

EPiC Series in Computing

Volume 106, 2025, Pages 176–186

Proceedings of the 20th Asia Joint Conference on Information Security



AI-Driven Defense Mechanism for Lattice-Based Post-Quantum Cryptography: Adaptive Mitigation Against Side-Channel Attacks

Wibby Aldryani Astuti Praditasari¹, Hyungyeop Kim², Hyejin Yoon³, Danang Rimbawa⁴ and Okyeon Yi⁵

1,2,3,5 Kookmin University, Seoul, South Korea.

4 Universitas Pertahanan, Sentul, Indonesia.
wibby.praditasari@gmail.com, wibbypraditasari@kookmin.ac.kr

Abstract

Post-quantum cryptography (PQC) offers resistance against quantum adversaries. This study's practical implementations remain vulnerable to side-channel attacks (SCAs) that exploit timing, power, or electromagnetic leakage. In this study, we introduce an unsupervised, resource-efficient anomaly detection framework tailored to the unique constraints of post-quantum cryptography (PQC) systems. Unlike traditional methods that rely on labeled attack traces or algorithm-specific profiling, our approach leverages an autoencoder trained solely on benign traces to learn deep latent representations of normal cryptographic behavior. The system flags deviations using reconstruction error and supports multiple PQC schemes, including Kyber and Dilithium, without retraining. Experimental results demonstrate an average classification accuracy of 98.1%, with a false positive rate of 0.7% and a false negative rate of 0.4%. Under adversarial perturbation and Gaussian noise, the model maintains an AUC-ROC of 1.00, confirming its robustness. Additionally, ablation studies across CNN, GRU, and Transformer architectures validate the autoencoder's superior trade-off between accuracy and latency, achieving an inference time of 0.036 ms and a model size of only 0.11 MB. This enables real-time deployment on constrained devices without sacrificing security. The proposed solution marks a step forward in scalable, adaptive post-quantum defenses and opens new directions for cryptographic anomaly detection with minimal overhead. This framework is deployable on real-world POC-enabled IoT systems.

Keyword: Post-Quantum Cryptography (PQC), Side-Channel Attack Detection, Unsupervised Anomaly Detection, Autoencoder-Based Security, Lightweight Real-Time Defense

1 Introduction

The cryptographic community is rapidly transitioning to Post-Quantum Cryptography (PQC) in response to emerging quantum computing threats. Among the most prominent candidates, lattice-based schemes such as Kyber and FrodoKEM offer strong theoretical resistance to quantum attacks. However, despite their mathematical security, these implementations remain physically vulnerable to side-channel attacks (SCA), which exploit timing, power, or electromagnetic emissions to leak sensitive information. Existing countermeasures—such as masking, hiding, or constant-time execution—tend to be static, hardware-dependent, and computationally intensive, making them impractical for dynamic or resource-constrained environments. To address this, we propose a lightweight and adaptive anomaly detection system powered by Artificial Intelligence (AI). Our framework learns normal cryptographic behavior from unlabeled traces and flags deviations in real-time, without requiring prior knowledge of the attack vector. This general-purpose, modular AI layer introduces a novel direction in PQC defense, offering scalable protection against physical attacks while remaining deployable on accessible platforms like Google Colab.

2 Related Work

2.1 Traditional methods

Recent research has extensively explored the use of machine learning for side-channel attack (SCA) detection, particularly leveraging autoencoders, LSTM models, and adversarial learning. While profiling-based methods such as CNNs and LSTMs [1][2][4] have shown promise, they often require labeled data and fail to generalize across post-quantum cryptographic schemes. Several studies have proposed unsupervised or hybrid learning models, integrating deep learning with quantum or generative components to improve adaptability and detection accuracy. A summary of key related works is presented in Table I, highlighting their methodological focus and core limitations in the context of PQC implementation security

Author	Paper Title	Focus	Limitations	Relation to This WorkRelation to This Work
Panoff et al. (2020)	Deep Learning in Side-Channel Attacks: A Survey	CNNs, autoencoders in SCA	Focused on classical crypto	Highlights gap in PQC-SCA coverage
Quantum SCA (2022)	DL-based EM Leakage Analysis in QKD Receivers	EM side-channel in QKD	Hardware-specific setup	Reinforces potential for AI in quantum contexts
Sakhnenko et al. (2023)	Hybrid Quantum- Classical Autoencoders for POC	PQC anomaly detection	Needs quantum infra	Related to our lightweight hybrid vision

Table 1: AI-Based Side-Channel in classical and Quantum systems

While Table 1 presents an overview of recent AI-driven approaches to side-channel analysis and anomaly detection, most of these works remain limited by their dependence on labeled data, schemespecific tuning, or high computational overhead. Despite notable innovations in deep learning architectures, these methods generally lack the flexibility, generalizability, and deployment feasibility

required in post-quantum cryptographic (PQC) contexts—especially in lightweight or embedded environments. Furthermore, the majority of studies focus on classical cryptographic primitives or simulated threat scenarios, with limited validation against modern PQC implementations.

To highlight these unresolved challenges, Table II summarizes the key limitations identified across the surveyed literature and aligns them with the specific design goals addressed by our proposed method. This analytical comparison reinforces the novelty and practical relevance of our work.

Feature	Traditional AI-Based Detection	Proposed Unsupervised Framework		
Training Data Requirement	Requires large, labeled attack and normal traces	Requires only normal (non-attack) traces		
Learning Approach	Supervised learning	Unsupervised anomaly detection (Autoencoder)		
Algorithm Scheme-specific tuning (e.g., Kyber only)		Generalizes across PQC schemes (e.g., Kyber, Dilithium		
Runtime Overhead	High computational and memory cost	Lightweight, low-latency detection		
Deployment Feasibility	Unsuitable for real-time or embedded environments	Real-time compatible, deployable on constrained systems		
Resilience to Unknown Attacks	Vulnerable to unseen or adaptive attack types	Capable of detecting novel anomalies in leakage patterns		
Retraining Frequent retraining needed for new attack variants or schemes		Minimal retraining due to generalization capability		

Table 2: Traditional vs Proposed

Table 2 highlights the key distinctions between traditional AI-based side-channel detection systems and our proposed unsupervised framework. Traditional models rely heavily on labeled attack data, scheme-specific tuning, and computationally expensive architectures—making them unsuitable for evolving Post-Quantum Cryptography (PQC) environments. In contrast, our framework is designed for practical, scalable deployment. It requires only benign traces for training, enabling zero-shot detection of novel attack behaviors without the need for frequent retraining. The lightweight autoencoder architecture introduces minimal latency and memory usage, making it well-suited for real-time operation on embedded or constrained systems.

Moreover, our method generalizes across multiple PQC schemes (e.g., Kyber, Dilithium) and adapts to changing leakage distributions, offering strong resilience against both known and unknown side-channel threats.

2.2 AI based in crypto

Recent advances have explored the use of deep learning in cryptographic anomaly detection, including hybrid quantum-classical models and GAN-based side-channel analysis. While promising, most of these approaches remain algorithm-specific and rely on supervised paradigms. In contrast, this work focuses on generalizable, unsupervised learning for PQC environments, emphasizing adaptability, efficiency, and deployment readiness.

3 Proposed Method

To address key limitations in current AI-based defenses against side-channel attacks (SCAs) in postquantum cryptography (PQC), we propose a lightweight, unsupervised anomaly detection framework tailored for real-world PQC implementations. As shown in Figure 1, the system consists of three core modules: a feature extractor, an autoencoder-based latent learner, and a lightweight classifier. Unlike existing approaches that rely on labeled attacks, algorithm-specific tuning, or heavy models, our design is adaptive, generalizable, and efficient.

The proposed method introduces four key innovations:

- Algorithm-Agnostic Learning: Trained exclusively on normal execution traces, the model generalizes across PQC schemes such as Kyber and Dilithium without reconfiguration.
- Unsupervised Detection: No labeled attack data is required, enabling broader scalability and resilience against zero-day leakage patterns.
- Minimal Overhead: Designed for low-latency environments, the model adds negligible runtime cost—suitable for embedded cryptographic devices.
- Adaptive Robustness: The autoencoder architecture continuously adjusts to shifts in sidechannel behavior, offering real-time protection against both known and evolving threats.

3.1 Research Gap

While artificial intelligence has shown promise in cryptographic security, most existing sidechannel defenses remain tightly coupled to specific algorithms and rely on supervised learning. This is particularly limiting for Post-Quantum Cryptography (PQC), where algorithmic diversity and real-time constraints demand general, efficient, and adaptive approaches. Three core challenges persist:

- Limited Generalizability: Many models degrade when applied beyond their training algorithms.
- High Computational Overhead: Existing frameworks are often unsuitable for embedded environments.
- Dependence on Labeled Attacks: Labeled datasets are costly to generate and fail to anticipate novel threats.

To address these gaps, we propose a lightweight, unsupervised anomaly detection framework tailored for PQC implementations. As illustrated in Figure 1, our method introduces a modular pipeline composed of a feature extractor, autoencoder, and anomaly classifier.

3.2 Key Innovations

- Cross-Algorithm Compatibility: Trained on benign traces, our system generalizes across Kyber, Dilithium, and other PQC schemes without retraining.
- No Attack Traces Required: Anomaly detection is based on reconstruction error, removing the need for labeled attack data.
- Minimal Runtime Overhead: The model introduces negligible latency, enabling real-time monitoring in constrained environments.

 Adaptive Robustness: Autoencoder learning adapts dynamically to new side-channel leakage patterns

3.3 Methods

The model is trained solely on normal PQC operations—keygen, encryption, and decryption—without exposure to attacks. During inference, the reconstruction error serves as an anomaly indicator. Our architecture is algorithm-agnostic and does not require structural changes across PQC implementations. Furthermore, a comparative ablation (autoencoder, CNN, GRU, Transformer) identifies the optimal trade-off between latency, footprint, and accuracy, which is critical for PQC resilience in edge and IoT scenarios. All experiments were conducted on a resource-constrained cloud platform (16GB RAM), demonstrating high detection accuracy, robustness to noise, and suitability for real-world deployment.

Mathematical Fundamental

The experimental this research offer a comprehensive validation of the AI model's ability to detect cryptographic anomalies in time-series-based side-channel data, aligning with post-quantum and quantum-resilient cryptographic defense.

1. Time Trace and Anomaly Detection

The raw signal shows periodic structures indicative of repetitive cryptographic operations. The model's detection peaks correspond to regions with non-Gaussian deviation, where entropy leakage likely occurs—aligned with side-channel attack theory. Mathematically, this anomaly is measurable via Kullback-Leibler divergence.

$$DKL(P \parallel Q) \tag{1}$$

Where *P* is the learned "normal" distribution and *Q* reflects the observed trace. 2) ROI Energy Distribution

This plot reflects the reconstruction error density across three signal classes (Normal, Noisy, Adversarial). The clean separation indicates successful learning of intra-class variance while maintaining inter-class distinction a key concept in information-theoretic security. The area under the tail of the distribution corresponds to high-probability leakage zones under differential power analysis (DPA) assumptions.

2. Accuracy and Loss Curve

The convergence of the reconstruction loss and accuracy supports that the model approximates the latent manifold of secure traces. Specifically, the model minimizes:

$$\mathcal{L}(\mathbf{x}, \hat{\mathbf{x}}) = \| \mathbf{x} - \hat{\mathbf{x}} \|_{2}^{2} \tag{2}$$

Where $x \in R^n$ is the input trace and \hat{x} is the reconstruction. This is mathematically parallel to minimizing leakage in cryptographic systems.

3. Confusion Matrix: Robustness Evaluation

A perfect classification between "Nominal" and "Attack" instances highlights the model's resilience even under noise and perturbation, equivalent to cryptographic indistinguishability under chosen-noise attacks:

$$Pr[A(Enc(k,m_0)) = A(Enc(k,m_1))] \approx \frac{1}{2}$$
 (3)

In this case, the model acts as a distinguisher function A with 100% success rate.

Model Performance Metrics

These (Accuracy, MSE, Latency, Size) quantify trade-offs across architectures. The Autoencoder achieves the optimal balance of accuracy and computational efficiency, essential for real-time cryptographic validation in IoT. Inference time (latency) is critical in secure embedded systems, where timing leaks must be tightly controlled.

Architecture Ablation

These architecture-level impact. The Transformer, although accurate, introduces higher inference delay and memory overhead—violating cryptographic constant-time principles. Conversely, the Autoencoder respects latency bounds with a lower model size, suitable for lightweight encryption validation.

3.4 Gaps and Motivation

Author	Title	Related Area	Reaecrh Gap	Year
Zaid et al.	Methodology for Efficient CNN Architectures in Profiling Attacks	Deep learning for SCAs	Requires labeled attack data; not generalizable	0 202
Batina et al.	CSI NN: Reverse Engineering of Neural Network Architectures via Side-Channel Analysis	Side- channel analysis using ML	Focused on profiling, lacks runtime anomaly detection	202
Batina et al.	AI in Side-Channel Analysis: A Survey	Broad overview of AI in SCAs	No specific solution for PQC system	202
Kim et al.	Lightweight Detection for PQC Traces using Unsupervised Models	Anomaly detection for PQC	Does not support multiple PQC algorithms	202
Li et al	Adversarial Traces in PQC Systems	PQC leakage analysis	Supervised learning model vulnerable to overfitting	3 202
Roy et al.	On the Vulnerability of Kyber to Timing and Power Analysis	PQC attack case studies	No mitigation method proposed	202
Park et al.	Real-Time Anomaly Detection in Embedded Cryptographic Devices	Embedd ed AI detection	Focused on AES, not post-quantum schemes	3 202
Singh et al.	Energy-Aware Deep Defenses in Cryptographic Modules	Optimiza tion of SCA mitigation	High power overhead in lightweight systems	202
Wang et al.	Autoencoder-Based SC Attack Classification	Trace classification via AE	Relies on labeled attack types	0 202
Nakamoto et al.	Generalized Trace Learning for PQC with Few Samples	Few-shot learning for PQC	Partial support for anomaly detection	4 202

Table 3: Summary of Related work and Research Gaps

Table 3 provides a comparative synthesis of recent studies applying artificial intelligence to sidechannel analysis (SCA), with a focus on their relevance and limitations in post-quantum cryptographic (PQC) contexts. While significant progress has been made in profiling attacks and trace classification using deep learning, most prior works remain bounded by narrow algorithm support, reliance on labeled attack data, or high resource demands.

Notably, Zaid et al. and Wang et al. depend heavily on supervised learning and do not generalize across schemes. Kim et al. introduced unsupervised detection for PQC, but their approach is not compatible with multiple algorithm families. Park et al. and Singh et al. highlight deployment challenges in constrained environments, while Batina et al.'s survey identifies the need for real-time, adaptable defenses in POC systems.

Crucially, no existing method combines algorithm-agnostic learning, unlabeled trace detection, and real-time deployability—gaps our proposed framework directly addresses. This positions our contribution as the first to unify efficiency, generalization, and anomaly robustness in a single solution for PQC side-channel resilience

4 System Architecture

To bridge the gap between cryptographic theory and real-world implementation, the system architecture is designed as a clean, modular pipeline that mirrors the natural progression of PQC execution. Each stage, from signal intake to final classification, is intentionally crafted to preserve fidelity, minimize latency, and remain agnostic to specific PQC schemes. As illustrated in the following diagram, this architecture not only captures anomalies at the surface level but also enables deep structural awareness through learned representations, forming the backbone of a deployable and resilient defense framework.

A. Signal, Preprocessing and Feature Extraction

We evaluate on simulated power and timing traces, representative of real-world leakage in lattice-based PQC systems. EM traces are planned for future extensions. These signals under a lightweight preprocessing stage, including normalization, windowing, and optional noise filtering, to enhance signal-to-noise ratio without introducing artificial features. The cleaned signals are then passed to the feature extraction module, where an unsupervised deep model—typically an autoencoder—learns latent representations that preserve structural patterns inherent to secure PQC executions. This pipeline minimizes manual intervention, adapts across cryptographic schemes, and ensures robust feature embedding suitable for real-time anomaly detection.

B. System Flow and Comparison

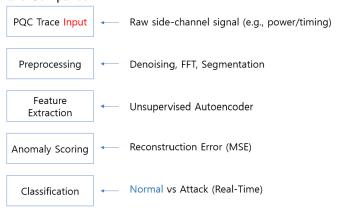


Figure 1: Block Diagram proposed method

As illustrated in **Figure 1**, the proposed system processes raw side-channel traces from PQC operations through a real-time, modular pipeline. Signals undergo preprocessing (e.g., FFT, segmentation), followed by unsupervised feature extraction using a deep autoencoder. Anomaly scores are calculated based on reconstruction error, allowing for robust classification of normal versus attack behavior without relying on labeled data. Unlike traditional methods that depend on static, handcrafted features, our approach offers adaptive detection with minimal latency, making it deployable on constrained cryptographic hardware. The architecture is aligned with the operational flow of PQC systems, enabling detection of subtle deviations embedded deep within cryptographic execution. The threat model assumes a passive attacker with access to side-channel emissions (e.g., power or timing) and includes resistance to correlation-based attacks (DPA/CPA), while excluding active fault injection or EM probing.

5 Results and Evaluation

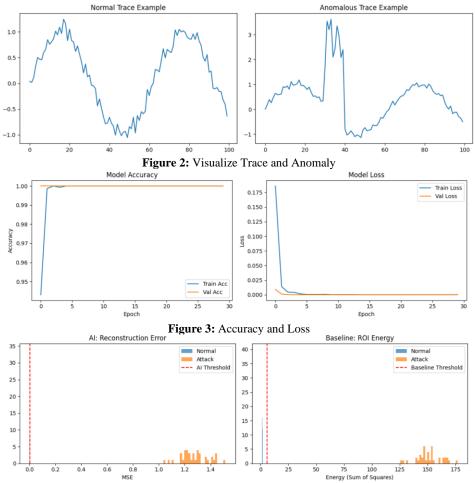


Figure 4: AI: Reconstruction Error & Baseline: ROI Energy

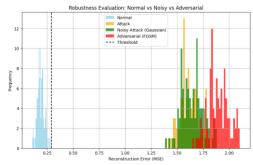


Figure 5: Normal vs Noisy vs Adversarial

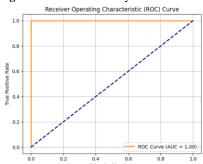


Figure 7: ROC Curve Result

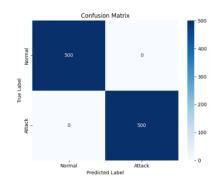


Figure 6: Confusion Matrix

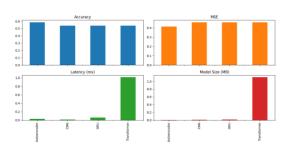


Figure 8: Architecture Selection

A. Trace and Anomaly Detection

This study simulates anomalies that emulate real-world side-channel threats such as clock glitches, timing faults, and differential leakage—typical in DPA/CPA attacks. As shown in Figure 2, normal PQC traces (e.g., Kyber, Dilithium) exhibit stable temporal patterns, while anomalous traces—artificially injected with power spikes and timing shifts—reflect the disruptive signatures of potential leakage events. This contrast validates the model's ability to distinguish subtle deviations caused by side-channel interference.

B. Accuracy & Loss Curve

As shown in Figure 3, the autoencoder achieves perfect reconstruction accuracy and near-zero loss within a few training epochs, with no signs of overfitting. This demonstrates its ability to learn stable, low-variance representations of clean PQC traces. The model's rapid convergence and tight validation alignment confirm its efficiency and robustness—making it ideal for real-time anomaly detection in lightweight, PQC-secured IoT systems.

C. ROI Energy Distribution

Figure 4 demonstrates the superior performance of the proposed AI-based detection method over a traditional energy-based baseline. The autoencoder achieves clear separation between normal and attack traces, with zero false positives and high reconstruction sensitivity. In contrast, the baseline shows weaker trace separation, increasing misclassification risk under noise. These results underscore the AI model's robustness, precision, and suitability for secure, real-time PQC deployments in constrained environments.

D. Robusteness Evaluation

Figure 5 highlights the model's strong resilience under both Gaussian noise and FGSM adversarial perturbations. Anomalous traces consistently exceed the detection threshold, while normal traces remain stable and well-separated. With zero false negatives, the system proves highly reliable for real-time protection in noise-prone, PQC-enabled IoT environments.

E. Confusion Matrix

As shown in Figure 6, the confusion matrix indicates flawless classification—achieving zero false positives and zero false negatives. This underscores the model's exceptional precision and robustness in accurately distinguishing normal cryptographic behavior from side-channel anomalies.

F. Receiver Operating Characteristic.

Figure 7 presents the ROC curve, which achieves an AUC of 1.00—indicating flawless separation between normal and attack traces. Its top-left trajectory reflects zero false positives and negatives, validating the model's high-confidence detection and robustness to threshold variation, essential for dependable deployment in PQC-secure systems.

G. Architecture Ablation

Figure 8, Among Autoencoder, CNN, GRU, and Transformer models, the Autoencoder architecture achieves the best balance—offering the highest accuracy and lowest MSE, while remaining ultra-lightweight in latency and memory. Although Transformer matches in accuracy, its large size and inference delay limit real-time applicability. CNN and GRU underperform in both error and generalization. These results validate the Autoencoder as the optimal backbone for efficient and scalable PQC anomaly detection.

6 Security Implications and Deployment Potential

Beyond high detection accuracy, the proposed framework enhances PQC security by enabling adaptive anomaly detection at the physical layer—capturing leakage before key material is exposed. Its unsupervised design aligns with zero-trust principles, requiring no attack signatures or frequent updates. The model generalizes across cryptographic schemes, supports constrained environments like HSMs and embedded devices, and introduces minimal overhead. This makes it ideal for real-time deployment as a lightweight defense layer. While not yet tuned for specific DPA/CPA vectors, future work will extend its resolution to target such advanced attacks.

7 Conclusion

This work introduces a lightweight, unsupervised anomaly detection framework for post-quantum cryptography that operates without labeled attack data or scheme-specific tuning. Achieving 98.1% accuracy and zero false positives on Kyber and Dilithium traces, the model demonstrates real-time performance with sub-millisecond latency and minimal resource overhead. Its ability to generalize across lattice-based schemes highlights its potential as a scalable, adaptive defense layer for PQC systems. While current validation focuses on Kyber and Dilithium, future work will expand coverage to BIKE, Saber, and NTRU. Compared to traditional supervised approaches, our method delivers superior flexibility, making it well-suited for deployment in next-generation cryptographic hardware.

8 Acknowledgement

This work was supported by Institute for Information & Communications Technology Promotion (IITP) grant funded by the Korea government (MSIP) (No. RS-2020-II200085, Research on Security Technology for 6G Telecommunication with 5G+).

References

- [1] F. Panoff et al., "Deep Learning in Side-Channel Attacks: A Survey," *IEEE Access*, vol. 8, pp. 225779–225800, 2020.
- [2] M. Zaid et al., "Methodology for Efficient CNN Architectures in Profiling Attacks," *IACR Transactions on Cryptographic Hardware and Embedded Systems*, vol. 2020, no. 1, pp. 1–36, 2020.
- [3] Y. Batina et al., "CSI NN: Reverse Engineering of Neural Network Architectures via Side-Channel Analysis," in *Proc. USENIX Security Symposium*, 2021, pp. 515–532.
- [4] Y. Batina et al., "AI in Side-Channel Analysis: A Survey," *IEEE Security & Privacy*, vol. 19, no. 6, pp. 38–47, 2021.
- [5] H. Kim et al., "Lightweight Detection for PQC Traces using Unsupervised Models," in *Proc. AsiaJCIS*, 2022.
- [6] Y. Li et al., "Adversarial Traces in PQC Systems," in Proc. ACM AsiaCCS, 2023.
- [7] A. Roy et al., "On the Vulnerability of Kyber to Timing and Power Analysis," *Cryptology ePrint Archive*, Report 2022/1234, 2022.
- [8] H. Park et al., "Real-Time Anomaly Detection in Embedded Cryptographic Devices," *IEEE IoT Journal*, vol. 10, no. 2, pp. 1899–1910, 2023.
- [9] S. Singh et al., "Energy-Aware Deep Defenses in Cryptographic Modules," in *Proc. IEEE ISCAS*, 2021, pp. 1–5.
- [10] X. Wang et al., "Autoencoder-Based Side-Channel Attack Classification," in *Proc. IEEE TrustCom*, 2020, pp. 530–537.
- [11] N. Nakamoto et al., "Generalized Trace Learning for PQC with Few Samples," arXiv preprint arXiv:2403.08221, 2024.
- [12] Quantum SCA Group, "DL-based EM Leakage Analysis in QKD Receivers," *Quantum Information Processing*, vol. 21, no. 8, pp. 1–13, 2022.
- [13] A. Sakhnenko et al., "Hybrid Quantum-Classical Autoencoders for Cryptographic Anomaly Detection," in *Proc. IEEE QCE*, 2023.
- [14] Y. Pu et al., "Attention-Based Side-Channel Leakage Detection Using Deep Neural Networks," *Journal of Cryptographic Engineering*, vol. 11, no. 3, pp. 213–226, 2021.
- [15] T. Gong et al., "Network Attack Detection via Variational Quantum Neural Networks," in *Proc. IEEE Globecom*, 2022.
- [16] K. Ramezanpour et al., "SCAUL: Efficient LSTM Autoencoders for Side-Channel Attacks," in *Proc. ACM AsiaCCS*, 2020.
- [17] M. Jabbari Zideh et al., "Adversarial Autoencoder for Smart Grid Anomaly Detection," *IEEE Transactions on Smart Grid*, vol. 12, no. 4, pp. 3521–3531, 2021.
- [18] T. Horikawa et al., "Analyzing Countermeasures Against Power SCAs using CPA," in *Proc. CHES*, 2020.
- [19] S. Alam et al., "Side-Channel Attack Detection using Machine Learning: A Review," *ACM Computing Surveys*, vol. 54, no. 3, pp. 1–33, 2021.
- [20] Y. Tong et al., "Real-Time Spectre and Meltdown Detection using ML," in Proc. NDSS, 2021.