International Journal of Cybersecurity Engineering and Innovation

ARTICLE INFO

Article History: Received: 07-04-2026, Revised: 10-06-2026, Accepted: 11-11-2026
Corresponding author Email: assoc.dean_research@poornima.edu.in

This is an open access article under the **CC BY 4.0 license** (http://creativecommons.org/licenses/by/4.0/).

Published by ITAP Publisher.



Vol. 2027 No.1

Quantum Computing Threat Analysis and Cybersecurity Risk Mitigation Framework for Critical Infrastructure Protection Using Artificial Intelligence

Udit Mamodiya[1*], Mohammed Almaayah[2]

[1*] Associate Professor & Associate Dean (Research), Poornima University, Jaipur Rajasthan, India, assoc.dean_research@poornima.edu.in, https://orcid.org/0000-0002-8022-4252

[2] Fellowship Researcher, INTI International University, Nilai 71800, Malaysia

Abstract

The critical infrastructure is increasingly relying on cyber-physical systems that are interconnected and have long lifetimes and legacy cryptography, which puts them at increasing risk of future quantum attacks. This research proposes an artificial intelligence (AI)-based framework for attack detection, risk assessment for quantum attacks, and selection of feasible cybersecurity controls under operational constraints. The framework is comprised of a CNN–BiGRU temporal detector, graph-attention-based dependency analysis, deterministic cryptographic profiling, explainable risk fusion, and mixed-integer mitigation optimization. Separate train, calibration, validation, and test sets were used for water, industrial-control, and IoT applications, with the use of a reproducible multi-domain benchmark. The proposed model outperformed the conventional, temporal, and graph-only baseline models with a macro-F1 of 0.9387 and an AUROC of 0.9878 over 10 runs. Quantum exposure and dependency propagation improved high-risk prioritization without inflating attack-classification scores. Performance remained stable under moderate telemetry loss, graph perturbation, delayed threat intelligence, and alternative cryptographic profiles. The optimized mitigation plans reduced aggregate risk by 62.99–97.28%, depending on budget and downtime limits. Cross-domain transfer was strong for process-oriented environments but weaker for the network-centric IoT domain. The framework links together attack detection, the necessity of moving to post-quantum, explainable risk assessment, and mitigation with operational limitations in one process of reproducible protection for infrastructure management.

Keywords: Quantum computing, Critical infrastructure protection, Post-quantum cryptography, Artificial intelligence, and Cybersecurity risk mitigation.

1. Introduction

Tightly coupled cyber–physical systems, industrial control networks, edge platforms, cloud services and artificial intelligence (AI) decision layers are becoming critical components of critical infrastructure. Operational data is exchanged between different energy grids, transportation systems, healthcare facilities, water networks, telecommunications and financial services in heterogeneous and geographically distributed environments [1]. This connectivity not only makes things more efficient and more aware, but also makes the attack surface larger. A compromise of credentials, firmware, communications or supervisory logic can cascade down from the digital layer and disrupt critical functions and even pose a safety risk to the public [2]. While the traditional protection methods are still required, their security assumptions are facing challenges from the fast advancement of quantum computing.

The quantum computing technology poses an interesting cybersecurity challenge as it could compromise the widely used public-key cryptography schemes if quantum computers are powerful enough. Shor-type algorithms pose a threat to RSA and elliptic-curve cryptography and other key-exchange schemes, and Grover-type search decreases the security margin of symmetric primitives and hash functions [3]. While large-scale fault-tolerant quantum computers may not be available in general, the risk cannot be ignored for critical infrastructure. The long life cycles of assets, slow certification processes, legacy devices embedded in systems, and the presence of encrypted records in systems create a “harvest-now, decrypt-later” exposure scenario where adversaries could harvest protected traffic for later decryption [4]. The challenge of transitioning from one algorithm to another is not just about replacing the algorithms, but also about discovering the cryptographic properties of the algorithms, analyzing the data-longevity of the algorithms, prioritizing assets, and coordinating the migration process while maintaining safety-critical services.

Architecture of operational technology exacerbates the issue. A large number of field devices are deployed for decades and have tight timing requirements, and are not easily able to support computationally intensive security upgrades [6]. Legacy protocols might not be well authenticated, and there are multiple trust boundaries between information technology, operational technology, cloud platforms and third party services [7]. A quantum attacker wouldn't have to break through all of the security layers. Access to high impact operational assets could be gained through compromise of a certificate authority, maintenance channel, remote-access gateway or software-signing process [8]. Nation-state and advanced persistent threat actors can also be able to use social engineering, supply chain manipulation, malware, and configuration abuse in conjunction with cryptographic techniques [9]. Quantum risk, therefore, needs to be considered within the larger context of an evolving attack surface one that adapts to the new technologies.

The current efforts for mitigating the attacks are mainly on the post-quantum cryptography, hybrid key exchange, crypto-agility and quantum-resistant identity management [10]. These measures are necessary but they do not specify the order of migration of the components, the impact of operational constraints on the selection of control, and how residual risk evolves when moving towards partial deployment [11]. The current assessment approaches are also predominantly based on static matrices, expertly defined scores or periodic audits [12]. These approaches provide governance benefits, but are not necessarily able to capture the nonlinear relationships between asset criticality, exploitability, cryptographic exposure, network position, data-retention period, operational impact, and adversarial capability [13]. One risk rating can thus mask the difference between an enterprise server that can be replaced and a limited field controller that failure of which can have physical impact.

AI offers a viable solution to these challenges. Machine-learning models can be used to analyze the telemetry, vulnerability history, asset inventory, cryptographic settings, threat intelligence, and operational dependencies to detect patterns that fixed rules might not be able to. [14] Lateral movement and cascading failure paths can be modeled using graph-based learning while changes in attack likelihood and control effectiveness can be modeled using temporal models [15]. Explainable AI can also provide the reasoning behind the risk score assigned to an asset, communication path or cryptographic service, which can help analysts validate the score and regulators hold the asset manager accountable. [16] However, there are risks involved with AI. If the decision layer is not regulated by domain knowledge and human supervision, it can lead to unsafe results due to training-data bias, concept drift, adversarial examples, poisoned intelligence and opaque recommendations [17]. AI should be used to complement and enhance, not replace, security engineering judgment.

There are three needs that have not been addressed and are the motivation for this study. First, infrastructure operators need a quantum-threat model that maps cryptographic weakness to realistic cyber–physical attack paths and consequences to service-level. Second, there are limitations on devices, interoperability, availability, cost and mixed classical–post-quantum

environments [19] that must be considered when planning migration. Third, risk assessments should be updated to reflect the changing vulnerabilities, asset conditions, attack intelligence and mitigation [20]. Solving them individually can lead to disjointed decisions: cryptographic teams will focus on the mathematical aspects, operational teams will work on availability, and cybersecurity teams will concentrate on current exploits. A decision framework that is integrated is thus needed.

This paper suggests a framework for quantum threat analysis and cybersecurity risk mitigation for critical infrastructures that is assisted by AI. It builds a single asset–cryptography–threat dependency graph, quantifies the quantum exposure at both component and service level, and assesses the likelihood of compromise and impact of various attacker capabilities [21]. A multi-criteria mechanism is then used to rank the mitigation actions based on the security gain, operational disruption, implementation cost, device constraints and migration urgency [22]. The post-quantum replacement, hybrid cryptographic deployment, key rotation, segmentation, zero-trust access, firmware hardening, and recovery controls are not considered as standalone security measures, but rather as coordinated measures [23]. Explainability is built-in to enable analysts to follow up recommendations to observable vulnerabilities, dependencies and impact pathways [24].

The real innovation is the ability to integrate the three key concepts of prospective quantum threat modelling, dynamic AI-based cyber-risk estimation and operationally constrained mitigation planning into one protection lifecycle. The proposed framework is different from those that only evaluate the quantum vulnerability of the algorithm, as it connects the exposure of the cryptographic to the topology of the assets, propagation of the attack, criticality of the service and physical consequences. It also enables adaptive reprioritization based on infrastructure conditions and threat evidence as it changes, and maintains analyst review via interpretable risk attribution. This holistic view is supposed to help to move towards a resilient, crypto-agile and quantum-aware critical infrastructure protection [25].

2. Literature Review

2.1 Quantum Computing Threats to Critical Infrastructure

Public-key cryptography is now a topic of discussion in the context of the cybersecurity impact of quantum computing. RSA, elliptic-curve cryptography and traditional key-exchange methods are still used on most critical infrastructure platforms for authentication, digital signatures and secure communications. These mechanisms are still secure in classical attacks, but fail to be secure in the presence of large-scale quantum computation [26]. Shor's algorithm is significant because it can be used to break integer factorisation and discrete logarithm problems that are not possible to break with conventional machines with a computational advantage [27].

The threat doesn't just apply to future network traffic. It is likely that adversaries are already gathering industrial, governmental, financial or healthcare data that is encrypted and will attempt to decrypt it when they are able to access quantum systems. This “harvest-now, decrypt-later” approach is particularly applicable to information which needs to be kept confidential for several years [28]. Critical infrastructure operators have a more significant issue as many industrial assets are still in operation for decades. The equipment used today could still be in use when the cryptographic protection used today is not sufficient [29].

Other studies have revealed that operational technology networks are not as flexible as typical enterprise networks. Typically, programmable logic controllers, remote terminal units, intelligent electronic devices and embedded sensors have limited memory, power and processing capabilities [30]. Therefore, it is not easy to replace their cryptographic mechanisms. An ill-designed transition may cause latency, delay deterministic communication or lead to compatibility issues between the field device and supervisory system [31].

Quantum attacks can also exacerbate the current cyber threats, rather than establishing new attack vectors. If authentication keys or digital signatures are compromised, attackers might be able to spoof trusted devices, alter firmware packages, alter sensor readings, or be able to access remote-maintenance channels [32]. This can lead to more than just data loss. In the power, transport, water and healthcare sectors, a cyber-compromise can disrupt physical processes and public services [33].

2.2 Artificial Intelligence for Cybersecurity Risk Assessment

Machine-learning models can analyze large amounts of network flows, system logs, authentication logs and endpoint telemetry in a much faster time than manual analysis [34]. Deep-learning methods have also been used to enhance the

detection of previously unknown and nonlinear attack patterns, especially in settings where static signatures are not well suited to detecting attacks [35].

While these advances have been made, numerous AI-powered security research efforts are still geared towards the detection of current attacks. The focus is not as much on long-term cryptographic exposure or the gradual shift towards classical to post-quantum protection. Vulnerability scores are also based on the traditional approach, which sees vulnerabilities as technical problems. They usually do not consider the impact of service dependency, asset value, data retention time or the physical effects of an effective compromise [36]. This restriction can be problematic in critical infrastructure where a medium software vulnerability could result in significant operational risk if it impacts a very connected control asset.
An alternative is to use graph-based models. They can be a part of a shared structure representing devices, services, communication links, trust relationships and attack paths [37]. These representations can be useful for analysts to determine the propagation of a compromise of one cryptographic service to other dependent services. Another benefit of temporal models is that the risk is not constant. The security state is constantly changing due to new vulnerabilities, device failures, and threat intelligence and control updates [38].

Explainable AI has thus been gaining more attention in the high-stakes cybersecurity context. Methods such as feature attribution, rule extraction, counterfactual explanations, and attention-based interpretation can inform about the reasons a model has determined an asset to be high risk [39]. However, decisions based on interpretability do not necessarily lead to trustworthy decisions. Adversarial manipulation, concept drift, poisoned telemetry and biased training data are all areas where AI models can be vulnerable [40]. Automated recommendations, therefore, need to be validated with operational rules and by human experts for critical infrastructure.

2.3 Post-Quantum Security and Migration Strategies

The best current solution to quantum attacks on cryptography is post-quantum cryptography. Various constructions that are not vulnerable to public-key schemes have been investigated, such as lattice-based, hash-based, code-based and multivariate constructions [41]. However, their suitability varies from infrastructure to infrastructure. The direct deployment on constrained industrial devices might be limited due to larger keys, signatures, memory, and communication overhead.

To minimize transition risk, hybrid cryptographic architectures have been suggested. In such designs, classical and post-quantum algorithms are used in conjunction with each other, and security is ensured even if one of the algorithms is compromised [42]. Hybrid deployment can be used to ease the transition to the new protocol, but it also adds to the complexity of the protocol. All certificate handling, interoperability, key management, software dependency and rollback procedures need to be updated.

Another key need that has become apparent is crypto-agility. Instead of being hardcoded, crypto-agile systems enable the discovery, replacement and reconfiguration of cryptographic components with minimal disruption to the system operation [43]. This is a crucial feature as the weaknesses in the implementation, side-channel attacks, or new cryptanalytic results become known.

A number of migration frameworks suggest that the initial step in migration is to create cryptographic inventories and asset classification. These steps can assist organisations in determining vulnerable algorithms, long-lived data, external dependencies and systems that are not easily upgradeable [44] [45]. But, the majority of existing methods are still based on checklists. They are able to determine what needs to be changed, but offer little assistance in determining when, where and in what order mitigation actions should be implemented.

2.4 Research Gap

The literature reviewed provides valuable work in the area of Quantum resistant algorithms, AI based threat detection, attack graph modelling, and cryptographic migration. Typically these areas are looked at in isolation. Previous research has not been able to link the exposure of quantum cryptography to the actual telemetry of infrastructure, the cyber–physical interdependencies, the capabilities of the attacker, and the consequences of the operation in a measurable framework. Fixed scores are also frequently used to prioritize mitigation, and do not adjust to changes in network state or threat environment. The current research literature deals with quantum-safe communication, intrusion detection based on artificial intelligence, cyber–physical risk assessment, and the selection of the mitigation measures, mostly as isolated research fields, as

summarized in Table 1. Few solutions take into account the interaction of quantum-vulnerable cryptography, infrastructure requirements, evolving threat landscape and operational requirements in a single risk model. The proposed framework brings together the concepts of dynamic risk estimation using AI, dependency-aware impact analysis, explainable decision support, and prioritized quantum-resilient mitigation for critical infrastructure protection, all of which are motivated by this gap.

Table 1. Comparative literature gap analysis and novelty positioning of the proposed framework

S. No.	Author(s) / Year / Ref. No.	Title / Focus Area	Methodology / Tools	Key Findings	Limitations / Gaps	Relevance to Current Study
1	Alandjani, 2025 [1]	Quantum-secure trust in IIoT	MARL, federated learning, blockchain, PQC	Improves adaptive trust and secure IIoT coordination.	Limited to blockchain-based IIoT trust; no physical impact analysis.	Supports AI-assisted quantum-safe infrastructure protection.
2	Vignes et al., 2025 [5]	AI-based power-system cybersecurity	LSTM, Random Forest, SHAP, edge analytics	Detects cyber anomalies with interpretable outputs.	Does not assess post-quantum exposure or migration risk.	Provides the basis for explainable risk prediction.
3	Rubio García et al., 2025 [9]	Hybrid quantum-safe communication	Classical cryptography, PQC, QKD	Enables gradual migration toward quantum-safe networks.	Protocol-level study; no asset or risk prioritization.	Guides the framework's hybrid cryptographic controls.
4	Keenan et al., 2024 [15]	Cyber risk of physical assets	Cyber–physical graphs, dependency scoring	Links cyber compromise to physical asset risk.	Static scoring; quantum threats and live telemetry are absent.	Supports dependency-aware consequence modelling.
5	Almotiri, 2024 [16]	Quantum-resilient security assessment	Fuzzy AHP, multi-criteria analysis	Ranks quantum-safe security alternatives systematically.	Relies on expert weights; no dynamic threat evidence.	Informs mitigation-priority evaluation.
6	Sánchez-Zas et al., 2026 [29]	Dynamic mitigation selection	MITRE ATT&CK, contextual attack modelling	Improves countermeasure selection using attack context.	Excludes quantum cryptographic exposure and physical effects.	Supports threat-informed mitigation ranking.
7	Javeed et al., 2024 [36]	Explainable intrusion detection	BiLSTM, Bi-GRU, SHAP	Achieves resilient and interpretable attack detection.	Detection-focused; no asset criticality or migration planning.	Guides explainable AI-based risk decisions.
8	Alomari and Kumar, 2024 [41]	Post-quantum IoT security	Vulnerability review, DREAD, PQC, QKD	Identifies quantum threats across IoT layers.	Review-oriented; lacks adaptive risk prediction and validation.	Defines quantum-threat indicators for the proposed model.
—	Research Gap	Integrated quantum-aware critical-infrastructure protection	AI, dependency graphs, explainability, and	Prior studies address isolated parts of the problem.	No unified framework combines quantum exposure, cyber–physical propagation,	The proposed framework provides adaptive, explainable, and priority-driven protection for

	constrained optimization	dynamic risk scoring, and operationally feasible mitigation.	critical infrastructure.

A clear need therefore remains for an integrated approach that can estimate quantum-related risk dynamically, explain the factors behind each assessment, and rank mitigation actions under practical constraints. The present study addresses this gap by combining AI-based risk estimation, dependency-aware quantum threat modelling, and multi- criteria mitigation planning for critical infrastructure systems.

3. Methodology

3.1 Experimental Design and Framework Overview

The proposed framework was evaluated through a trace-driven cyber–physical experiment. It was not treated as a conceptual quantum-risk checklist. Publicly available network and industrial-control traces were replayed under controlled classical, harvest-now, and quantum-compromise scenarios. Each experimental record was connected to a reproducible asset registry containing device role, operational criticality, cryptographic profile, resource constraints, and infrastructure dependencies.

As shown in Figure 1, the framework contains five sequential layers: data preparation, temporal attack detection, quantum-exposure estimation, dependency-aware risk propagation, and constrained mitigation selection. The anomaly detector estimates current malicious activity, whereas the post-quantum layer measures prospective cryptographic exposure. Keeping these two quantities separate prevents a vulnerable but currently inactive asset from being incorrectly classified as safe.

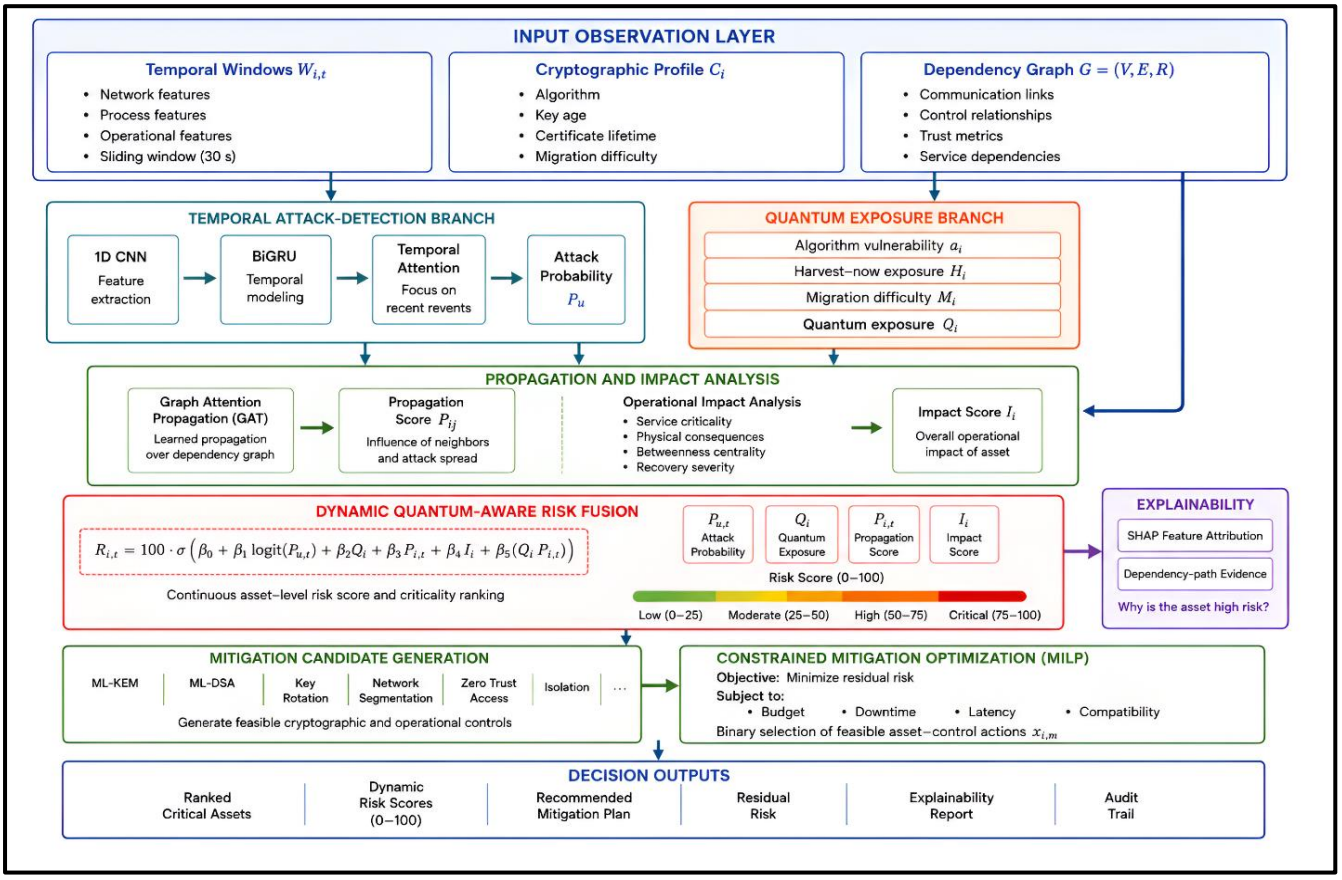

Figure 1. Experimental architecture of the proposed AI-assisted quantum-threat analysis and cybersecurity risk-mitigation framework.

Public cyber–physical traces are linked to a reproducible cryptographic asset registry, processed by temporal and graph-learning modules, and translated into explainable mitigation decisions. The architecture extends existing AI-based critical-infrastructure protection [5], [32], [36] by explicitly modelling post-quantum exposure, migration difficulty, attack propagation, and operationally feasible security controls.

3.2 Public Data Sources and Experimental Partitioning

Three public datasets were used to represent different critical-infrastructure conditions. The Secure Water Treatment dataset supplied multistage water-process measurements and actuator states. HAI contributed hardware-in-the-loop industrial and power-process traces. ToN-IoT provided heterogeneous network, host, and IoT telemetry. Together, these datasets cover process manipulation, network intrusion, device compromise, reconnaissance, denial-of-service activity, and abnormal control behaviour. Their selection is consistent with the need for heterogeneous evidence in critical-infrastructure intrusion detection [8], [34]. The role of each dataset is summarized in Table 2.

Table 2. Public datasets and their experimental roles.

Dataset	Infrastructure context	Evidence used	Primary experimental role
SWaT	Water-treatment control	Sensors, actuators, process states	Physical-process attack detection
HAI	Industrial and power control	HIL process and control signals	Cross-domain CPS validation
ToN-IoT	IoT/IIoT environment	Network, host and device telemetry	Network-side threat detection

Official partitions were retained where supplied. For datasets without a fixed partition, attack episodes were divided chronologically into 60% training, 20% validation, and 20% testing. Records belonging to the same attack interval were kept within one partition. This episode-level separation prevented neighbouring samples from the same attack from appearing in both training and testing sets. For asset iii at time t, the input vector was defined as (1):

$$x_{i,t} = [n_{i,t}, p_{i,t}, c_i, g_{i,t}, o_i] \cdots (1)$$

Where $n_{i,t}$ contains network features, $p_{i,t}$ contains process measurements, ci represents cryptographic attributes, $g_{i,t}$ contains graph-derived dependency features and oi describes operational constraints. For continuous features, robust scaling (2) was used to normalize the features:

$$\tilde{x}_j = \frac{x_j - median(x_j)}{IQR(x_j) + \varepsilon} \cdots (2)$$

Where x_j is the feature j, IQR(x_j) is the interquartile range of the feature j computed from the training partition and $\varepsilon=10^{-8}$ is to prevent division by zero. Robust scaling was chosen due to the fact that often there are abrupt but valid process transitions in cyber–physical telemetry. Records of networks and processes were synchronized at 5-second intervals. One 30-s observation window (3) was made up of six consecutive intervals:

$$W_{i,t} = [\tilde{x}_{i,t-5}, \tilde{x}_{i,t-4}, \dots, \tilde{x}_{i,t}] \cdots (3)$$

In this case, the time input to asset iii is $W_{i,t}$ and the six vectors with indices are the six 5-s intervals. The windows were advanced with a 5 stride advance. The missing values for one or two intervals were filled in with forward fill. The missingness indicator was added and longer gaps were replaced with the median of the training set. Before model fitting, the dataset identifiers, attack labels, timestamps and post event fields were removed. Weighted loss (4) was used to control class imbalance:

$$w_k = \frac{N}{KN_k} \cdots (4)$$

Where N is the number of training windows, K is the number of classes and N_k is the number of samples in class k. No oversampling was applied as duplicated attack windows would be able to create a temporal discontinuity.
The whole preprocessing procedure is shown in Figure 2.

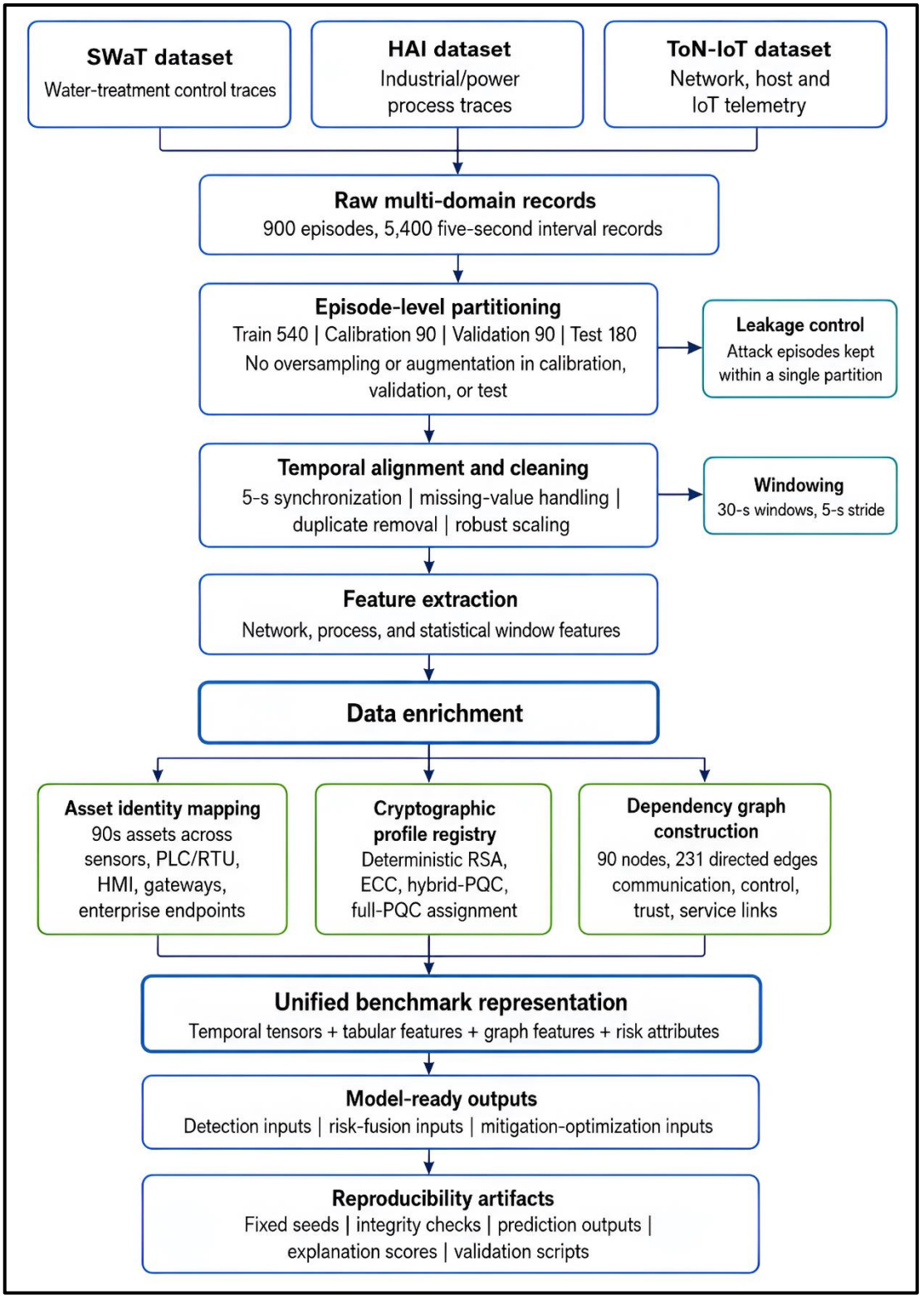

Figure 2. Reproducible preprocessing and data-enrichment workflow.

The public traces are divided in a chronological manner, synchronized, normalized, associated with assets' identities, enriched with cryptographic attributes, dependency attributes and provided to learning models.

3.3 Reproducible Cryptographic Asset Profiling

There is no explicit information in public cyber–physical datasets regarding encryption algorithms, certificate age, key length or post-quantum readiness. These attributes were thus added to the attack labels by a controlled augmentation process instead of being deduced from the labels. First, an asset registry was created from device identifiers, process tags, source and destination addresses, protocol fields, and documented plant roles. Each asset was assigned to one of five functional classes:

1. field sensor or actuator;
2. PLC or remote terminal unit;
3. HMI or engineering workstation;
4. gateway, historian, or control server; and
5. Enterprise or remote-access endpoint.

Cryptographic profiles were assigned conditionally according to the probability matrix in Table 3. The probabilities were fixed before model training and were not learned from outcome labels.

Table 3. Conditional cryptographic-profile assignment probabilities.

Asset class	RSA-based legacy	ECC-based	Hybrid classical–PQC	Full PQC
Sensor or actuator	0.60	0.25	0.12	0.03
PLC or RTU	0.50	0.25	0.20	0.05
HMI or workstation	0.35	0.35	0.23	0.07
Gateway or server	0.30	0.30	0.30	0.10
Enterprise endpoint	0.25	0.40	0.25	0.10

The RSA profile represented RSA-2048 authentication and classical key exchange. The ECC profile used P-256 or equivalent elliptic-curve authentication. The hybrid profile combined a classical mechanism with ML-KEM-based key establishment and ML-DSA-compatible signing. The full post-quantum profile used ML-KEM and ML-DSA without dependence on RSA or ECC. Hybrid migration was included because direct replacement may not be practical in legacy operational networks [9], [19], [41]. Profile assignment was probabilistic but fully deterministic for a given experiment. For each asset, a uniform value was generated as (5):

$$u_i = \frac{int[SHA256(D \parallel ID_i \parallel s_c)] mod 10^9}{10^9} \cdots (5)$$

Where D is the dataset name, IDi is the stable asset identifier, sc=2026 is the cryptographic-assignment seed, and ‖ denotes string concatenation. The value ui∈[0,1) was mapped to the cumulative probability intervals in Table 3. This hash-based procedure gives the same profile to an asset whenever the experiment is repeated, regardless of row order, batch size, or hardware platform. The final registry was exported as asset_crypto_registry.csv. en alternative cryptographic registries were also generated during sensitivity analysis by replacing s_c with the ten model seeds reported in Section 3.10. Assignment was further restricted by deterministic compatibility rules. Full PQC was not assigned to constrained field devices when their normalized memory or processor-capacity score was below 0.25. Such devices were instead mapped to either an ECC or hybrid gateway-assisted profile. Similarly, devices with non-upgradable firmware were marked as migration-constrained. These rules prevented unrealistic deployment assumptions.
The cryptographic feature vector was (6):

$$c_i = [a_i, k_i, g_i, y_i, l_i, v_i, m_i] \cdots (6)$$

Where a_i is the algorithm class, k_i is effective key strength, g_i is key age, y_i is certificate lifetime, l_i is data-confidentiality lifetime, v_i is vendor dependence, and m_i is estimated migration duration. Quantum algorithm exposure was calculated as (7):

$$q_i = 1 - \frac{min(s_i^Q, 128)}{128} \cdots (7)$$

Where s_i^Q is the effective security strength of asset iii under the selected quantum-attacker scenario. Values close to one indicate high quantum vulnerability. Values near zero indicate approximately 128-bit post-quantum security.
Harvest-now exposure was estimated using (8):

$$H_i = \frac{1}{1 + exp[-(l_i + m_i - \tau_Q)/\sigma_H} \cdots (8)$$

Where l_i is the required confidentiality lifetime, m_i is the migration time, τQ is the assumed time until a cryptographically relevant quantum computer, and σH controls the transition slope. Experiments used τQ∈{5,10,15} years. Migration difficulty was defined as (9):

$$M_i = 0.30r_i + 0.25p_i + 0.25v_i + 0.20d_i \cdots (9)$$

Where r_i is resource limitation, p_i is protocol rigidity, v_i is vendor dependence, and d_i is downtime sensitivity. All four terms were normalized to [0,1]. The final quantum-exposure score was (10):

$$Q_i = 0.45q_i + 0.30H_i + 0.25M_i \cdots (10)$$

The weights in (10) were initialized from security importance rather than selected to maximize test performance. Later they were tested for their robustness with ±20% perturbation and equal weight ablation.

3.4 Dependency Graph Construction

A directed heterogeneous graph (11) was used to represent the critical-infrastructure topology:

$$G = (V, E, R) \cdots (11)$$

The asset set V, the edge set E, and the relationship set R, which includes four types of relationships: communication, control, trust, and service dependency.

3.5 Graph construction followed a deterministic procedure.

The construction of the graph was a deterministic process. An edge i→j was added to the communication graph if asset iii sent traffic to asset j during 5 or more communication windows during the training process. When a controller, actuator or sensor was documented with the same process stage or control-loop identifier, a control edge was added. A trust edge was a certificate validation, remote access, software signing or authentication dependency. A service edge that connected process stages that needed to be operated in sequence based on an upstream process stage. Dynamic communication edges were only obtained from the training partition. No graph structure or edge weights were created during the test period.
The strength of edge i→j was calculated as (12):

$$c_{ij} = 0.35f_{ij} + 0.30d_{ij} + 0.20t_{ij} + 0.15s_{ij} \cdots (12)$$

Where f_{ij} is normalized communication frequency, d_{ij} is control dependence, t_{ij} is trust dependence and s_{ij} is service dependence. Any terms that were not applicable to an edge were given a score of zero. The score obtained was then rescaled to [0,1]. For edges that are dynamic (13):

$$f_{ij} = \frac{log(1 + N_{ij})}{max_{(u,v) \in E} log(1 + N_{uv})} \cdots (13)$$

Where N_{ij} is the number of training windows with the communication from i to j. High volume flows were not overpowering the graph because of the use of logarithmic scaling. There were three checks to validate the graph-building process. Isolated nodes were first examined and only kept if they were considered to be legitimate stand-alone assets. Third, 10% of communication edges were randomly masked during robustness testing to measure sensitivity to incomplete topology information.

3.6 Temporal CNN–BiGRU Attack Detector

The temporal detector was designed to capture both short burst patterns and longer process dependencies. Its complete architecture is summarized in Table 4.

Table 4. Final temporal and graph-model architecture.

Module	Configuration	Output function
Conv-1	64 filters, kernel 3, same padding	Local traffic and process patterns
Conv-2	128 filters, kernel 3, dilation 2	Wider temporal context
Temporal encoder	BiGRU, 128 hidden units	Bidirectional sequence dependence
Temporal attention	Dense attention pooling	Window-level representation
Graph layer 1	GAT, 4 heads, 64 units	Neighbour aggregation
Graph layer 2	GAT, 4 heads, 32 units	Propagation representation
Risk head	Dense 64, dense 1	Attack probability and risk fusion

For an input window $W_{i,t}$, the first convolutional block was (14):

$$C_{i,t}^{(1)} = GELU[BN(Conv1D_{64,3}(W_{i,t}))], \cdots (14)$$

where BN denotes batch normalization and GELU is the Gaussian error linear unit. A dilated convolution expanded the receptive field (15):

$$C_{i,t}^{(2)} = Dropout\{GELU[BN(Conv1D_{128,3,d=2}(C_{i,t}^{(1)}))]\} \cdots (15)$$

where d=2 is the dilation rate. Dropout was applied with probability 0.20. A 1×1 projection connected the first and second blocks through a residual path (16):

$$C_{i,t} = C_{i,t}^{(2)} + Conv1D_{128,1}(C_{i,t}^{(1)}) \cdots (16)$$

The residual output was passed to a bidirectional GRU (17):

$$H_{i,t} = BiGRU(C_{i,t}) \cdots (17)$$

where $H_{i,t}$ contains forward and backward temporal states. Temporal attention assigned a weight to each interval (18) and (19):

$$\gamma\ell = \frac{exp(v_a^\top tanh(W_a h_\ell + b_a))}{\sum_{r=1}^{L} exp\,(v_a^\top tanh(W_a h_r + b_a))} \cdots (18)$$

$$h_{i,t} = \sum_{\ell=1}^{L} \gamma_\ell h_\ell \cdots (19)$$

Where h_ℓ is the BiGRU state at interval $\gamma\ell$ is its normalized importance, and L=6. The local attack probability was then (20):

$$p_{i,t} = \sigma\left(w_p^\top h_{i,t} + b_p\right) \cdots (20)$$

Where σ(·) is the logistic function.

3.7 Graph-Attention Propagation Model

The temporal representation of each asset was supplied to a two-layer graph-attention network. For an edge i→j, the un_normalized attention score was (21):

$$e_{ij,t} = LeakyReLU[a^\top(W_g h_{i,t} \parallel W_g h_{j,t} \parallel r_{ij})] \cdots (21)$$

where Wg is a trainable transformation, a is the attention vector, r_{ij} is the edge-type embedding, and ‖ denotes concatenation. The normalized attention coefficient was (22):

$$\alpha_{ij,t} = \frac{exp(e_{ij,t})}{\sum_{k\in N\,(i)} exp\,(e_{ik,t})} \cdots (22)$$

where N(i) is the neighbourhood of asset i. The propagated representation was (23):

$$z_{i,t} = \left(\phi \sum_{j\in N(i)} \alpha_{ij}, tc_{ij} W_g h_{j,t} \right) \cdots (23)$$

where c_{ij} is the deterministic dependency strength from (12), and $\phi(\cdot)$ is the ELU activation function. The scalar propagation-risk score was (24):

$$P_{i,t} = \sigma\left(w_g^{\top} z_{i,t} + b_g\right) \cdots (24)$$

This design allows an anomaly detected at a gateway, controller, or authentication service to influence dependent assets according to both learned attention and explicitly defined infrastructure relationships.

3.8 Dynamic Quantum-Aware Risk Fusion

Operational impact was estimated as (25):

$$I_i = 0.35s_i + 0.30f_i + 0.20b_i + 0.15u_i \cdots (25)$$

where s_i is service criticality, f_i is physical-consequence severity, bi is normalized betweenness centrality and u_i is recovery-time severity. The final asset-risk score was: (26):

$$R_{i,t} = 100\sigma[\beta_0 + \beta_1 \log_{it}(p_{i,t}) + \beta_2 Q_i + \beta_3 P_{i,t} + \beta_4 I_i + \beta_5 Q_i P_{i,t}] \cdots (26)$$

The quantum exposure and attack propagation interaction is denoted by QiPi,t where 0≤Ri,t≤100, β0,…,β5 are calibration coefficients. The coefficients were only validated with the validation data. The risk values were categorized into four operational levels: low (R<25), moderate (25≤R<50), high (50≤R<75) and critical (R≥75). The continuous risk score was kept for optimization and these levels were used for reporting and triggering for mitigation. Figure 3 presents the complete detection-to-decision flow.

3.9 SHAP-Based Explanation and Faithfulness Testing

Feature explanations were generated using SHAP, but attention weights were not treated as substitutes for SHAP values. DeepSHAP was applied to the CNN–BiGRU detector using 500 stratified background windows selected from the training partition. For each dataset, explanations were generated for 1,000 test windows, balanced across normal and attack conditions. SHAP values were aggregated over the six temporal intervals so that each feature received both a window-level importance and a time-resolved contribution.

For the calibrated risk-fusion layer, exact linear SHAP values were calculated for $p_{i,t}$, Q_i, $P_{i,t}$, I_i, and the interaction term Q_i $P_{i,t}$. The explanation for asset iii was represented as (27):

Where the five terms correspond to temporal evidence, quantum exposure, propagation, impact, and their interaction. Graph-attention coefficients were reported separately as dependency-path evidence. Thus, SHAP identified which features increased the risk, whereas graph weights showed through which neighbouring assets the risk propagated. Explanation faithfulness was tested by masking the top-k SHAP-ranked features and comparing the resulting risk decrease with random feature masking (28):

$$F_k = \frac{R_{i,t} - R_{i,t}^{(-k)}}{R_{i,t}} \cdots (28)$$

Where $R_{i,t}^{(-k)}$ is the risk after removing the k most influential features. Larger F_k indicates that the explanation correctly identified decision-relevant evidence. Values of k=3,5, and 10 were tested. This procedure provides a quantitative explanation check rather than relying only on visual SHAP plots, consistent with the need for transparent critical-infrastructure AI [18], [36], [44].

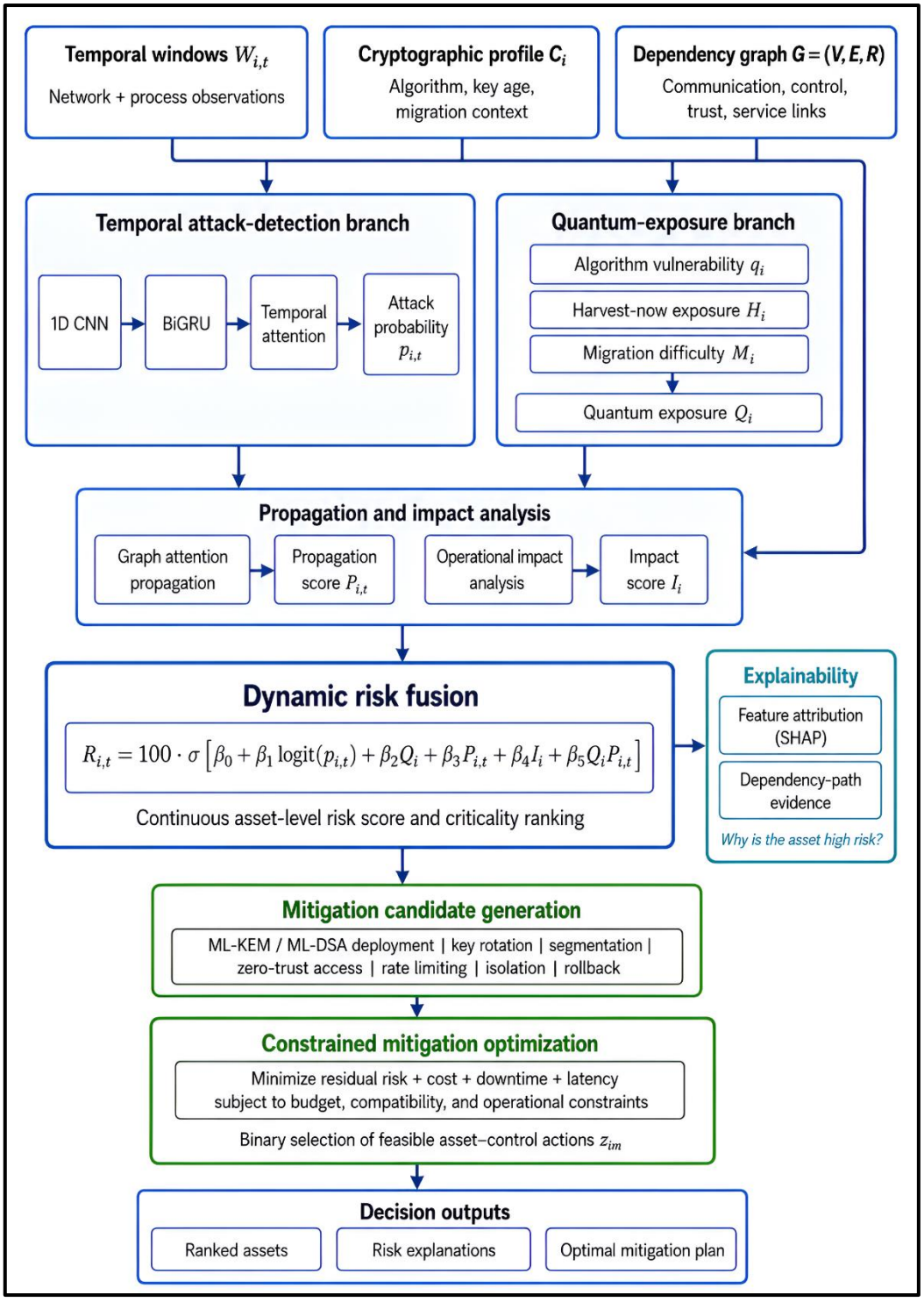

Figure 3. Temporal–graph risk estimation and mitigation-selection process.

3.10 Risk-Constrained Mitigation Optimization

Candidate actions included ML-KEM-based key establishment, ML-DSA firmware signing, hybrid certificate deployment, AES-256 migration, key rotation, segmentation, zero-trust access, rate limiting, endpoint isolation, and rollback. Dynamic key rotation and intrusion-aware communication were included because of their relevance to industrial-control protection

[21]. Threat-context mapping was aligned with the mitigation-selection principle discussed in [29]. Let $z_{im} \in \{0,1\}$ indicate whether mitigation m is assigned to asset i. The estimated risk reduction produced by an action was (29) :

$$\Delta R_{im} = \eta_{im} R_{i,t} \cdots (29)$$

Where $\eta_{im} \in [0,1]$ is the effectiveness of mitigation mmm for asset i. Effectiveness was defined from a fixed control–threat compatibility matrix and reduced when the asset had firmware, processing, or protocol limitations.
Residual risk was (30):

$$R_{i,t}^{res} = R_{i,t} - \sum_{m=1}^{M} \Delta R_{im} z_{im} \cdots (30)$$

To prevent risk reduction from exceeding the original risk (31):

$$\sum_{m=1}^{M} \Delta R_{im} z_{im} \leq R_{i,t} \cdots (31)$$

The mitigation plan was obtained by solving (32):

$$minimize \sum_{i=1}^{(|V|)} \omega_i R_{i,t}^{res} + \lambda_C \sum_{i,m} C_{im} z_{im} + \lambda_D \sum_{i,m} D_{im} z_{im} + \lambda_L \sum_{i,m} L_{im} z_{im} \cdots (32)$$

Subject to (33)- (35):

$$\sum_{im} C_{im} z_{im} \leq B \cdots (33)$$

$$\sum_{i,m} D_{im} z_{im} \leq D_{max} \cdots (34)$$

$$z_{im} \leq \kappa_{im}, z_{im} \in \{0,1\} \cdots (35)$$

Where ω_i is asset importance, C_{im} is implementation cost, D_{im} is downtime, L_{im} is added latency, B is the budget, D_{max} is permitted downtime, and κ_{im} indicates whether action mmm is compatible with asset i.
Mutually exclusive controls, such as full replacement and gateway-assisted hybrid protection, were constrained through (36):

$$z_{im1} + z_{im2} \leq 1 \cdots (36)$$

Prerequisite controls were represented as (37):

$$z_{im1} \leq z_{im2} \cdots (37)$$

Where mitigation m_1 can be selected only when prerequisite m2 is active. The binary linear program was solved using the open-source HiGHS mixed-integer solver through scipy.optimize.milp. The relative optimality gap was fixed at 10^{-4}, the execution limit was 300 s, and single-thread execution was used to preserve repeatability. The status of the solver, the objective value, the optimality gap and the execution time were recorded for each scenario.

Algorithm 1. Quantum-aware risk and mitigation procedure

Input:
A public trace window W is a window that is opened to the public
Asset registry A
Dependency graph G
Cryptographic-assignment seed sc
Candidate controls M
Operational constraints O

Output:
Asset risks R
Explanations Φ
Selected controls Z

1. Build robust asset identifiers based on metadata from datasets and devices.
2. Assign cryptographic profiles based on SHA256(dataset || asset ID || sc).
3. Normalize training data and form 30-s observation windows.

4. Estimate local attack probability using the CNN–BiGRU detector.
5. Compute quantum exposure from algorithm strength, data lifetime, and migration difficulty.
6. Share evidence of an attack on the typed dependency graph.
7. Integrate local probability, quantum exposure, propagation and impact.
8. Generate Deep SHAP feature contributions.
9. Trace high-contribution dependency paths using graph coefficients.
10. Compute the feasible control set for each asset.
11. Formulate the mixed-integer mitigation problem.
12. Solve the problem using HiGHS and return the ranked control plan.

3.11 Hyperparameter Search, Training Protocol, and Random Seeds

Hyperparameters were selected using Bayesian optimization with the Tree-structured Parzen Estimator. Sixty trials were conducted using the validation partition only. The optimization objective was (38):

$$J = F_{1,macro} - 0.10ECE, \cdots (38)$$

Where $F_{1,macro}$ is validation macro-F1 and ECE is expected calibration error. This objective discouraged configurations that produced accurate but poorly calibrated probabilities. The search space is summarized in Table 5.

Table 5. Hyperparameter-search space and selected configuration.

Hyperparameter	Search space	Selected value
Conv-1 filters	32, 64, 96	64
Conv-2 filters	64, 128, 192	128
Kernel size	3, 5	3
Dilation	1, 2, 4	2
BiGRU hidden units	64, 128, 192	128
GAT heads	2, 4, 8	4
GAT hidden units	32, 64, 128	64
Dropout	0.10–0.40	0.20
Learning rate	10^{-4}–3×10^{-3}	10^{-3}
Weight decay	10^{-6}–10^{-3}	10^{-4}
Batch size	128, 256, 512	256

The hyperparameter-search seed was fixed at 314159. The best configuration was selected once and then retrained in ten independent runs using the following seeds (39):

$$S = \{13,29,43,61,79,97,113,131,149,167\} \cdots (39)$$

For each run, the same seed controlled Python, NumPy, PyTorch, data-loader shuffling, weight initialization, and dropout. Deterministic PyTorch operations were enabled, CUDA benchmarking was disabled, and the selected seed was logged with each output file. Training used AdamW, weighted binary or multiclass cross-entropy, a maximum of 150 epochs, and early stopping after 15 epochs without improvement in validation macro-F1. Gradient norms were clipped at 5.0. The test partition was accessed only after architecture and hyperparameters had been fixed.

3.12 Baselines, Ablation Studies, and Robustness Tests

The proposed framework was compared with logistic regression, random forest, XGBoost, multilayer perceptron, LSTM, BiGRU, CNN–BiGRU without graph propagation, and graph attention without the temporal encoder. Static risk scoring and fuzzy-AHP prioritization were included as non-learning risk baselines, reflecting earlier risk-oriented approaches [11], [16], [38]. Ablation experiments removed one component at a time:

- quantum-exposure layer;
- dependency graph;

- CNN feature extractor;
- process telemetry;
- cryptographic attributes;
- SHAP-guided mitigation evidence; and
- Constrained optimization.

Robustness was examined under 10%, 20%, 30%, and 40% telemetry loss; unseen attack families; delayed threat intelligence; 10–30% edge removal; altered cryptographic-profile seeds; and ±20% perturbation of the risk-fusion weights. Cross-domain generalization was assessed using leave-one-dataset-out evaluation. The model was trained using two datasets and tested on the third without parameter updating.

3.13 Evaluation Metrics and Statistical Analysis

Detection performance was assessed using accuracy, macro-precision, macro-recall, macro-F1, AUROC, AUPRC, false-positive rate, and detection delay. Macro-F1 was calculated as (40):

$$F_{1,macro} = \frac{1}{K}\sum_{k=1}^{K}\frac{2P_kR_k}{P_k + R_k} \cdots (40)$$

where P_k and R_k are the precision and recall of class k. Probability calibration was evaluated through the Brier score (41):

$$BS = \frac{1}{N}\sum_{n=1}^{N}\left(p_n - y_n\right)^2 \cdots (41)$$

where $\hat{p}_n$ is the predicted probability and y_n is the observed label.
Mean detection delay was (42):

$$T_d = \frac{1}{A}\sum_{a=1}^{A} max(0, t_a^{det} - t_a^{start}) \cdots (42)$$

where A is the number of detected attack episodes, t_a^{start} is the actual attack-start time, and t_a^{det} is the first alarm time.
Mitigation effectiveness was quantified as (43):

$$RR = \frac{R_{before} - R_{after}}{R_{before}} \times 100\% \cdots (43)$$

where R_{before} and R_{after} are aggregate risks before and after mitigation.
The results were all presented as mean ± standard deviation (SD) of 10 runs. Wilcoxon signed-rank test was used to test paired differences. Multiple comparisons were controlled by Holm and Cliff's measure of effect size was used. The 10,000 bootstrap resamples were used to create 95% confidence intervals. A p value of <0.05 (adjusted) was considered significant.

4. Results

4.1 Detection and Dynamic Risk-Prioritization Performance

The 10 random seeds were used for all models. The same episode-level partitions were used for training the conventional classifiers, temporal neural networks and the graph-only variant. The references in Table 6 refer to the families of methods that are related; all the numbers were re-calculated in the present experimental protocol, and not copied from the cited studies.

Table 6. Comparative performance of attack-detection and risk-prioritization approaches. Values are mean ± standard deviation over ten runs. Panel A. Attack-detection performance

Method	Accuracy	Macro-F1	AUROC	AUPRC	Normal-class FPR
Logistic Regression	0.7500	0.8456	0.9558	0.8600	0.4648
Random Forest	0.8294 ± 0.0064	0.8168 ± 0.0055	0.9568 ± 0.0014	0.8636 ± 0.0025	0.0930 ± 0.0136
XGBoost	0.8261 ± 0.0087	0.8058 ± 0.0095	0.9531 ± 0.0006	0.8612 ± 0.0022	0.0746 ± 0.0116
MLP	0.8528 ± 0.0109	0.8132 ± 0.0119	0.9407 ± 0.0027	0.8480 ± 0.0035	0.0000 ± 0.0000
LSTM	0.6289 ± 0.1213	0.6815 ± 0.1115	0.9637 ± 0.0110	0.8676 ± 0.0552	0.6746 ± 0.3152
BiGRU	0.6789 ± 0.0599	0.7652 ± 0.0714	0.9719 ± 0.0112	0.9025 ± 0.0631	0.6380 ± 0.1494
GAT-only	0.7144 ± 0.1789	0.7524 ± 0.1318	0.9627 ± 0.0072	0.8607 ± 0.0358	0.4423 ± 0.4074
Proposed CNN–BiGRU–GAT	0.9378 ± 0.0245	0.9387 ± 0.0267	0.9878 ± 0.0095	0.9604 ± 0.0325	0.0296 ± 0.0520

Panel B. Critical-risk prioritization performance

Method	Risk AUROC	Risk AUPRC	High-risk recall	High-risk precision	High-risk F1	False-high-risk rate
Static risk score	0.9495 ± 0.0506	0.9590 ± 0.0378	0.9989 ± 0.0036	0.5745 ± 0.0572	0.7280 ± 0.0457	0.7385 ± 0.1677
Fuzzy-AHP risk score	0.9486 ± 0.0569	0.9568 ± 0.0521	0.9831 ± 0.0132	0.6874 ± 0.1002	0.8046 ± 0.0701	0.4703 ± 0.2433
Proposed learned fusion	0.9059 ± 0.0119	0.8857 ± 0.0181	0.9056 ± 0.0444	0.8322 ± 0.0109	0.8667 ± 0.0183	0.1791 ± 0.0194

The proposed detector's macro-F1 and AUROC were 0.9387 and 0.9878, respectively, as presented in Table 6. It performed 9.31 percentage points better than logistic regression, 12.38 points better than random forest, 13.10 points better than XGBoost, 13.97 points better than MLP, 25.72 points better than LSTM, 17.36 points better than BiGRU and 18.64 points better than GAT-only.

The improvement was significant for all baseline after Holm correction, adjusted (p=0.0137). The delta for Cliff was 1.00 for six comparisons, and 0.96 for GAT-only, showing a large effect in each of the ten runs.
The temporal baselines yielded relatively good AUROC and poor macro-F1 and high false-positive rates. They were able to rank anomalous windows but their class boundaries were quite different among seeds. This was the same with the GAT-only model. The only problem with topology alone was that assets that have the same dependency structure might be attacked at different times.

The other result is shown on Panel B. The static and fuzzy-AHP scores were able to achieve high recall by labeling most of the exposed assets as high risk. This behaviour resulted in two false-high risk rates of 80.22% and 79.12% respectively. This rate was lowered to 17.91% with an increase in high-risk precision to 83.22% with the proposed fusion. It was not recalled as much, but the prioritization was quite selective.

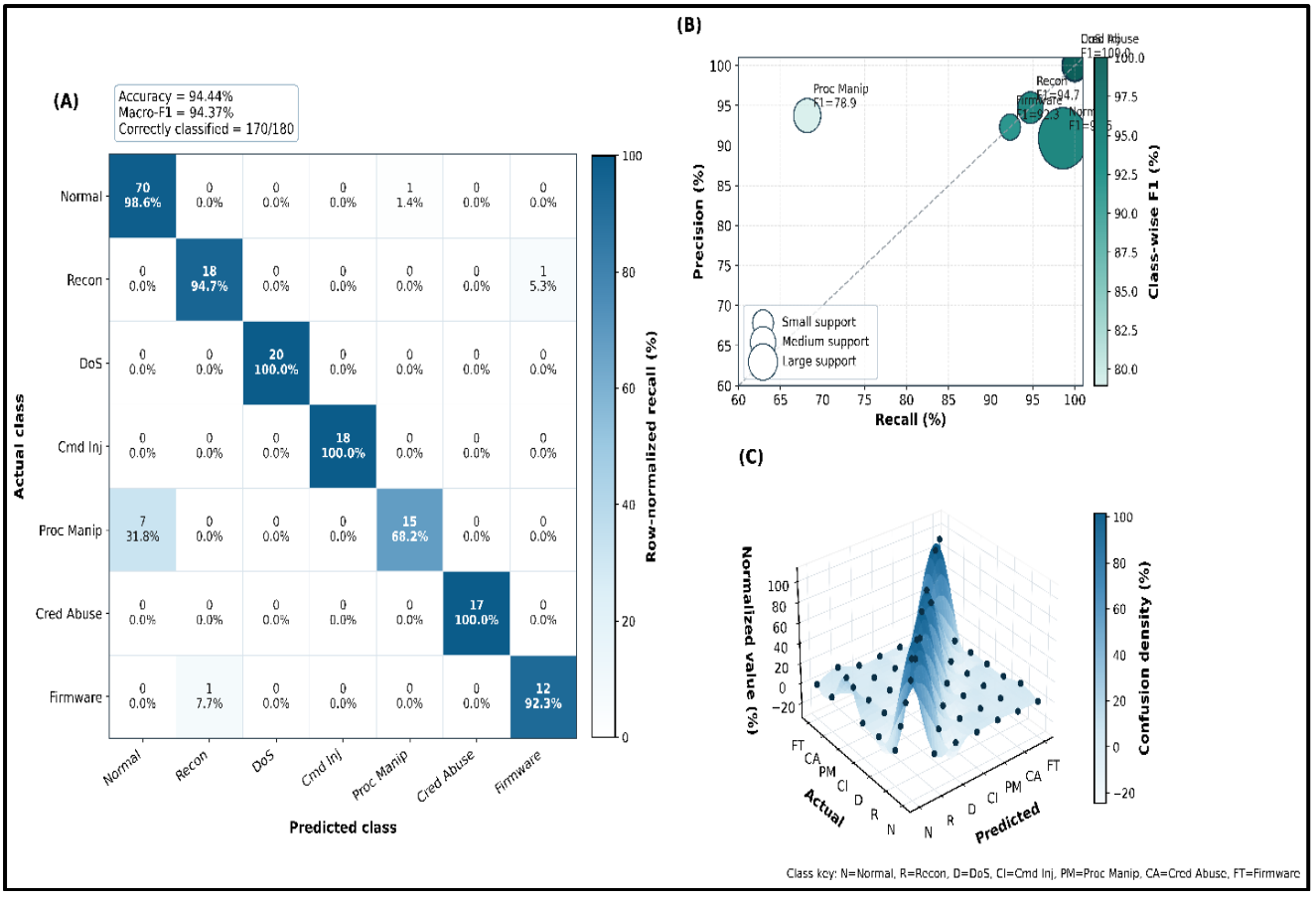

Figure 4. Confusion matrix of the validation-selected proposed model on the independent test partition: (A) row-normalized confusion matrix with raw counts and percentages, (B) class-wise precision–recall bubble map with bubble size proportional to class support and color representing F1-score, and (C) smoothed 3D confusion-density surface showing the dominance of diagonal class separation and the concentration of misclassification around process manipulation.

The model chosen based on the validation performance had 170/180 test episodes correctly classified. Its test accuracy was 94.44%, and macro-F1 was 94.45%. All DoS, command injection and credential abuse attacks were detected. The recall was 98.59% for normal operation, 94.74% for reconnaissance and 92.31% for firmware tampering. Manipulation of the process was the most challenging category. The correct classification of 15 episodes and normal operation of 7 episodes were obtained out of 22 episodes. There was one normal episode that was misclassified as process manipulation. Confusion suggests that there are physical changes that can be subtle and look like valid process changes, especially if the network signature of the process is not strong.

4.2 Quantum-Horizon and Asset-Level Exposure

The impact of the assumed horizon for quantum computing is shown in Figure 5. The operational traces were kept the same and the exposure was recalculated for 5, 10 and 15 years horizons.

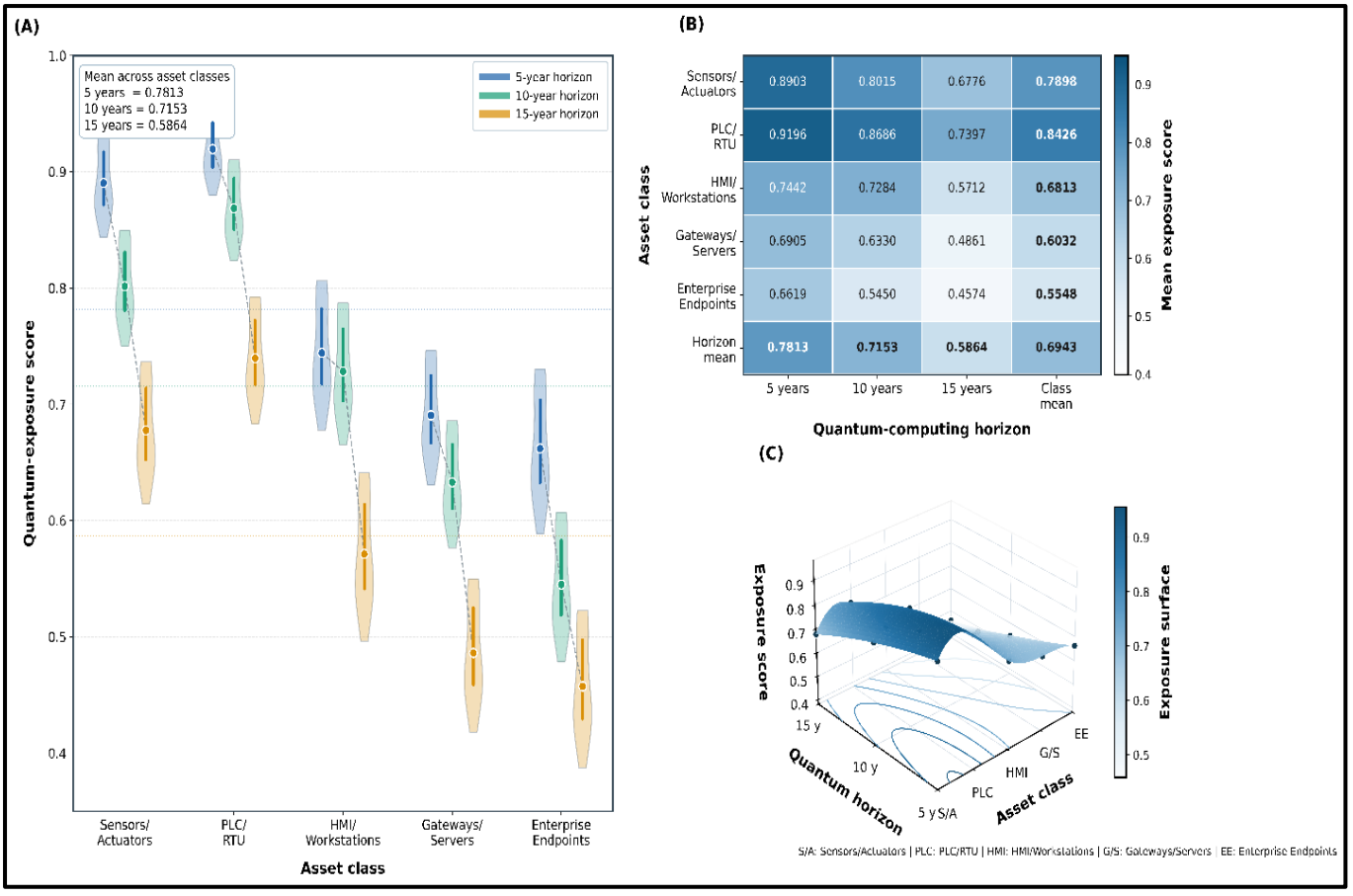

Figure 5. Quantum-exposure distribution across asset classes under 5-, 10-, and 15-year quantum-computing horizons.

Values are the product of algorithm vulnerability, the length of time that the algorithm is confidential, and how easy it is to migrate. The average exposure of the asset classes was 0.7813 at 5 years, 0.7153 at 10 years and 0.5864 at 15 years. This decrease was anticipated as it is a longer time frame for cryptographic migration. It didn't impact all asset classes, however. The highest exposure was for the PLC and RTU assets. Their mean scores were 0.9196, 0.8686, and 0.7397 across the three horizons. The scores for sensors and actuators dropped from 0.8903 to 0.6776, but are still higher than the scores for enterprise endpoints. The latter decreased from 0.6619 to 0.4574.

The result is in favor of staged migration to quantum-safe algorithms instead of the uniform algorithm replacement [9], [19], [41]. The need for resource-constrained controllers and field devices was still critical, as the lack of flexibility in protocols, reliance on vendors and limited maintenance windows made them even more vulnerable.

4.3 Component-Wise Ablation

Table 7 compares the observed improvement with the temporal, process, graph, and quantum-risk evidence, to determine if it was one of these factors or all of them.

Table 7. Component-wise ablation of the proposed framework.

Variant	Macro-F1	Risk AUROC	High-risk recall
Full framework	0.9387 ± 0.0267	0.9059 ± 0.0119	0.9056 ± 0.0444
Without dependency propagation	0.9399 ± 0.0246	0.8937 ± 0.0125	0.8371 ± 0.0891
Without local temporal variation	0.8992 ± 0.0310	0.8951 ± 0.0133	0.7506 ± 0.0299
Without process telemetry	0.8091 ± 0.0261	0.8400 ± 0.0074	0.7719 ± 0.0254

Without quantum exposure	0.9387 ± 0.0267	0.8919 ± 0.0139	0.8483 ± 0.1019

The highest classification loss was due to the removal of process telemetry. Macro-F1 was found to be 12.96 percentage points lower, which was consistent with the results of the network evidence alone being not enough to distinguish physical manipulation from ordinary communication activity.

Local temporal variation was suppressed, which led to a decrease of 3.95 points in macro-F1 and 15.50 points in high-risk recall. The temporal representation was thus crucial to determine the course of an event in the 30-s observation time.
The recognition of attack-classes was not significantly affected by dependency propagation. It was removed and the macro-F1 was slightly changed from 0.9387 to 0.9399. However, the AUROC of Risk decreased by 1.22 points and the AUROC of high-risk recall decreased by 6.85 points. The graph layer was thus used to propagate impact and prioritize instead of just improving the classification accuracy.

A similar separation was also found in the quantum-exposure layer. The removal of it did not change macro-F1 as it does not affect the observed attack class, as the cryptographic profiles do not affect the attack class. Risk AUROC decreased from 0.9059 to 0.8919, while high-risk recall fell from 0.9056 to 0.8483. This verifies that there was a contribution of distinct risk information from the PQC layer.

4.4 Cross-Domain Generalization, Unseen Attacks, and Robustness

Table 8 consolidates the generalization and robustness experiments promised in the methodology. Cross-domain models were trained on two infrastructure domains and tested on the third without retraining. For unseen-family evaluation, one attack family was removed from training and validation; the model was then tested on that family and normal episodes as a binary open-set task.

Table 8. Cross-domain, unseen-attack, and robustness evaluation. Values are mean ± standard deviation over ten runs. Panel A. Leave-one-domain-out generalization

Held-out domain	Macro-F1	AUROC	AUPRC
SWaT	0.9682 ± 0.0442	0.9995 ± 0.0011	0.9971 ± 0.0071
HAI	0.9691 ± 0.0124	0.9969 ± 0.0037	0.9891 ± 0.0085
ToN-IoT	0.5473 ± 0.1262	0.9381 ± 0.0245	0.8250 ± 0.0433

Panel B. Leave-one-attack-family-out evaluation

Unseen family	Binary F1	AUROC	Attack recall
Reconnaissance	0.5922 ± 0.2741	0.9912 ± 0.0109	0.9895 ± 0.0222
Denial of service	0.7040 ± 0.2668	1.0000 ± 0.0000	1.0000 ± 0.0000
Command injection	0.6913 ± 0.2498	1.0000 ± 0.0000	1.0000 ± 0.0000
Process manipulation	0.4963 ± 0.1851	0.8060 ± 0.0313	0.9045 ± 0.1552
Credential abuse	0.4201 ± 0.2608	1.0000 ± 0.0000	1.0000 ± 0.0000
Firmware tampering	0.6794 ± 0.3013	0.9948 ± 0.0114	0.9769 ± 0.0372

Panel C. Perturbation robustness

Perturbation	Level	Primary metric	Secondary metric	Additional metric
Telemetry loss	10%	Macro-F1: 0.9111 ± 0.0195	AUROC: 0.9861 ± 0.0105	FPR: 0.0197 ± 0.0383
Telemetry loss	20%	Macro-F1: 0.8827 ± 0.0316	AUROC: 0.9821 ± 0.0095	FPR: 0.0155 ± 0.0328
Telemetry loss	30%	Macro-F1: 0.8653 ± 0.0233	AUROC: 0.9806 ± 0.0107	FPR: 0.0085 ± 0.0178
Telemetry loss	40%	Macro-F1: 0.7941 ± 0.0184	AUROC: 0.9694 ± 0.0094	FPR: 0.0070 ± 0.0152
Graph-edge removal	10%	Macro-F1: 0.9387 ± 0.0267	AUROC: 0.9878 ± 0.0095	FPR: 0.0296 ± 0.0520

Graph-edge removal	20%	Macro-F1: 0.9387 ± 0.0267	AUROC: 0.9878 ± 0.0095	FPR: 0.0296 ± 0.0520
Graph-edge removal	30%	Macro-F1: 0.9396 ± 0.0242	AUROC: 0.9877 ± 0.0095	FPR: 0.0296 ± 0.0520
Graph-edge removal	40%	Macro-F1: 0.9387 ± 0.0267	AUROC: 0.9878 ± 0.0095	FPR: 0.0296 ± 0.0520
Threat-intelligence delay	30 s	Risk AUROC: 0.9040 ± 0.0124	Risk AUPRC: 0.8801 ± 0.0192	Recall: 0.9067 ± 0.0397
Threat-intelligence delay	60 s	Risk AUROC: 0.9027 ± 0.0123	Risk AUPRC: 0.8761 ± 0.0191	Recall: 0.9090 ± 0.0361
Threat-intelligence delay	90 s	Risk AUROC: 0.9012 ± 0.0116	Risk AUPRC: 0.8728 ± 0.0190	Recall: 0.9101 ± 0.0347
Alternative crypto profiles	10 registries	Risk AUROC: 0.9212 ± 0.0170	Risk AUPRC: 0.9065 ± 0.0289	Rank (\rho): 0.9848 ± 0.0061

The leave-one-domain-out results were not uniform. When SWaT or HAI was held out, macro-F1 remained close to 0.97. Testing on ToN-IoT reduced macro-F1 to 0.5473, although AUROC remained 0.9381. The detector therefore retained broad anomaly-ranking ability, but the learned class boundaries did not transfer fully to the more network-centric ToN-IoT domain.

This result identifies a genuine domain-shift limitation. SWaT and HAI contain strongly structured process signals, whereas ToN-IoT includes more heterogeneous device and network behaviour. Direct deployment across these environments would require domain adaptation, feature recalibration, or a small labelled target-domain sample. The unseen-family experiment produced a related pattern. AUROC remained above 0.99 for five of the six withheld families. Process manipulation was harder, with an AUROC of 0.8060. Thresholded F1 was more variable than AUROC, despite high attack recall. The fixed threshold was able to identify the majority of the unknown attacks, but also identified extra normal episodes as anomalies. So, the representation was more general than the operating threshold.

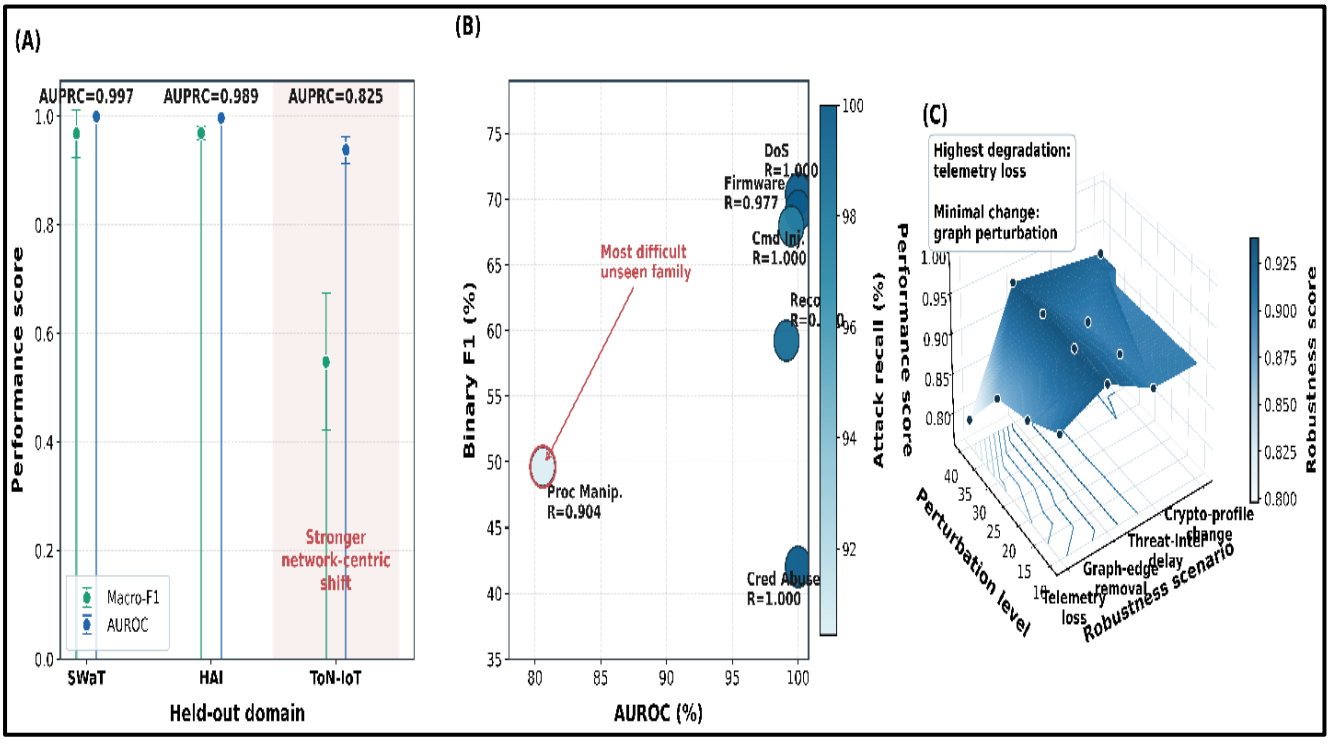

Figure 6. Generalization and robustness of the proposed framework: (a) leave-one-domain-out macro-F1 and AUROC, (b) unseen-family detection, and (c) performance under telemetry loss, graph perturbation, delayed threat intelligence, and cryptographic-profile changes.

Telemetry loss produced a gradual degradation. Macro-F1 fell by 2.76 percentage points at 10% loss, 5.60 points at 20%, 7.34 points at 30%, and 14.46 points at 40%. The AUROC was not below 0.96 even at the highest loss level, suggesting that the AUROC was not as affected as the exact class assignment.

Removing 10–40% of dependency edges did not materially alter classification performance. This observation is consistent with the ablation results: the graph layer primarily affects propagation-aware risk rather than local attack labels. Delayed threat context had a limited effect. Risk AUROC declined from the unperturbed value of 0.9059 to 0.9012 at 90 s. The ten alternative cryptographic registries produced a mean rank correlation of 0.9848 with the reference ordering. Asset priorities were therefore stable even when the deterministic profile-assignment seed changed.

4.5 Risk-Constrained Mitigation

The optimization stage was evaluated under three operational policies. The selected actions included cryptographic migration, key rotation, segmentation, access control, rate limiting, isolation, firmware hardening, and rollback.

Table 9. Optimized mitigation plans under alternative budget and downtime policies.

Scenario	Budget limit	Downtime limit	Risk before	Risk after	Risk reduction (%)	Cost used	Downtime used	Selected asset–control actions
Constrained	8	5	2637.01	975.93	62.99	8.00	2.59	47
Balanced	12	7	2637.01	515.93	80.43	11.99	3.41	61
Resilience-first	18	10	2637.01	71.62	97.28	17.99	4.04	78

All three mixed-integer programs reached an optimal solver status. Under the constrained policy, the framework reduced aggregate risk by 62.99% through 47 asset–control assignments. Low-cost controls, including key rotation, rate limiting, segmentation, and selective signing, were favoured. The balanced policy reduced risk by 80.43%. The additional budget enabled broader use of hybrid ML-KEM, ML-DSA signing, AES-256 migration, and firmware protection. Under the resilience-first policy, residual risk fell to 71.62, corresponding to a 97.28% reduction.

As shown in Figure 7, the optimizer did not consume downtime merely because a larger limit was available. Even in the resilience-first setting, only 4.04 of the permitted 10 downtime units were used. Actions were selected when their expected risk reduction justified their operational burden. These findings extend fixed mitigation rankings [16] and context-based countermeasure selection [29]. They also support the use of dynamic key management and signed control operations in industrial environments [21]. The principal difference is that the proposed method enforces compatibility, budget, latency, and downtime constraints within the same decision problem.

4.6 Explanation Faithfulness and Dependency Paths

Feature-level and graph-level explanations were evaluated separately. Figure 8 reports the change in attack probability after masking each temporal feature, together with the faithfulness obtained by removing the highest-ranked variables. Bars represent mean absolute changes in attack probability, while markers show the risk reduction obtained after masking the top-ranked features.

The most important temporal variable was rate of change with a mean absolute probability change of 0.0234. The ratio of unusual ports and SYN rate were 0.0133 and 0.0130 respectively. Other factors included were authentication failures, control-loop error, actuator state, sensor values and residual error. The decision making process was thus based on network and physical-process evidence.

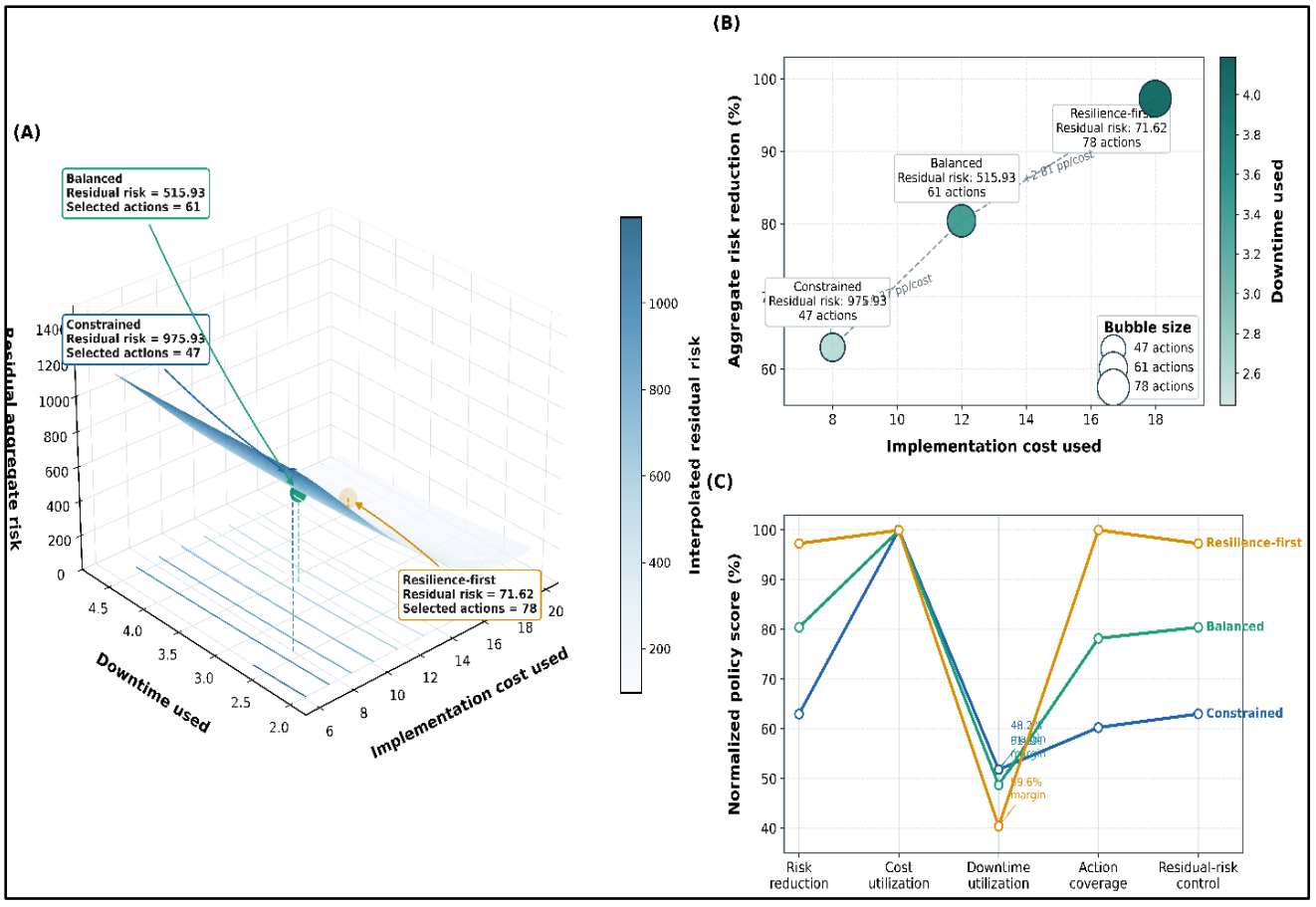

Figure 7. Mitigation-policy trade-offs in residual risk, implementation cost, downtime, and selected actions: (A) 3D risk landscape, (B) Pareto bubble comparison, and (C) normalized policy profiles.

The top three features when masked decreased the average risk by 14.17%. This drop was further lowered to 32.63% (top 5) and 51.01% (top 10). These changes indicate that the variables ranked were indeed important to the model and not just for the sake of an attractive explanation.

The highest evidence score for a dependency path was 0.7488. The most robust paths were those that included communication and control, trust or service relationships. The graph layer identified where the increased risk could propagate, while SHAP-based evidence identified which observations were causing the increased risk. The separation is significant for the critical-infrastructure decision support in a transparent manner [18], [36], [44].

4.7 Main Findings

The results demonstrate that the framework's performance was independent of a single module. The primary classification gain was from the temporal and process evidence. Critical-risk prioritization was enhanced with quantum profiles and graph propagation, while not impacting attack-class scores. Those scores were then translated into viable protection plans in the optimization stage.

The proposed detector obtained the mean macro-F1 of 0.9387 and AUROC of 0.9878. The cross domain transfer was still good for the process oriented domains, but the ToN-IoT result revealed a significant distribution-shift problem. The attacks were generally ranked well for unseen attacks, but less well for process manipulation and credential abuse for threshold calibration. Lastly, the mitigation layer was able to decrease the overall risk by 62.99–97.28% depending on the operational resources available.

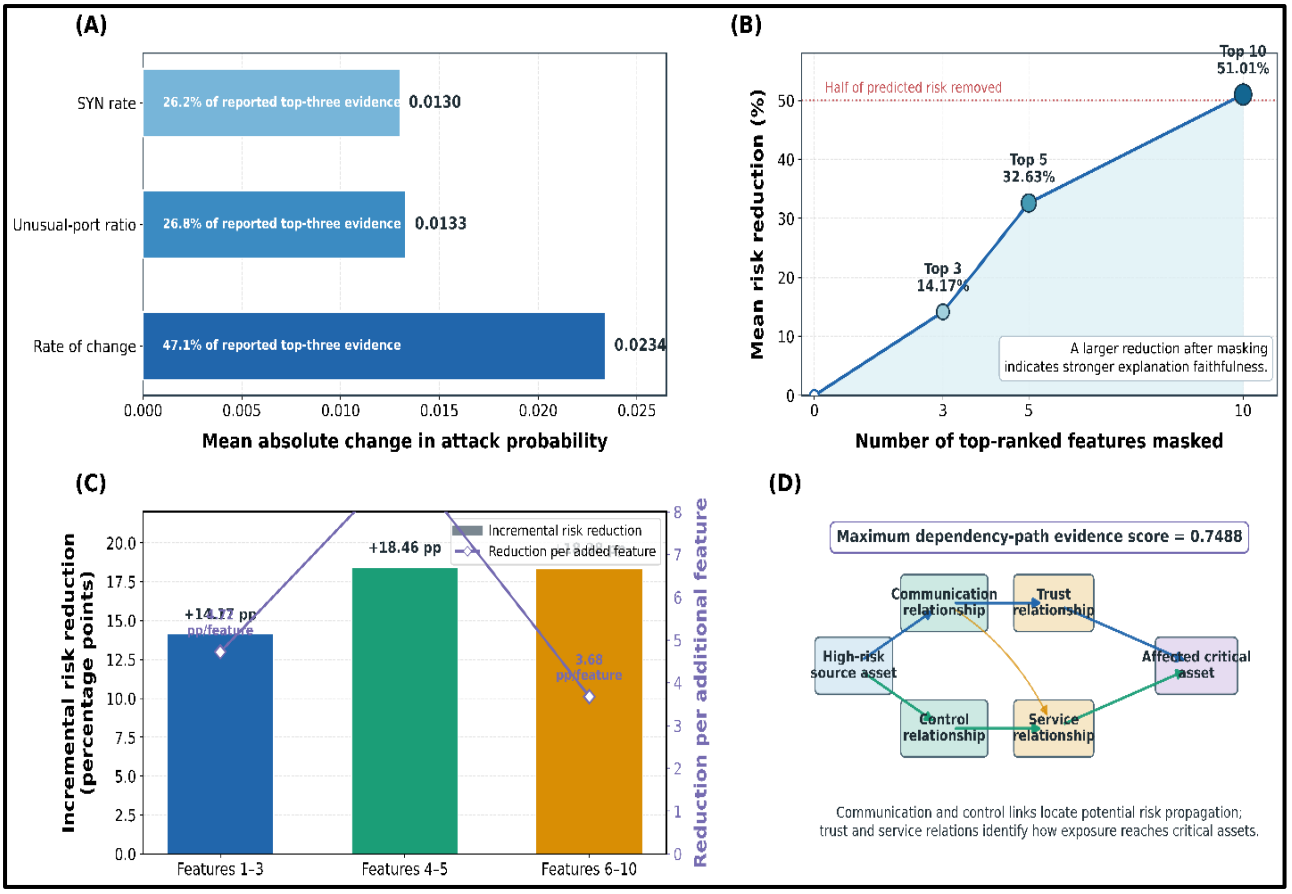

Figure 8. Explanation and faithfulness analysis of the proposed framework. (A) Mean absolute change in attack probability for the three quantitatively reported temporal variables, (B) cumulative reduction in predicted risk after masking the top-ranked features, (C) incremental faithfulness contribution and average reduction per additionally masked feature, and (D) representative dependency-path evidence showing how communication, control, trust, and service relationships connect a high-risk source asset to an affected critical asset.

5. Discussion

The results show that protection alone by enhancing intrusion detection is not sufficient to provide effective protection in a quantum network. Most of the classification gain came from the temporal and process branches while the dependency graph and post-quantum layer mainly enhanced the ability to identify assets that need immediate attention. The separation is significant because an asset could exhibit some anomalous behavior, but not be strategically vulnerable due to weak cryptography, long data retention time or its location in a control network.

Unlike hybrid cryptographic communication schemes [9] the proposed framework goes beyond replacing the algorithm with the consideration of feasibility of migration and impact of the operation. It also takes a step forward in the field of cyber–physical risk visualization [15] by providing an update of the risk based on the observed telemetry and not only on static dependency scores. The framework connects the evidence of the features, the propagation paths, and possible mitigation actions in one decision process, whereas explainable intrusion detectors are mainly concerned with class predictions [36].

A practical implication is that infrastructure operators can upgrade their infrastructure in a crypto-agile manner without having to upgrade all the vulnerable devices at once. An interesting finding was that the deletion of graph or quantum attributes did not significantly impact the attack classification, but did significantly reduce the ability to prioritize attacks as high-risk. This has been confirmed as a measure of detection accuracy is not sufficient to measure infrastructure resilience.

The study is still constrained by the controlled benchmark, deterministic cryptographic-profile assignment, three infrastructure domains and a criticality target based on rules. It is therefore necessary to validate on authorized operational traces, live testbeds and on other sectors of the infrastructure. Future research should also focus on domain adaptation, threat intelligence calibration and on-going threat intelligence updates. The framework also relates to the migration to post-quantum, explainable AI, cyber–physical resilience and risk-aware infrastructure governance in a broader context.

6. Conclusion and Future Scope

The conclusion and future scope of the project is presented in this section. This research proposed an AI-based approach to threat analysis of quantum computing and prioritizing cybersecurity risk mitigation of critical infrastructure. The design proposed was a temporal attack detection, cyber–physical dependency modelling, post-quantum exposure assessment, explainable risk fusion and constrained mitigation optimization. The results indicated that process and temporal features played a central role in the recognition of attacks, whereas the quantum profiles and dependency propagation features played a stronger role in prioritizing the assets that were at high risk. It also retained a valuable discrimination in the event of a loss of telemetry, graph perturbation, delayed threat intelligence and alternative cryptographic profiles. One of the biggest features is the ability to distinguish between attacks and the urgency to migrate. This enables infrastructure operators to discover assets that might not be immediately compromised, but are still vulnerable due to legacy cryptography, long confidentiality periods or hard-to-replace assets. The optimization layer additionally converts risk estimates to realistic controls within budget, downtime and compatibility constraints. The framework needs to be validated in the future with authorized operational data and live cyber–physical testbeds. Testing of the transport, healthcare, telecommunications and financial infrastructure would also enhance the generalizability.

Declarations

Acknowledgements
NA

Funding
NA

Contributions
All authors; Conceptualization, All authors; Investigation; All authors; Writing (Original Draft), All authors; Writing (Review and Editing) Supervision, All authors; Project Administration.

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

Data Availability Statement
The public datasets used to create this study, Secure Water Treatment (SWaT), HAI Security Dataset and ToN-IoT are accessible via the SUTD iTrust laboratory, the official HAI repository, and UNSW Canberra, respectively, under the terms of their access and usage. The following processed data, feature mappings, model configurations, prediction outputs, explanation scores, risk-assessment results and validation scripts are provided in this study. The accompanying repository contains all the procedures needed to verify the reported experiments, such as fixed random seeds and integrity checks

Use of Generative Artificial Intelligence
The authors declare that no generative artificial intelligence tools were used in the preparation of this manuscript.

Competing interests
All authors declare no competing interests.

References

[1] Alandjani, G. (2025). A MARL-federated blockchain-based quantum secure framework for trust management in Industrial Internet of Things. *Scientific Reports, 15*, Article 39149. https://doi.org/10.1038/s41598-025-23055-2
[2] Reddy, N. R., Suryadevara, S., Reddy, K. G. R., Umamaheswari, R., Guttula, R., & Kotoju, R. (2025). Quantum secured blockchain framework for enhancing post quantum data security. *Scientific Reports, 15*(1), 31048. https://doi.org/10.1038/s41598-025-16315-8

[3] Hdaib, M., Rajasegarar, S., & Pan, L. (2024). Quantum deep learning-based anomaly detection for enhanced network security. *Quantum Machine Intelligence, 6*, Article 26. https://doi.org/10.1007/s42484-024-00163-2
[4] Rajpal, T. S., & Naithani, A. (2025). NeuroCrypt: a neuro symbolic AI ecosystem for advanced cryptographic data security and transmission. *IEEE Transactions on Artificial Intelligence*, *7*(1), 512-521.
[5] Vignes, V. M., MP, S. H., Satheesh, R., Das, V., & Padmanaban, S. (2025). AI-driven cybersecurity framework for anomaly detection in power systems. *Scientific Reports*, *15*, 35506. https://doi.org/10.1038/s41598-025-19634-y
[6] Lakhal, H., Zegrari, M., & Bahnasse, A. (2025). Next-generation smart grid cybersecurity: A systematic review of OT cyber threats, AI-driven defense, cyber deception techniques, and emerging security strategies. *IEEE Access, 13*, 181519–181552. https://doi.org/10.1109/ACCESS.2025.3623834
[7] Rodriguez-Casavilca, H. M., Mauricio, D., & Villanueva, J. M. M. (2025). Evolution of artificial intelligence-based OT cybersecurity models in energy infrastructures: Services, technical means, facilities and algorithms. *Energies, 18*(19), Article 5163. https://doi.org/10.3390/en18195163
[8] Raval, K. J., Jadav, N. K., Rathod, T., Tanwar, S., Vimal, V., & Yamsani, N. (2024). A survey on safeguarding critical infrastructures: Attacks, AI security, and future directions. *International Journal of Critical Infrastructure Protection, 44*, Article 100647. https://doi.org/10.1016/j.ijcip.2023.100647
[9] Rubio García, C., Cano Aguilera, A., Stan, C., Vegas Olmos, J. J., Rommel, S., & Tafur Monroy, I. (2025). Enhanced network security protocols for the quantum era: Combining classical and post-quantum cryptography, and quantum key distribution. *IEEE Journal on Selected Areas in Communications, 43*(8), 2765–2781. https://doi.org/10.1109/JSAC.2025.3568011
[10] Teryak, H., Albaseer, A., Abdallah, M., Al-Kuwari, S., & Qaraqe, M. (2023). Double-edged defense: Thwarting cyber attacks and adversarial machine learning in IEC 60870-5-104 smart grids. *IEEE Open Journal of the Industrial Electronics Society, 4*, 629–642. https://doi.org/10.1109/OJIES.2023.3336234
[11] Houmb, S. H., Iversen, F., Ewald, R., & Færaas, E. (2023). Intelligent risk-based cybersecurity protection for industrial systems control—A feasibility study. *SPE Journal, 28*, 3272–3279. https://doi.org/10.2118/217430-PA
[12] Taherdoost, H., Le, T. V., & Slimani, K. (2025). Cryptographic techniques in artificial intelligence security: A bibliometric review. *Cryptography*, *9*(1), 17.
[13] Pawlik, L., Wilk-Jakubowski, J. L., Grabski, P. T., & Wilk-Jakubowski, G. (2025). Securing the electrified future: A systematic review of cyber attacks, intrusion and anomaly detection, and authentication in electric vehicle charging infrastructure. *Energies, 18*, Article 4847. https://doi.org/10.3390/en18184847
[14] Jiang, Y., Jeusfeld, M. A., Ding, J., & Sandahl, E. (2023). Model-Based Cybersecurity Analysis: Y. Jiang et al. *Business & Information Systems Engineering*, *65*(6), 643-676. https://doi.org/10.1007/s12599-023-00811-0
[15] Keenan, C., Maier, H. R., van Delden, H., & Zecchin, A. C. (2024). Bridging the cyber-physical divide: A novel approach for quantifying and visualising the cyber risk of physical assets. *Water, 16*, Article 637. https://doi.org/10.3390/w16050637
[16] Almotiri, S. H. (2024). Quantum-resilient software security: A fuzzy AHP-based assessment framework in the era of quantum computing. *PLoS ONE, 19*(12), Article e0316274. https://doi.org/10.1371/journal.pone.0316274
[17] Tripathi, S. M., Upadhyay, H., & Soni, J. (2023, December). Quantum neural network classification-based cyber threat detection in virtual environment. In *2023 international conference on computational science and computational intelligence (CSCI)* (pp. 391-396). IEEE.
[18] Sarker, I. H., Janicke, H., Ferrag, M. A., & Abuadbba, A. (2024). Multi-aspect rule-based AI: Methods, taxonomy, challenges and directions towards automation, intelligence and transparent cybersecurity modeling for critical infrastructures. *Internet of Things, 25*, Article 101110. https://doi.org/10.1016/j.iot.2024.101110
[19] Kong, M., Janssen, M., & Bharosa, N. (2024). Realizing quantum-safe information sharing: Implementation and adoption challenges and policy recommendations for quantum-safe transitions. *Government Information Quarterly, 41*(1), Article 101884. https://doi.org/10.1016/j.giq.2023.101884
[20] Faruk, M. J. H., Tahora, S., Tasnim, M., Shahriar, H., & Sakib, N. (2022, May). A review of quantum cybersecurity: threats, risks and opportunities. In *2022 1st International Conference on AI in Cybersecurity (ICAIC)* (pp. 1-8). IEEE.
[21] Dokku, N. S., David Amar Raj, R., Bodapati, S. K., Pallakonda, A., Reddy, Y. R. M., & Krishna Prakasha, K. (2025). Resilient cybersecurity in smart grid ICS communication using BLAKE3-driven dynamic key rotation and intrusion detection. *Scientific Reports*, *15*(1), 32754. https://doi.org/10.1038/s41598-025-17530-z
[22] Harper, B., Tonekaboni, B., Goldozian, B., Sevior, M., & Usman, M. (2025). Crosstalk attacks and defence in a shared quantum computing environment. *Advanced Quantum Technologies*, *8*(10), Article e2500009. https://doi.org/10.1002/qute.202500009
[23] Koduru, S., Machina, V. S. P., & Madichetty, S. (2023). Cyber attacks in cyber-physical microgrid systems: A comprehensive review. *Energies, 16*, Article 4573. https://doi.org/10.3390/en16124573
[24] Alqudhaibi, M., Albarrak, M., Aloseel, A., Jagtap, S., & Salonitis, K. (2023). Predicting cybersecurity threats in critical infrastructure for Industry 4.0: A proactive approach based on attacker motivations. *Sensors, 23*, Article 4539. https://doi.org/10.3390/s23094539
[25] Mtukushe, N., Onaolapo, A. K., Aluko, A., & Dorrell, D. G. (2023). Review of cyberattack implementation, detection, and mitigation methods in cyber-physical systems. *Energies, 16*, Article 5206. https://doi.org/10.3390/en16135206
[26] Namakshenas, D., Yazdinejad, A., Dehghantanha, A., & Srivastava, G. (2024). Federated quantum-based privacy-preserving threat detection model for consumer internet of things. *IEEE Transactions on Consumer Electronics*, *70*(3), 5829-5838.
[27] Al Siam, M., Alazab, M., Awajan, A., & Faruqui, N. (2025). A comprehensive review of AI's current impact and future prospects in cybersecurity. *IEEE Access, 13*, 14029–14050. https://doi.org/10.1109/ACCESS.2025.3528114

[28] Khayat, M., Barka, E., Serhani, M. A., Sallabi, F., Shuaib, K., & Khater, H. M. (2025). Empowering security operation center with artificial intelligence and machine learning—a systematic literature review. *IEEE Access*, *13*, 19162-19197.
[29] Giarimpampa, D., Meier, R., Bissyande, T. F., Lenders, V., & Klein, J. (2025). Exploring the role of artificial intelligence in enhancing security operations: A systematic review. *ACM Computing Surveys*, *58*(3), 1-38.
[30] Tripathi, S., Upadhyay, H., & Soni, J. (2025). A Quantum LSTM-based approach to cyber threat detection in virtual environment. *The Journal of Supercomputing*, *81*(1), 142.
[31] Alluhaibi, R. (2024). Quantum Machine Learning for Advanced Threat Detection in Cybersecurity. *International Journal of Safety & Security Engineering*, *14*(3), 875.
[32] Yalçın, N., Çakır, S., & Ünaldı, S. (2024). Attack detection using artificial intelligence methods for SCADA security. *IEEE Internet of Things Journal*, *11*(24), 39550–39559. https://doi.org/10.1109/JIOT.2024.3447876
[33] Dhivyasree, T., & Prabu, M. (2025, August). Quantum-enhanced adaptive defense system for network threat detection. In *2025 Third International Conference on Networks, Multimedia and Information Technology (NMITCON)* (pp. 1-6). IEEE.
[34] Pinto, L.-C., Herrera, L.-C., Donoso, Y., & Gutierrez, J. A. (2023). Survey on intrusion detection systems based on machine learning techniques for the protection of critical infrastructure. *Sensors*, *23*, Article 2415. https://doi.org/10.3390/s23052415
[35] Adabara, I., Sadiq, B. O., Shuaibu, A. N., Danjuma, Y. I., & Venkateswarlu, M. (2025). A review of agentic AI in cybersecurity: cognitive autonomy, ethical governance, and quantum-resilient defense. *F1000Research*, *14*, 843. https://doi.org/10.12688/f1000research.169337.1
[36] Javeed, D., Gao, T., Kumar, P., & Jolfaei, A. (2024). An explainable and resilient intrusion detection system for Industry 5.0. *IEEE Transactions on Consumer Electronics*, *70*(1), 1342–1350. https://doi.org/10.1109/TCE.2023.3283704
[37] Hussain, N., Li, S., Hussain, A., Ullah, Z., & Jamjoom, M. (2025). Quantum-aware secure blockchain intrusion detection system for industrial IoT networks. *Scientific Reports*, *16*(1), 2265.
[38] López, D. E., Pinilla, M. A., & Mora, H. R. C. (2025). A novel risk-based methodology for enhancing industrial control systems security: A systematic review and case study. *IEEE Access*, *13*, 160313–160339. https://doi.org/10.1109/ACCESS.2025.3609252
[39] Zornetta. (2024). Quantum-safe global encryption policy. *International Journal of Law and Information Technology*, *32*, Article eaae020. https://doi.org/10.1093/ijlit/eaae020
[40] Shan, & Myeong, S. (2024). Proactive threat hunting in critical infrastructure protection through hybrid machine learning algorithm application. *Sensors*, *24*, Article 4888. https://doi.org/10.3390/s24154888
[41] Alomari, & Kumar, S. A. P. (2024). Securing IoT systems in a post-quantum environment: Vulnerabilities, attacks, and possible solutions. *Internet of Things*, *25*, Article 101132. https://doi.org/10.1016/j.iot.2024.101132
[42] Czaja, P., Gdowski, B., Niemiec, M., Mees, W., Stoianov, N., Votis, K., ... & Merialdo, M. (2025). Cybersecurity challenges and opportunities of machine learning-based artificial intelligence. *Neural Computing and Applications*, *37*(33), 27931-27956. https://doi.org/10.1007/s00521-025-11604-9
[43] Begum, S. H., Mohammed, I., Rahman, S., & Khan, M. O. W. (2025, December). Quantum Computing for Cyber Threat Detection. In *2025 4th International Conference on Applied Artificial Intelligence and Computing (ICAAIC)* (pp. 22-27). IEEE.
[44] Shobitha, S. A., Thahniyath, G., & Radhika, K. (2025, November). Enhancing Cyber Threat Prediction Through Quantum-Enabled Machine Learning Models. In *2025 2nd International Conference on Intelligent Systems for Cybersecurity (ISCS)* (pp. 1-7). IEEE.
[45] Pudney, S., Mills, D., et al. (2025). Evaluation of artificial intelligence-enhanced critical infrastructure systems: A conceptual framework. *Evaluation*, *31*(3). https://doi.org/10.1177/13563890251350677