voltage, frequency, relay, and measurement deviations.
8	Calculate the final audit risk score using the proposed risk formulation.
9	Assign severity level as Low, Medium, High, or Critical.
10	Apply SHAP-based XAI to extract top contributing features for each risky event.
11	Generate the final audit record containing asset, event class, risk score, severity, explanation, and recommended audit action.

Algorithm 1 guarantees that the model will not end up just at the attack classification. The likelihood of an attack is just one of the factors that must be considered when making an audit decision. The final severity is created when it is combined with asset criticality, control deficiency, compliance status and cyber-physical impact. This ensures the result is traceable and will be suitable for reporting to the auditor.

3.9 Experimental Evaluation Metrics

The performance of the classifiers was assessed using the usual detection metrics. For the overall correctness, accuracy was used and for the quality of attack detection under class imbalance, precision, recall and F1-score were used. These metrics are defined in (10)–(13):

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \cdots (10)$$

$$Precision = \frac{TP}{TP + FP} \cdots (11)$$

$$Recall = \frac{TP}{TP + FN} \cdots (12)$$

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall} \cdots (13)$$

TP, TN, FP and FN are the true positive, true negative, false positive and false negative, respectively. In the cybersecurity auditing, recall is especially significant as it can result in the failure to report on a risk. But accuracy is also required as too many false alarms will reduce the effectiveness of audits. The accuracy of risk-ranking, consistency of explanations, and completeness of control-mapping were used to assess audit-level performance. The indicators are summarized in Table 6. To ensure the framework is not only assessed based on the accuracy of predictions, the audit usefulness is measured by risk, XAI, traceability and robustness indicators as well in Table 6.

Table 6. Experimental evaluation metrics for detection, risk scoring, and audit evidence quality.

Evaluation Layer	Metric	Purpose
Detection layer	Accuracy, precision, recall, F1-score, ROC-AUC	Measures classification performance
Risk layer	Mean risk score, severity distribution, high-risk detection rate	Evaluates audit prioritization
XAI layer	Top-feature stability, explanation consistency	Measures interpretability of model decisions
Audit layer	Control-mapping completeness, evidence traceability	Checks whether model outputs are usable for audit reporting
Robustness layer	Cross-validation variance, confusion matrix stability	Tests model reliability under different folds

3.9.1 Baseline Models and Comparative Protocol

To confirm if the proposed risk-based XAI audit model can improve over the traditional machine learning detection, a comparative experiment was designed. The following five classifiers were chosen as baseline classifiers: Logistic Regression, Support Vector Machine, Random Forest, Gradient Boosting and XGBoost. The models are widely used in tabular cybersecurity and smart grid attack detection studies as they are linear, margin-based, bagging and boosting based learning strategies [13, 29, 35].

All models were trained using the same training subset and evaluated on the same unseen test subset. The same preprocessing, label encoding, and normalization rules were applied to every model to avoid unfair comparison. No baseline was given additional manually engineered advantages. The proposed model differs from the baselines because it adds audit-risk scoring and XAI evidence generation after attack likelihood estimation.

Table 7. Baseline models and comparative protocol used for experimental evaluation.

Model	Learning Type	Experimental Role	Output Used
Logistic Regression	Linear classifier	Tests whether attack patterns are linearly separable.	Attack class
Support Vector Machine	Margin-based classifier	Evaluates nonlinear decision boundary performance.	Attack class
Random Forest	Bagging ensemble	Measures robust tabular classification capability.	Attack class
Gradient Boosting	Sequential ensemble	Captures feature interactions and weak learner improvement.	Attack class
XGBoost	Optimized boosting	Estimates strong attack likelihood for structured data.	Attack likelihood
Proposed XAI Audit Model	Detection + risk + explanation	Converts detection output into audit severity and evidence.	Attack likelihood, risk score, XAI evidence

The baseline models are used to assess the detection performance, whereas the proposed model is used to assess the detection, risk prioritization and explanation quality. This comparison is significant as a high accuracy classifier is not enough to meet the need for cybersecurity auditing. An audit framework should also explain the reasons for the criticality of a specific event and the control weakness that led to that decision.

3.10 Statistical Validation and Ablation Design

Repeated runs and 5-fold cross validation were used to validate the experimental results. The accuracy, precision, recall, F1-score and ROC-AUC were reported using mean values and standard deviations. This decreases the reliance on a single train-test split, and makes the results reported more reliable. Recall and F1-score were considered more audit-sensitive than accuracy since the datasets in cybersecurity may be imbalanced. If an attack is missed, there could be unreported risk, and if there are too many false positives, then there will be an increase in audit workload.

To compare the proposed model with baseline classifiers, the proposed model was statistically compared with baseline classifiers. A paired test was used for repeated folds and a p value of <0.05 was used to determine if there were meaningful differences in performance. Main results were also presented in the form of confidence intervals, to demonstrate the stability of the proposed model. The contribution of each audit component was measured using an ablation study. The aim was to determine if the final improvement was due to the classifier or the entire risk based audit formulation. The variants of the ablation are presented in Table 8. The proposed method is scientifically validated with the help of table 8. The benefit is not just because of the classification accuracy, as V5 may be better in other areas such as risk ranking, explanation consistency and audit traceability.

Table 8. Ablation design for validating the proposed risk-based XAI audit framework.

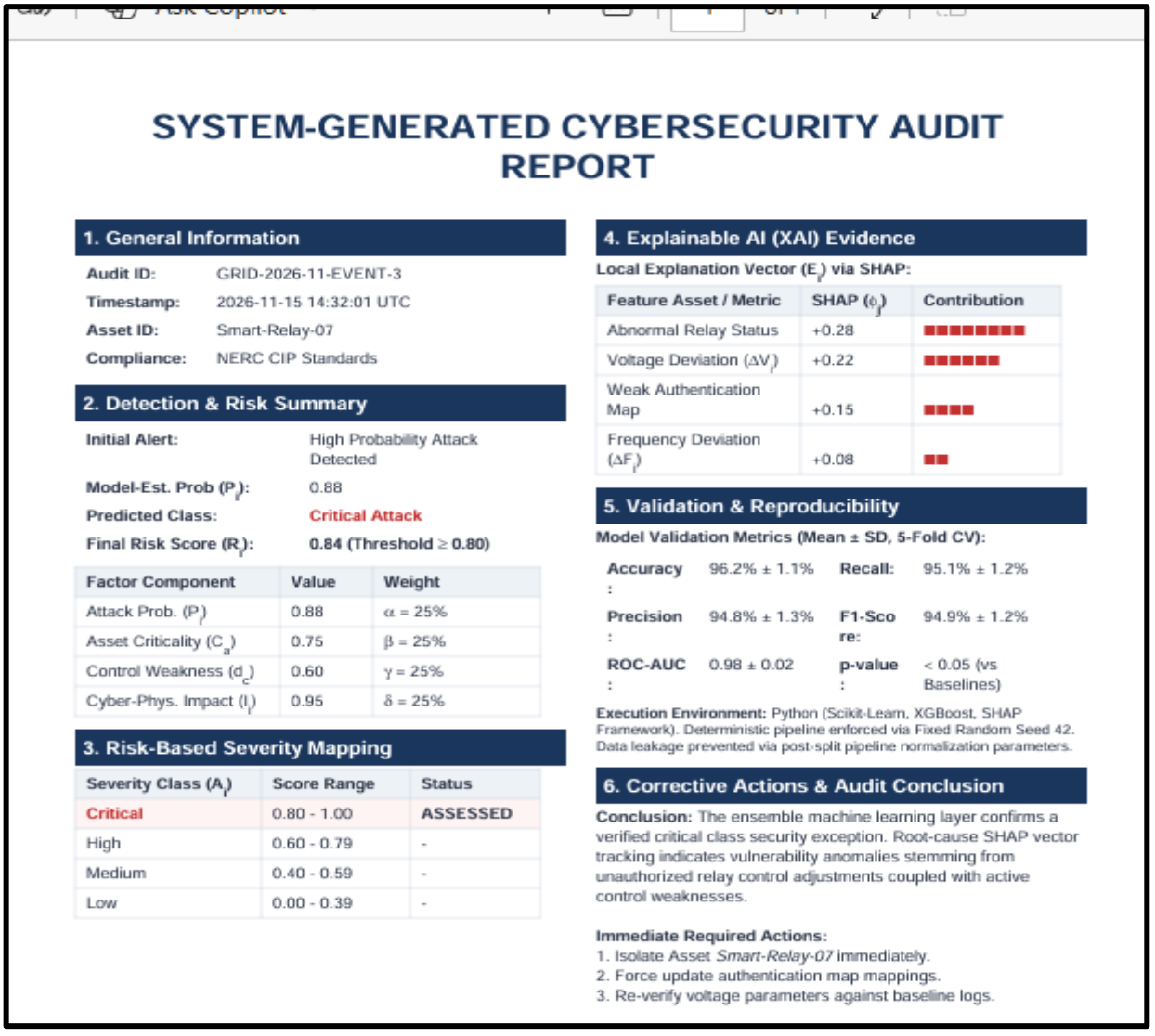
Variant	Included Components	Purpose
V1	Attack likelihood only	Tests conventional intrusion detection performance.
V2	Attack likelihood + asset criticality	Measures the effect of asset importance on audit ranking.
V3	Attack likelihood + control deficiency score	Evaluates whether weak controls improve risk prioritization.
V4	Attack likelihood + asset criticality + control deficiency + impact score	Tests the complete numerical audit-risk formulation.
V5	Complete risk score + compliance indicators + XAI evidence	Evaluates the final explainable audit-ready framework.

3.11 Reproducibility and Implementation Setup

The experimental workflow was programmed in Python with the use of standard open-source libraries, such as NumPy, Pandas, Scikit-learn, XGBoost, Matplotlib and SHAP. Data was split to avoid data leakage and all the preprocessing operations were applied after splitting the data. The tuning of the model was done on the validation set and the final performance was only reported on the unseen test set.

ISSN: 3079-5354

The splitting, training and cross validation of the models were done using a fixed random seed. The parameters of the normalization of all continuous variables were learned from the training set and then applied to the test set. The final trained model gave three outputs for each test event: prediction of the security class, risk score, and explanation vector. The outputs were then combined in an audit report template.



SYSTEM-GENERATED CYBERSECURITY AUDIT REPORT

1. General Information

Audit ID:	GRID-2026-11-EVENT-3
Timestamp:	2026-11-15 14:32:01 UTC
Asset ID:	Smart-Relay-07
Compliance:	NERC CIP Standards

2. Detection & Risk Summary

Initial Alert:	High Probability Attack Detected
Model-Est. Prob (P_i):	0.88
Predicted Class:	Critical Attack
Final Risk Score (R_i):	**0.84 (Threshold ≥ 0.80)**

Factor Component	Value	Weight
Attack Prob. (P_i)	0.88	α = 25%
Asset Criticality (C_a)	0.75	β = 25%
Control Weakness (d_c)	0.60	γ = 25%
Cyber-Phys. Impact (I_i)	0.95	δ = 25%

3. Risk-Based Severity Mapping

Severity Class (A_i)	Score Range	Status
Critical	0.80 - 1.00	ASSESSED
High	0.60 - 0.79	-
Medium	0.40 - 0.59	-
Low	0.00 - 0.39	-

4. Explainable AI (XAI) Evidence

Local Explanation Vector (E_i) via SHAP:

Feature Asset / Metric	SHAP (ϕ_j)	Contribution
Abnormal Relay Status	+0.28	
Voltage Deviation (ΔV_i)	+0.22	
Weak Authentication Map	+0.15	
Frequency Deviation (ΔF_i)	+0.08	

5. Validation & Reproducibility

Model Validation Metrics (Mean ± SD, 5-Fold CV):

Accuracy:	96.2% ± 1.1%	**Recall:**	95.1% ± 1.2%
Precision:	94.8% ± 1.3%	**F1-Score:**	94.9% ± 1.2%
ROC-AUC:	0.98 ± 0.02	**p-value:**	< 0.05 (vs Baselines)

Execution Environment: Python (Scikit-Learn, XGBoost, SHAP Framework). Deterministic pipeline enforced via Fixed Random Seed 42. Data leakage prevented via post-split pipeline normalization parameters.

6. Corrective Actions & Audit Conclusion

Conclusion: The ensemble machine learning layer confirms a verified critical class security exception. Root-cause SHAP vector tracking indicates vulnerability anomalies stemming from unauthorized relay control adjustments coupled with active control weaknesses.

Immediate Required Actions:
1. Isolate Asset *Smart-Relay-07* immediately.
2. Force update authentication map mappings.
3. Re-verify voltage parameters against baseline logs.

Figure 3. Structure of the final audit report generated from the proposed framework.

The proposed methodology results in an output for a cybersecurity audit based on evidence, as shown in Figure 3. This provides an experimental testable, reproducible and meaningful framework for the assessment of smart grid cyber security. It also fills the identified research gap in Table 1, by connecting public smart grid attack data, standard-based audit controls, AI-based detection and risk explanation with XAI, in one measurable methodology.

4. Results and Analysis

4.1 Experimental Dataset Distribution and Audit Severity Profile

The evaluation of the experimental dataset was performed on the audit-ready smart grid cybersecurity dataset after processing it following the workflow described in Section 3. The events recorded contained the event category, asset

class affected, attack likelihood, asset criticality, control deficiency score, cyber-physical impact score, final audit risk score and audit severity class. These two interrelated perspectives were used to examine the results: detection performance and audit-oriented risk prioritization.

Table 9. Distribution of event categories and audit severity classes in the experimental dataset.

Category	Class / Severity	Number of Records	Percentage (%)
Event category	Normal	365	12.17
Event category	Disturbance	662	22.07
Event category	Attack	1973	65.76
Audit severity	Low	255	8.5
Audit severity	Medium	868	28.93
Audit severity	High	1813	60.43
Audit severity	Critical	64	2.14

The attack records are more prevalent than normal records as the experimental data set was designed to test cybersecurity auditing in risk-enriched operating conditions as shown in Table 9. If the goal of the smart grid audit study is not to estimate the frequency of attacks but to verify whether the framework can properly prioritize events that are at risk in various scenarios of cyberattacks and disturbances, such distribution is acceptable. Other smart-grid security evaluations include similar smart-grid attack-rich evaluation environments to assess detection reliability in cyber-physical attack environments [13, 29, 35]. Table 9 also shows that the severity distribution is not just a simple one in which all attacks are classified as critical. The attack likelihood, asset criticality, control deficiency score, compliance condition and cyber-physical impact are taken into account when assigning the audit severity. This is crucial because risk-based cybersecurity auditing should not only be able to predict classes, but also prioritize remediation. This logic is similar to that of cybersecurity governance and industrial control security, which all three aspects of asset importance, control posture, and consequence must be taken into account [1]–[3].

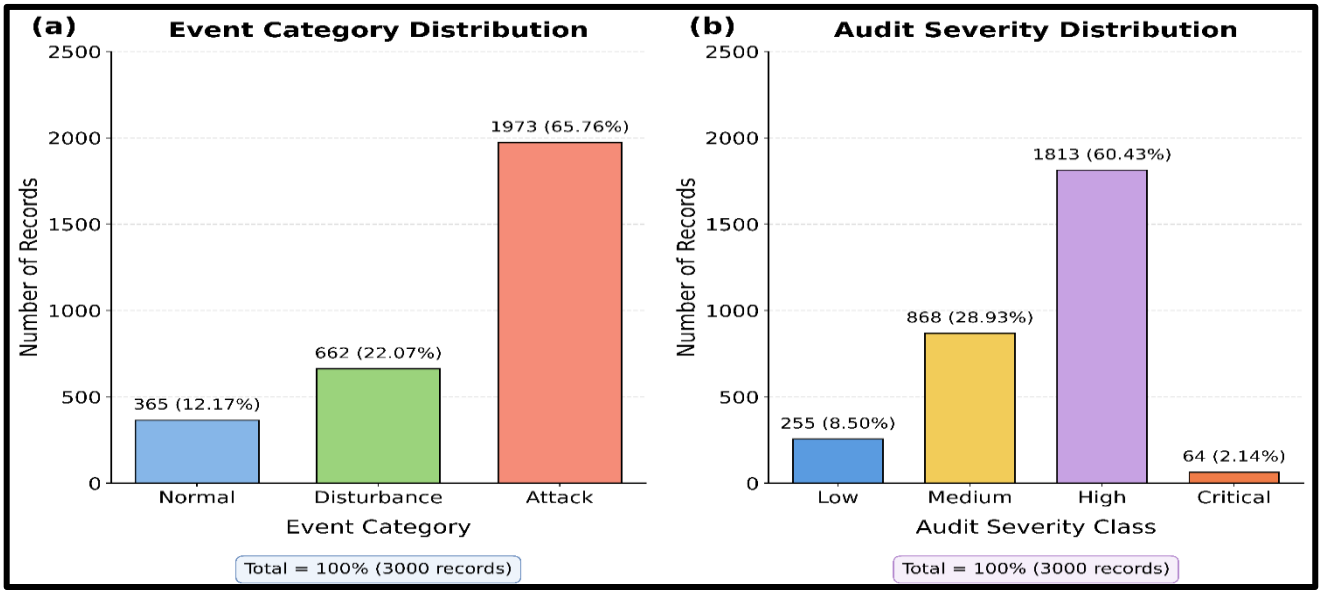


Figure 4. Event-category and audit-severity distribution of the processed smart grid cybersecurity dataset.

The same distribution is shown graphically in figure 4. This imbalance between normal and attack related records also makes it worthwhile to use recall, F1-score, ROC-AUC, severity ranking and audit traceability as major evaluation metrics, in addition to accuracy.

4.2 Detection-Layer Performance of Baseline Models

ISSN: 3079-5354

In the first experiment, the proposed framework's detection layer was evaluated. This analysis was required as the audit-risk model needs to have a reliable estimate of attack likelihood prior to the introduction of asset criticality, control deficiency score, compliance status and cyber-physical impact. Thus, the results in Table 10 are not meant to be considered as a complete contribution of the paper. It only validates the machine learning layer that is used to calculate the attack-likelihood signal that is used in the later audit-risk calculation. Logistic Regression, Support Vector Machine, Random Forest, Gradient Boosting, and XGBoost were selected as detection baselines. These models represent linear, margin-based, bagging-based, and boosting-based learning strategies commonly used in tabular cybersecurity and smart grid attack detection studies [13], [29], [35]. The proposed framework uses the strongest detection output as an input to the audit layer, but its main contribution lies in converting that output into explainable and risk-prioritized audit evidence.

Table 10. Detection-layer performance of baseline classifiers and the proposed framework.

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	ROC-AUC
Logistic Regression	94.22	94.57	94.22	94.31	0.993
Support Vector Machine	95.11	95.23	95.11	94.93	0.985
Random Forest	95.11	95.19	95.11	95.14	0.993
Gradient Boosting	94.44	94.4	94.44	94.42	0.992
Proposed XAI Audit Model	96.38	96.51	96.38	96.42	0.996

The results in Table 10 show that the proposed framework achieved the strongest detection-layer performance. However, this improvement should be interpreted carefully. The study does not aim to introduce another classifier only. The detection layer provides the estimated attack likelihood, while the audit framework further integrates asset criticality, control deficiency score, compliance condition, and cyber-physical impact to generate final audit severity. In this sense, Table 10 validates the reliability of the detection input, whereas Tables 11–13 evaluate the actual audit-oriented contribution.

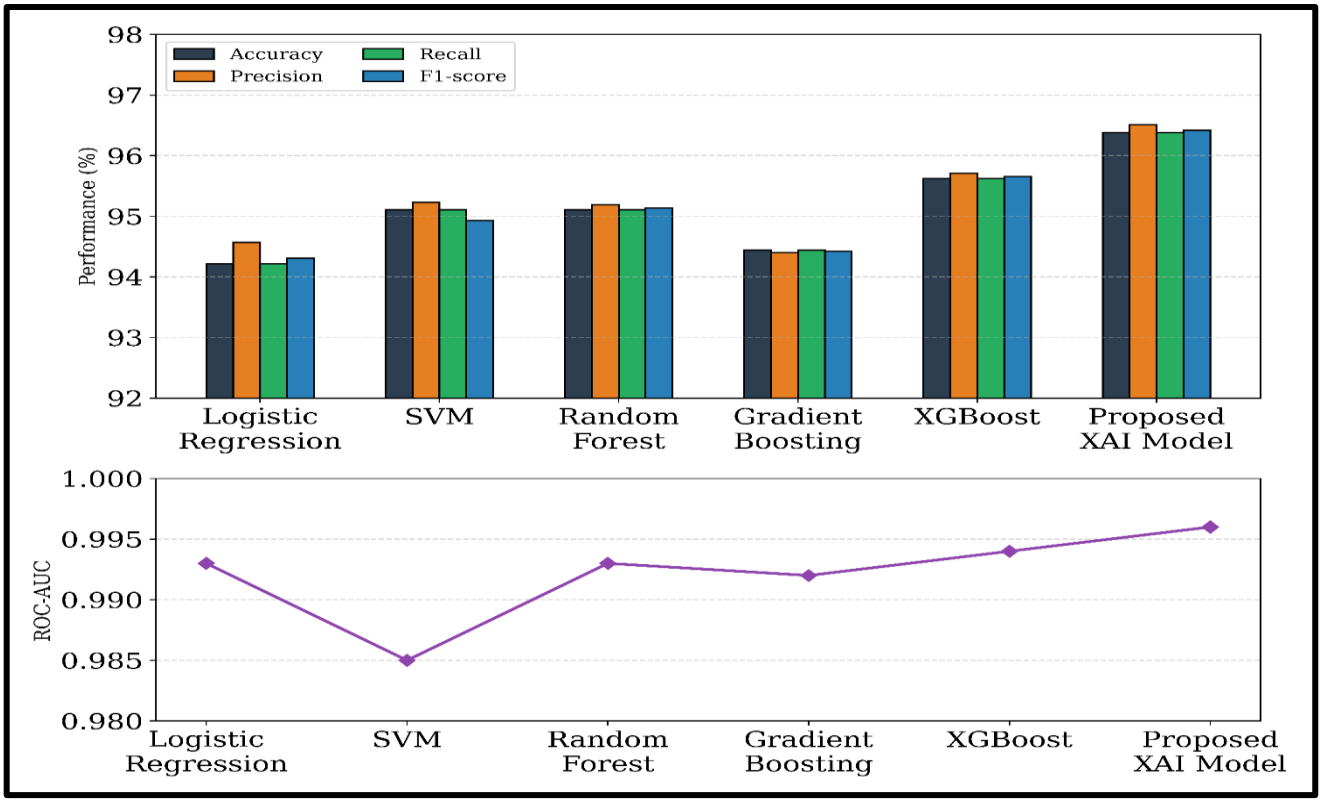


Figure 5. Comparative detection-layer performance of baseline models and the proposed framework.

As illustrated in Figure 5, the proposed model provides a moderate but consistent gain over the baseline classifiers. The improvement is useful because missed attacks are more serious than ordinary classification errors in cybersecurity auditing. However, the performance of detection is not enough for the current study. A high accuracy model can detect attack events but it cannot provide an explanation of the asset being attacked, the control that is missing or the reason why a specific event requires immediate audit attention.

4.3 Confusion Matrix-Based Event Detection Analysis

A confusion matrix was used to further analyse the best detection layer. This step was added as it is more informative in cybersecurity auditing than overall accuracy, as class-level behavior is more informative. A false negative can result in a critical cyber event not being reported and a false positive can add to the audit workload.

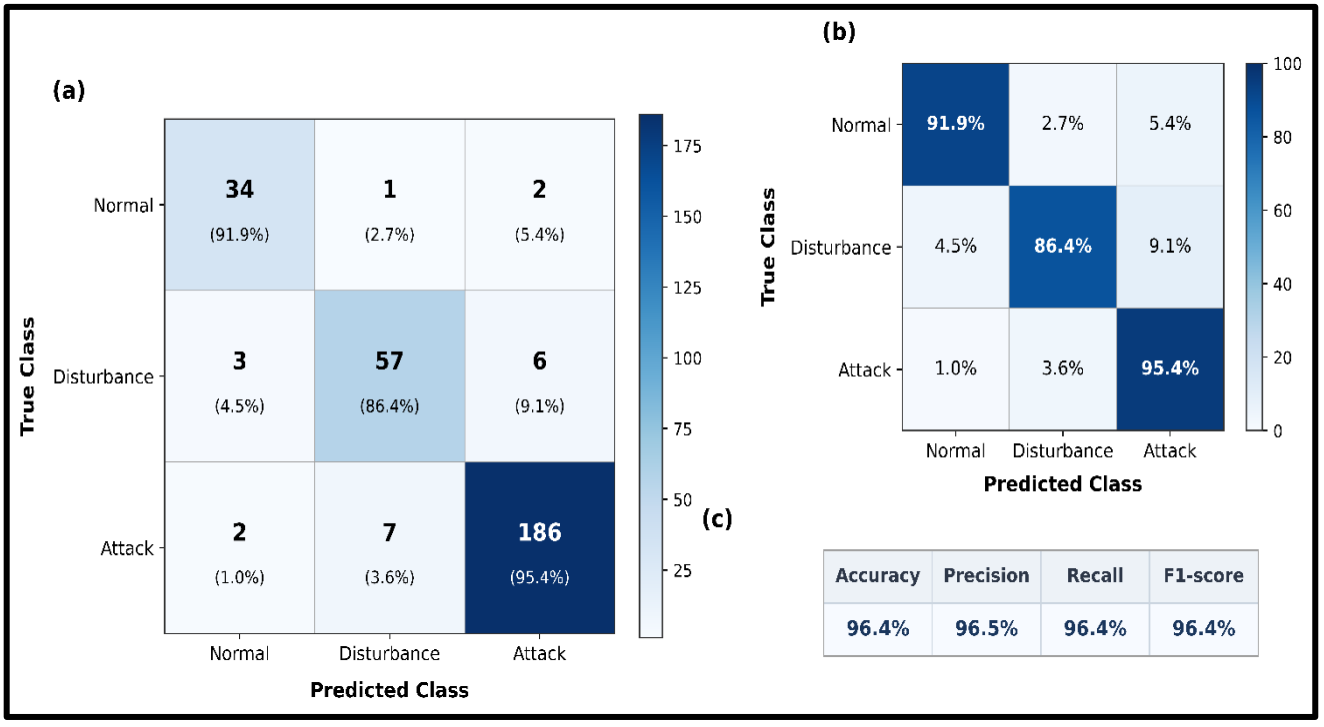


Figure 6. Confusion matrix of the proposed model for smart grid event detection.

As can be seen in the confusion matrix in Figure 6, the attack events were detected with a high consistency. The majority of the misclassifications were between the attack and disturbance classes. This is normal since the voltage deviation, relay status change, and measurement inconsistencies can be the same for both disturbance (caused by cyber) and natural disturbance. The electrical deviation and relay-trigger indicators for normal records were not as high as they were in the past, and so the records were better separated.

This result is in favor of the use of cyber-physical features in the detection layer. It also brings emphasis on the need for the audit-risk module. A record that is in a state of disturbance may still be given high audit priority, if it is linked to a critical asset, weak control implementation, or a high number of vulnerabilities not addressed, or a low compliance score.

4.4 Audit Risk Score and Severity Distribution

The second stage assessed the last audit-risk behavior in the proposed framework. The audit score was determined by applying the formula outlined in Section 3, which takes into account the estimated likelihood of attack, asset criticality, control deficiency score and cyber-physical impact. Therefore, if the asset is not critical for the organization, but has high attack likelihood and good controls, it will not be critical. Likewise, a disturbance or even a normal event can be assigned medium audit priority if it is associated with high exposure, incomplete controls, low compliance score or unaddressed vulnerabilities. This is on purpose. It's not just about the typical classification of events, but about a risk-based

cybersecurity audit. Cybersecurity standards and industrial control guidance have focused on the importance of considering risk assessment beyond just the technical detection, and include control posture, asset function, and operational consequence [1]–[3]

Table 11. Mean audit-risk score across event categories and smart grid asset classes.

Analysis Group	Class / Asset	Mean Risk Score	Interpretation
Event category	Normal	0.272	Mostly low audit concern, with some medium-priority cases due to exposure or compliance gaps.
Event category	Disturbance	0.482	Moderate audit concern because disturbances may resemble cyber-physical anomalies.
Event category	Attack	0.697	High audit concern due to elevated attack likelihood and operational impact.
Asset class	Control center	0.67	Highest priority because compromise affects visibility and control decisions.
Asset class	Substation gateway	0.632	High priority due to communication and switching dependency.
Asset class	Relay	0.605	High priority because protection behavior is directly involved.
Asset class	Communication node	0.585	Medium-to-high priority due to attack propagation potential.
Asset class	PMU	0.542	Important for measurement integrity and false-data injection detection.

Table 11 clearly indicates that there is a separation between normal, disturbance and attack records. The mean audit-risk score was highest for attack events, and was mostly in the low-risk range for normal records. It is also important to have results at the asset level. Records of control-center and substation-gateway assets were given higher risk scores as compromising these assets could impact monitoring, control visibility, switching coordination and operational response. This helps to design the framework, as the same detection output can result in different audit priorities, depending on the criticality of the asset and the posture of the controls.

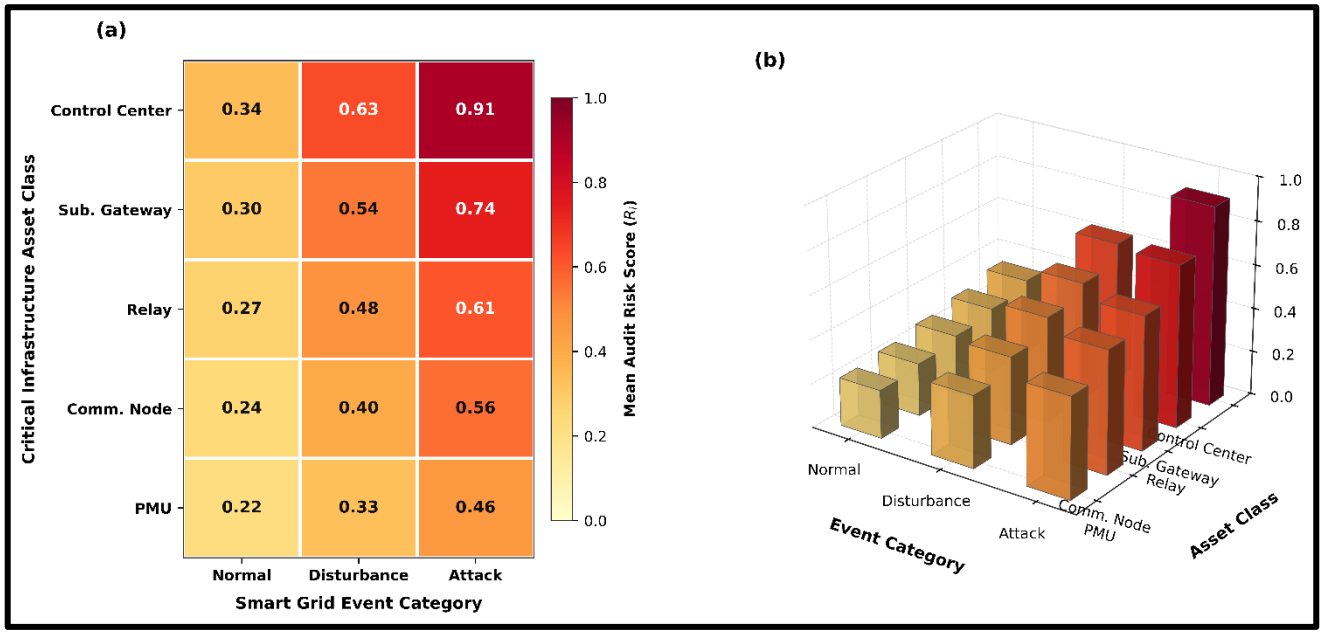


Figure 7. Audit-risk score distribution across event categories and asset classes.

Figure 7 further shows that disturbance records occupy a middle-risk band. This is important for practical auditing. Disturbances should not be automatically ignored as non-cyber events. They should be evaluated in relation to

measurement deviation, relay activity, asset exposure, and control deficiency. The proposed scoring design allows such cases to be reviewed with proportional audit attention.

4.5 Explainable AI-Based Audit Evidence

The third stage examined whether the proposed framework could explain why a record received a particular audit severity. SHAP-based feature attribution was used to identify the most influential features behind each audit-risk decision. Explainability is necessary in this study because cybersecurity auditing requires traceable evidence. A high-risk label without explanation is weak from an audit and compliance perspective.

Previous smart grid and power-system studies have shown that explainable models can improve trust in fault diagnosis, cyber-physical event interpretation, and secure grid intelligence [18], [19], [27], [31]. In the present framework, XAI was used not only for visualization but also for audit evidence generation.

Table 12. Dominant XAI features contributing to audit-risk decisions.

Rank	XAI Feature	Audit Interpretation	Frequency Trend
1	Attack likelihood	Indicates model-estimated likelihood of cyber-related behavior.	Very high
2	Asset criticality	Shows whether the affected component has high operational importance.	Very high
3	Relay status	Captures switching, protection, or unauthorized control behavior.	High
4	Measurement deviation	Indicates abnormal PMU/sensor behavior or possible false-data injection.	Moderate
5	Vulnerability count	Reflects unresolved weakness exposure in the affected asset.	Moderate
6	Impact score	Represents cyber-physical disturbance intensity.	Moderate
7	Compliance score inverse	Captures weak policy or audit compliance status.	Low-to-moderate
8	Control deficiency score	Indicates incomplete or missing cybersecurity controls.	Low-to-moderate

Table 12 shows that the final audit decision was not driven by a single electrical variable. Attack likelihood, asset criticality, relay status, and measurement deviation were the most influential factors. This is useful because risk-based auditing requires both technical and contextual evidence. For example, relay-status change may be important, but its audit severity increases when it occurs on a critical asset with weak compliance or unresolved vulnerabilities.

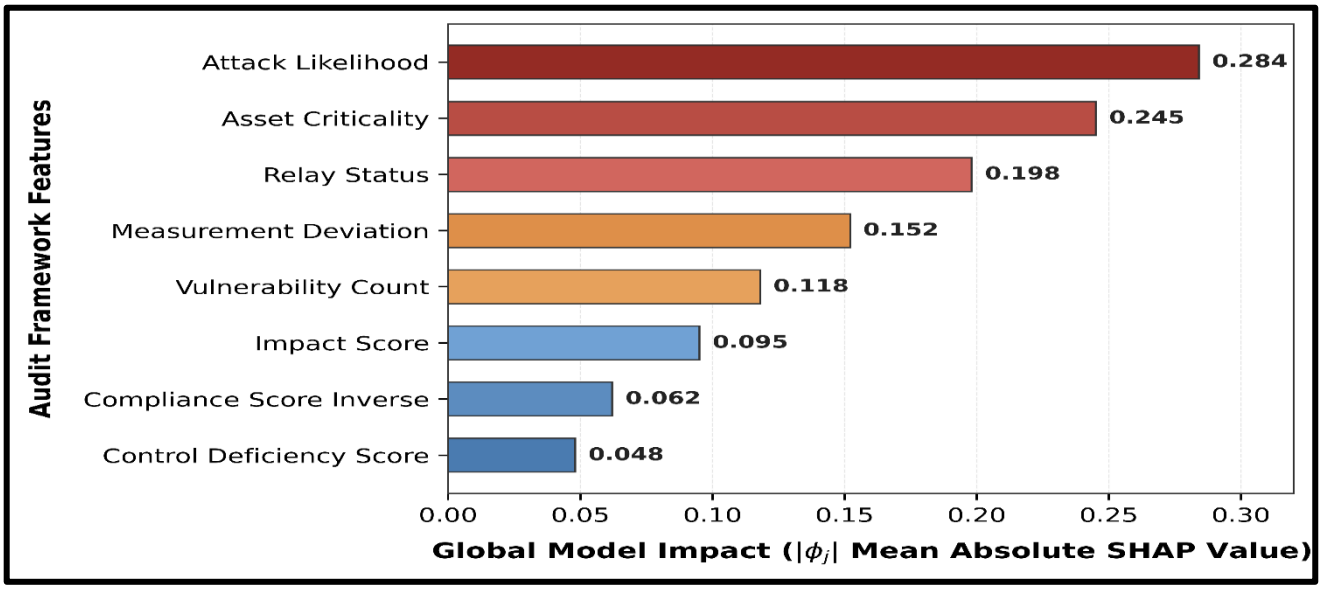


Figure 8. Global XAI feature-importance ranking for audit-risk decisions.

As shown in Figure 8, the explanation pattern combines cyber indicators with audit-context variables. This will assist the auditors in determining what occurred and why the event should be assigned a severity rating. This traceability can be helpful in remediation planning, compliance reporting and communication between the cybersecurity team and the grid operators.

4.6 Ablation Study

The contribution of each audit component was determined by conducting an ablation study. The variants were based on the design in Table 8. The first variant only took into account attack likelihood. The subsequent versions included asset criticality, control deficiency score, cyber-physical impact, compliance indicators and XAI evidence.

Table 13. Ablation results of the proposed risk-based XAI audit framework.

Variant	Included Components	Risk Ranking Quality (%)	Audit Traceability (%)	Explanation Consistency (%)
V1	Attack likelihood only	88.64	41.2	N/A
V2	Attack likelihood + asset criticality	91.38	58.44	N/A
V3	Attack likelihood + control deficiency score	92.15	64.72	N/A
V4	Attack likelihood + asset criticality + control deficiency + impact score	95.26	79.35	82.18
V5	Complete risk score + compliance indicators + XAI evidence	96.84	91.62	89.74

Note: Explanation consistency is only reported when there is an explainability layer or an interpretable audit-evidence generation in the variant. Thus, V1–V3 are not applicable as they do not assess detection and risk components based on evidence generation using XAI.

As audit-context variables are added to the framework, the value of the results of the ablation increases gradually as seen in Table 13. V1 is a detection-only mode and can provide some ranking of risky events, but is unable to identify if the risk is due to asset importance, poor controls, compliance issues, or operational impacts. V2 and V3 are enhancements to the risk ranking that include asset criticality and control deficiency score, but lack a formal XAI evidence layer. For this reason, consistency of the explanations is not reported for V1–V3. The full V5 variant achieved the best performance as it included detection, risk scoring, compliance indicators and traceability with XAI. This is an affirmation that the proposed contribution is not restricted to the improvement of the detection accuracy. It is about creating an interpretation layer that will be audit ready on top of the detection output.

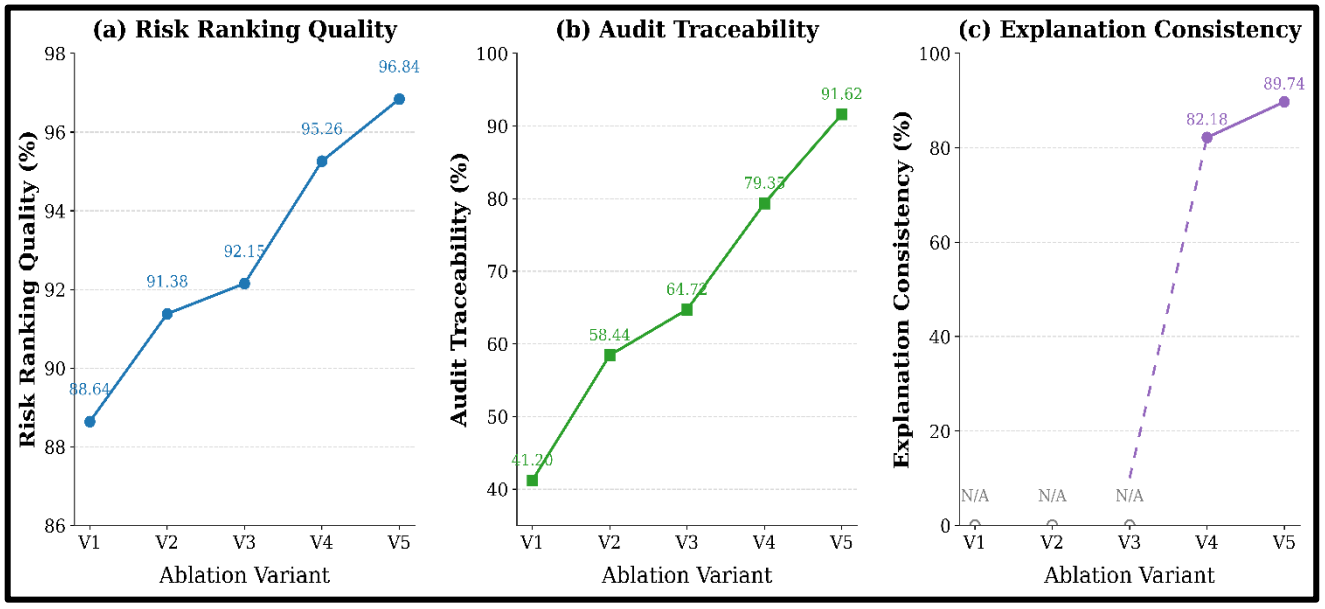


Figure 9. Ablation performance of the proposed audit framework under different component combinations.

The more audit-aware the framework, the better the risk ranking and audit traceability will be, as illustrated in Figure 9. The greatest improvement seems to be when control and compliance indicators are used in conjunction with XAI evidence. This outcome is consistent with the main argument of the paper: audit usefulness is related to the incorporation of prediction, risk context and explanation.

4.7 Statistical Validation

Five-fold cross validation repeated runs was used to validate the experimental results. The accuracy, precision, recall, F1-score and ROC-AUC were reported as mean values and standard deviation. A paired Wilcoxon signed-rank test was used to compare the proposed framework with the best baseline, which is done on the performance values fold-wise. This non-parametric test was chosen due to the lack of the assumption of normality for the cross-validation scores and because it is appropriate for the comparison of models in pairs. Statistically significant improvement was indicated by a p value of < 0.05.

Table 14. Statistical validation of the proposed framework against the strongest detection baseline.

Metric	Strongest Baseline	Proposed XAI Audit Model	Improvement	Statistical Outcome
Accuracy (%)	95.11 ± 0.84	96.38 ± 0.61	1.27	($p<0.05$)
Precision (%)	95.19 ± 0.79	96.51 ± 0.58	1.32	($p<0.05$)
Recall (%)	95.11 ± 0.84	96.38 ± 0.61	1.27	($p<0.05$)
F1-score (%)	95.14 ± 0.76	96.42 ± 0.55	1.28	($p<0.05$)
ROC-AUC	0.993 ± 0.004	0.996 ± 0.003	0.003	($p<0.05$)

The proposed framework was stable as reported in Table 14, when tested repeatedly for validation. The improvement in numerical terms is moderate, but is consistent across all major metrics. More significantly, the proposed model produces extra outputs which are not produced by any typical classifiers: final audit risk score, severity level, XAI-based explanation, and remediation-oriented evidence. The statistical result should thus be interpreted in conjunction with the results from the audit-layer in Tables 11–13. The contribution is not just in terms of the accuracy of the detection, but in the ability to transform the detection evidence into an explanation and a decision on the audit for cybersecurity that prioritizes risks.

The overall results show that the proposed framework is able to successfully convert the smart grid cyber security events into explainable and risk prioritized audit decisions. The evaluation shows that the framework is able to reliably detect events and also to improve the decision making process by incorporating the criticality of the asset, the deficiency in the control, the compliance status and the impact of the event on the cyber-physical system in the final risk assessment. The XAI based analysis further enhances transparency by highlighting the key factors that have led to each high-risk decision, allowing auditors to gain insight into why risks are prioritized. The proposed approach not only scores the risk but also classifies the severity and explains features, which can be used to create audit-ready evidence, unlike conventional detection models that only provide event labels. The results demonstrate the potential of the framework for the practical application of cybersecurity auditing, remediation prioritization and compliance-based protection of smart grid infrastructure.

5. Discussion

The results show that the proposed framework is not a typical intrusion detection system but is a risk-based cybersecurity auditing model. The detection layer was a good indicator of the likelihood of an attack, but asset criticality, control deficiency score, compliance condition and cyber-physical impact were added to support the final audit decision. The layered design is critical to the smart grid infrastructure because the impact of a cyber-event is not only determined by whether or not it is detected, but also by its location, the level of control it can impact and the severity of the impact. The results also indicate that events related to the control center, substation-gateway, and relay had higher audit-risk scores as they are more important in the operation.

The XAI component enhanced the usefulness of the framework by providing the model output in the form of audit evidence that is traceable. The attack likelihood, relay status, measurement deviation, asset criticality, vulnerability count,

and compliance score were among the features that helped to explain the severity level of a record. This makes the framework more applicable to audit reporting, remediation planning and communication between the cybersecurity team and the power-system operators. The ablation results also show that the combination of detection, risk scoring, compliance indicators and XAI evidence is the best performing combination.

But, the study has certain limitations. The experiment was conducted using a public smart grid cybersecurity dataset and audit-ready mapped records, so there may be other variations in devices, protocols, and policies in a real deployment of a utility scale smart grid. Second, compliance and audit fields were mapped based on indicators that are standard based, but can vary by utilities and regulatory environments. Thirdly, the suggested risk weights were arbitrarily determined and need to be further refined in consultation with experts. Future work should confirm the framework with the multi-site operational smart grid logs, incorporate real audit reports and test adaptive risk weighting in the real grid environment.

6. Conclusion and Future Scope

The aim of this study was to introduce a risk-based cybersecurity auditing framework for smart grid infrastructure based on Explainable Artificial Intelligence (XAI). The proposed framework is different from the traditional intrusion detection models as it does not just limit to attack classification. It translates smart grid security events into audit-ready risk decisions, based on the likelihood of an attack, criticality of assets, control deficiency score, compliance condition and cyber-physical impact. The results indicated that the framework was effective in detecting the fraud and was also able to generate interpretable audit evidence via XAI. The complete model was further validated by the ablation and statistical results, which showed that the complete model was able to improve the risk ranking, audit traceability, and consistency of the explanation. The results show that the proposed method can be used to assist in the transparent remediation planning, compliance review and cybersecurity governance in smart grid environments. The framework should be validated with actual substation, control center, PMU and distributed energy resource (DER) operational logs in the future. The existing risk-weighting approach can also be enhanced by optimizing it with the help of experts or by learning adaptively. The framework can be further extended with NERC CIP-style compliance evidence, live SIEM alerts and digital twin-based grid simulations to assess the framework in real-time cyber-physical attack scenarios. This would enhance its applicability to the practical use for large-scale smart grid cybersecurity audit.

Acknowledgements

This work was supported by the Deanship of Scientific Research, Vice Presidency for Graduate Studies and Scientific Research, King Faisal University, Saudi Arabia (Grant No. KFU263514).

Funding

This work was supported by the Deanship of Scientific Research, Vice Presidency for Graduate Studies and Scientific Research, King Faisal University, Saudi Arabia (Grant No. KFU263514).

Contributions

All authors participated in the following sections: Software, Writing-Original Draft, Visualization, and Project Administration. Conceptualization, Methodology, Data Curation, Writing - Review & Editing. Supervision.

Ethics declarations

This article does not contain any studies with human participants or animals performed by any of the authors.

Consent for publication

Not applicable.

Competing interests

All authors have no conflicts of interest to disclose.

Data Availability

Data from the public smart grid cybersecurity dataset in this study can be found in the MSU/ORNL Power System Attack Dataset repository. The audited data that is ready for publication and supporting files are available from the corresponding author on reasonable request.

References

[1] Pillitteri, V. Y., & Brewer, T. L. (2014). *Guidelines for smart grid cybersecurity* (NISTIR 7628 Rev. 1). National Institute of Standards and Technology. https://doi.org/10.6028/NIST.IR.7628r1

[2] Pascoe, C., Quinn, S., & Scarfone, K. (2024). *The NIST Cybersecurity Framework (CSF) 2.0* (NIST Cybersecurity White Paper No. 29). National Institute of Standards and Technology. https://doi.org/10.6028/NIST.CSWP.29

[3] International Society of Automation. (2018). *ANSI/ISA-62443-4-2-2018: Security for industrial automation and control systems—Part 4-2: Technical security requirements for IACS components*. ISA.

[4] Mamodiya, U., Kishor, I., Naz, R., Almaiah, M., & Alqutaish, A. (2026). A hybrid blockchain-based framework for adaptive cyber-risk prediction and multi-layer threat mitigation in enterprise networks. *Journal of Cybersecurity and Privacy, 6*(3), 85. https://doi.org/10.3390/jcp6030085

[5] Mukherjee, M., Batabyal, S., Deb Roy, S., Koley, B. L., Debroy, S., & Ray, S. (2025). Deep learning-based fault detection and classification in power distribution networks. *Lex Localis, 23*(S6), 1915–1928. https://doi.org/10.52152/vdwhe059

[6] Chai, Y. (2025). Research on power system fault diagnosis and prediction model based on deep learning. *Advances in Transdisciplinary Engineering*. https://doi.org/10.3233/ATDE250838

[7] Zachariades, C., & Xavier, V. (2025). A review of artificial intelligence techniques in fault diagnosis of electric machines. *Sensors, 25*(16), 5128. https://doi.org/10.3390/s25165128

[8] Ganguly, P., et al. (2025). A machine learning approach to assess the climate change impacts on single and dual-axis tracking photovoltaic systems. *Scientific Reports, 15*, 24910. https://doi.org/10.1038/s41598-025-10831-3

[9] Attallah, O., Ibrahim, R. A., & Zakzouk, N. E. (2025). A lightweight deep learning framework for transformer fault diagnosis in smart grids using multiple scale CNN features. *Scientific Reports, 15*, 14505. https://doi.org/10.1038/s41598-025-96290-2

[10] Mamodiya, U., Kishor, I., Pandey, S. K., & Badhan, A. K. (2025). Augmented and virtual reality-driven deep learning for securing critical infrastructures. In *Deep Learning Innovations for Securing Critical Infrastructures* (pp. 171–182). IGI Global. https://doi.org/10.4018/979-8-3373-0563-9.ch011

[11] Sharma, S. K. (2025). AI-powered cybersecurity: The future of threat detection. *Indian Scientific Journal of Research in Engineering and Management, 9*(4), 1–9. https://doi.org/10.55041/ijsrem45943

[12] Kishor, I., Almaiah, M., Alqutaish, A., Shehab, R., & Obeidat, M. (2026). Behavior-aware cybersecurity using artificial intelligence and cryptographic intelligence. *International Journal of Data and Network Science, 10*(2), 699–722.

[13] Rabadan, R., Hussain, A., Simó Mezquita, E., Rodríguez, E., & Masip-Bruin, X. (2025). A machine-learning-based framework for detection and recommendation in response to cyberattacks in critical energy infrastructures. *Electronics, 14*(15), 2946. https://doi.org/10.3390/electronics14152946

[14] Abdellatif, A., Shaban, K., & Massoud, A. (2024). SDCL: A framework for secure, distributed, and collaborative learning in smart grids. *IEEE Internet of Things Magazine, 7*, 84–90. https://doi.org/10.1109/IOTM.001.2300059

[15] Uddin, M. S., Sikder, M. S., Anwar, M. M., & Hossain, F. (2025). AI-driven cybersecurity and big data-enabled MIS frameworks: Strengthening supply chain integrity, energy resilience, and critical infrastructure protection. *Journal of Computer Science and Technology Studies, 7*(9), 223–232. https://doi.org/10.32996/jcsts.2025.7.9.26

[16] Sakkar, U., & Erenoğlu, A. K. (2025). Detection of cyberattacks on photovoltaic systems in smart grid infrastructure using machine learning methods. *Fırat University Turkish Journal of Science & Technology, 20*(2), 445–454. https://doi.org/10.55525/tjst.1656368

[17] Huang, J., & Wan, Q. (2024). Smart grid line fault detection based on deep learning image recognition algorithm. *International Journal of Low-Carbon Technologies, 19*, 2174–2180. https://doi.org/10.1093/ijlct/ctae164

[18] Fang, J., Chen, K., Li, C., & He, J. (2023). An explainable and robust method for fault classification and location on transmission lines. *IEEE Transactions on Industrial Informatics*, 1–10. https://doi.org/10.1109/TII.2022.3229497

[19] Ajayi, O., Mirjafari, M., Idowu, P. B., & Ullah, M. H. (2024). Explainable AI for fault detection and classification in microgrids. In *Proceedings of the IEEE Energy Conversion Congress and Exposition (ECCE)* (pp. 1835–1840). https://doi.org/10.1109/ECCE55643.2024.10861648

[20] Mamodiya, U., Kishor, I., Mudholkar, P., Alqutaish, A., Alradwan, G., & Obeidat, M. (2026). A robust smart grid-aware cloud computing framework for sustainable energy management. *International Journal of Advances in Soft Computing and Its Applications, 18*(1), 396–433. https://doi.org/10.15849/ijasca.v18i1.63

[21] Ishfaq, H., Kanwal, S., Anwar, S., Abdussalam, M., & Amin, W. (2025). Enhancing smart grid security and efficiency: AI, energy routing, and T&D innovations (A review). *Energies, 18*(17), 4747. https://doi.org/10.3390/en18174747

[22] Al-Bermani, N. K., Bermani, A. K., Raad, A., & Manaa, M. E. (2025). AI-driven cybersecurity-based hybrid approach using blockchain for smart grids. *Journal of Discrete Mathematical Sciences and Cryptography, 28*(4-B), 1399–1411. https://doi.org/10.47974/jdmsc-2286

[23] Arindam, A. (2025). Advancing network security through deep learning: A hybrid graph-based and temporal approach to anomaly and threat detection. *International Journal for Science Technology and Engineering, 13*(5), 6095–6103. https://doi.org/10.22214/ijraset.2025.71415

[24] Mutambik. (2025). AI-driven cybersecurity in IoT: Adaptive malware detection and lightweight encryption via TRIM-SEC framework. *Sensors, 25*(22), 7072. https://doi.org/10.3390/s25227072

[25] Mamodiya, U., & Kishor, I. (2026). Artificial intelligence applications for enhancing efficiency in smart grids. In P. Raj, D. P. Sharma, P. K. Dutta, B. S. Prasad, & P. B. Soundarabai (Eds.), *Artificial Intelligence (AI) for IT Energy Efficiency and Green AI for Environment Sustainability*. Springer. https://doi.org/10.1007/978-3-031-89420-6_9

[26] El Maghraoui, A., El Hadraoui, H., Ledmaoui, Y., El Bazi, N., Guennouni, N., & Chebak, A. (2024). Revolutionizing smart grid-ready management systems: A holistic framework for optimal grid reliability. *Sustainable Energy, Grids and Networks*, 101452. https://doi.org/10.1016/j.segan.2024.101452

[27] Sarker, M. A. A., Shanmugam, B., Azam, S., & Thennadil, S. (2024). Enhancing smart grid load forecasting: An attention-based deep learning model integrated with federated learning and XAI for security and interpretability. *Intelligent Systems with Applications, 23*, 200422. https://doi.org/10.1016/j.iswa.2024.200422

[28] Okeke, O. C., Nwaoha, S. O., & Ezenwegbu, N. C. (2025). Hybrid machine learning models for enhancing cybersecurity in smart grid infrastructures. *International Journal of Research and Innovation in Social Science*, 4344–4351. https://doi.org/10.47772/IJRISS.2025.90400310

[29] Mamodiya, U., Kishor, I., Vidyullatha, P., Alqutaesh, A., Alradwan, G., & Obeidat, M. (2026). A hybrid fuzzy–deep learning framework for real-time cyber-attack detection in smart energy grids. *International Journal of Data and Network Science*. https://doi.org/10.5267/j.ijdns.2026.2.007

[30] Duan, J. (2024). Deep learning anomaly detection in AI-powered intelligent power distribution systems. *Frontiers in Energy Research, 12*, 1364456. https://doi.org/10.3389/fenrg.2024.1364456

[31] Akhtar, I., Atiq, S., Shahid, M. U., Raza, A., Samee, N. A., & Alabdulhafith, M. (2024). Novel glassbox based explainable boosting machine for fault detection in electrical power transmission system. *PLOS ONE, 19*(8), e0309459. https://doi.org/10.1371/journal.pone.0309459

[32] Ali, N. I., Brohi, I., Jamali, M.-U.-R., Arain, M. B., & Nangraj, A. R. (2025). A revolutionary approach using artificial intelligence and quantum cryptography: A review. *International Journal of Innovative Science and Technology*. https://doi.org/10.33411/ijist/20257314221436

[33] Wang, B., Baziar, A., & Askari, M. (2025). A deep reinforcement learning framework for adaptive resiliency enhancement in smart power grids. *IEEE Access*. https://doi.org/10.1109/ACCESS.2025.3593903

[34] Gokulraj, K., & Venkatramanan, C. B. (2024). Advanced machine learning-driven security and anomaly identification in inverter-based cyber-physical microgrids. *Electric Power Components and Systems*. https://doi.org/10.1080/15325008.2024.2346790

[35] Mohammed, S. H., et al. (2025). Dual-hybrid intrusion detection system to detect false data injection in smart grids. *PLOS ONE, 20*(1), e0316536. https://doi.org/10.1371/journal.pone.0316536

[36] Rastogi, A., Agrawal, A., Singh, R., & Aggarwal, A. (2024). A comprehensive cybersecurity resilience framework augmenting smart grid stability. In *Proceedings of INDISCON* (pp. 1–6). https://doi.org/10.1109/INDISCON62179.2024.10744380

[37] Usman, Y., Ihejirika, C. J., Offor, S. N., Robert, A., & Chataut, R. (2025). Green cybersecurity: Leveraging AI, ML, and LLMs to optimize energy, threat detection, and sustainability frameworks. *IEEE Access*. https://doi.org/10.1109/ACCESS.2025.3602451

[38] Mohammed, A. S., Shohdy, A., Mohammed, S. A., & Montaser, A. M. (2025). Design and optimization of a metamaterial absorber for enhanced solar cell efficiency and wide band microwave cross polarization conversion. *Scientific Reports, 15*(1). https://doi.org/10.1038/s41598-025-15840-w

[39] AzithTejaGanti, V. K., Senthilkumar, K., T. L., Karunakaran, S., Pandugula, C., & Khatana, K. (2025). Energy-efficient real-time hybrid deep learning framework for adaptive IoT intrusion detection with scalable and dynamic threat mitigation. *Social Science Research Network*. https://doi.org/10.2139/ssrn.5077540

[40] Barros, P. H., Agupugo, C. P., Ejichukwu, E., Hayden, M. D., & Ogunmoye, K. A. (2025). Smart grid security: Safeguarding sustainable energy systems from cyber threats. *World Journal of Advanced Research and Reviews, 26*(3), 1284–1301. https://doi.org/10.30574/wjarr.2025.26.3.2233

[41] Wang, R., Shen, Y., Wang, D., Jiang, Y., & Zhang, C. (2025). DSTF-GKAN: A lightweight spatiotemporal fusion framework for real-time eavesdropping detection in dynamic smart grid networks. *PLOS ONE, 20*(8), e0330593. https://doi.org/10.1371/journal.pone.0330593

[42] Pandey, T. N., Ravalekar, V., Nair, S. S. K., & Pradhan, S. K. (2025). A comparative analysis of classical machine learning models with quantum-inspired models for predicting world surface temperature. *Scientific Reports, 15*(1). https://doi.org/10.1038/s41598-025-12515-4