# Hybrid Quantum-Classical Intrusion Detection with Quantum Feature Representations under NISQ Constraints

**Murilo Salem**[1], **Daniel Pontes**[1], **Luísa Böhm**[1]
**Henrique dos Reis** [1], **Karen Satie Ono**[1], **Anderson Priebe Ferrugem**[1]

[1]Centro de Desenvolvimento Tecnológico - Universidade Federal de Pelotas (UFPel)
96010-610 – Pelotas – RS – Brazil

{mcsalem, dhspbarretos, lcbohm, hdreis, ksono, ferrugem}@inf.ufpel.br

***Abstract.*** *Intrusion detection systems must balance predictive quality, robustness, and computational cost, yet the role of quantum representations under NISQ constraints remains unclear. This paper investigates whether quantum principal component analysis (QPCA) can provide useful features for IDS without relying on claims of end-to-end quantum superiority. We evaluate PCA- and QPCA-based pipelines combined with Logistic Regression, SVM, and Random Forest, and include a QPCA→VQC branch as a comparative quantum arm. Experiments on CICIDS2017 and NSL-KDD under* `nisq_preset` *and* `scaled_preset` *use simulator-based quantum execution, multi-seed evaluation, Wilcoxon–Holm tests, bootstrap confidence intervals, and cost analysis. Results show no universal advantage of QPCA, but selective ranking gains: ROC-AUC improves from 0.5397 to 0.8772 (CICIDS2017) and from 0.7453 to 0.8063 (NSL-KDD), both with corrected significance under matched data budgets. Overall, QPCA is most useful as a representation enhancer under NISQ-compatible, not hardware-validated, constraints.*

## 1. Introduction

Intrusion detection systems (IDSs) must identify malicious traffic under constraints of accuracy, robustness, latency, and cost [Sowmya and Mary Anita 2023, Abdallah et al. 2022]. Classical models such as Logistic Regression (LR), Support Vector Machines (SVM), and Random Forests (RF) remain strong baselines, but their performance depends heavily on feature representation [Abdallah et al. 2022, Sowmya and Mary Anita 2023]. This makes dimensionality reduction a central design component.

Quantum machine learning has been proposed as a way to enrich feature transformations [Biamonte et al. 2017, Havlíček et al. 2019], but its practical use is limited by NISQ constraints, including noise, shallow circuits, and high optimization cost [Preskill 2018, McClean et al. 2018]. Rather than claiming quantum advantage, we focus on a narrower question: can *QPCA* improve feature representations for IDS under realistic constraints.

We evaluate a hybrid framework combining QPCA with classical classifiers, alongside PCA baselines and a QPCA→VQC branch. Experiments on CICIDS2017 and NSL-KDD [Sharafaldin et al. 2018, Tavallaee et al. 2009] use two regimes (`nisq_preset`, `scaled_preset`) and a multi-seed protocol with statistical testing and cost analysis.

Results show that QPCA is not universally superior to PCA, but can improve representation quality in specific settings, especially in ranking metrics. For example, ROC-AUC increases from 0.5397 to 0.8772 in CICIDS2017 and from 0.7453 to 0.8063 in NSL-KDD, both with corrected significance. Across scenarios, hybrid pipelines provide the most robust performance-cost trade-off.

The contributions are a controlled hybrid quantum-classical IDS framework that isolates QPCA under matched PCA/QPCA budgets, a multi-seed statistical evaluation with cost analysis and explicit simulator-based assumptions, and evidence of selective representational gains without universal or hardware-level quantum advantage.

## 2. Related Work and Positioning

Classical machine learning remains central to IDS, with LR, SVM, and RF serving as structured-data baselines [Sowmya and Mary Anita 2023]. They cover interpretable linear modeling [Hosmer et al. 2013], discriminative classification, and nonlinear ensembles, while controlling for preprocessing, imbalance, and dataset artifacts [Arcos-Argudo et al. 2025]. Recent work therefore emphasizes reproducibility, rigorous evaluation, and performance–efficiency trade-offs [Shaikhanova et al. 2025].

Quantum ML for security typically follows either end-to-end quantum classification or hybrid quantum-classical pipelines, where quantum stages act as feature transformers [Sai et al. 2025, Peral-García et al. 2024, Wang and Liu 2024]. Under NISQ constraints, hybrid approaches are more practical because of noise, limited qubits, and training challenges [Preskill 2018]. We therefore compare PCA and QPCA with the same downstream LR, SVM, and RF models, include QPCA$\rightarrow$VQC as a quantum baseline, and use multi-seed tests, Holm correction, bootstrap intervals, and cost analysis [Wilcoxon 1945, Holm 1979].

Prior IDS studies on CICIDS2017 and NSL-KDD use different label mappings, cleaning rules, balancing strategies, and splits, so direct leaderboard-style comparison is fragile. We therefore position prior work as methodological context and make the main claims through matched internal contrasts: PCA and QPCA receive the same samples, classifiers, seeds, and metrics. The novelty is thus evaluative rather than algorithmic: identifying when a quantum representation stage is useful and when classical alternatives remain preferable. This positioning also clarifies the comparison with recent IDS papers that report high absolute scores on the same datasets: such numbers are not directly comparable unless preprocessing, sampling, binary conversion, and train/test partitions are aligned. For this reason, the paper prioritizes controlled internal contrasts and uses prior work mainly to motivate the design space and the need for reproducible evaluation.

This design choice addresses an important comparability issue raised by the evaluation of IDS benchmarks. Many papers on CICIDS2017 and NSL-KDD report strong absolute values, but they often differ in attack grouping, removal of duplicate flows, normalization, imbalance treatment, and whether validation data leaks into model selection. Instead of treating these numbers as a direct ranking, our protocol makes the comparison local and paired. Each seed induces the same split for PCA and QPCA, and each downstream model receives the same number of features. As a result, the reported deltas are interpreted as evidence about the representation stage, not as a claim that the full pipeline dominates all previously published IDS systems. This also makes negative results

informative: when QPCA fails to improve F1 or loses to Random Forest after PCA, the outcome is preserved rather than hidden.

## 3. Hybrid Quantum-Classical IDS Framework

### 3.1. Problem Formulation

We model IDS as binary classification over tabular network data, $\mathcal{D} = \{(\mathbf{x}_i, y_i)\}_{i=1}^{N}$, with $\mathbf{x}_i \in R^d$ and $y_i \in \{0, 1\}$. The goal is to learn $f : R^d \to \{0, 1\}$ that generalizes across seeds and splits under NISQ-compatible constraints.

We evaluate `CICIDS2017` and `NSL-KDD`: the former reflects modern traffic complexity, while the latter is a controlled IDS reference. The `nisq_preset` uses reduced samples and fixed quantum features; `scaled_preset` increases the data budget while preserving the pipeline, separating representation effects from scaling.

### 3.2. Pipelines Under Comparison

The framework compares four pipeline families so that the effects of feature representation, downstream classifier, and quantum decision stage can be examined separately. In Table 1, `LR` denotes Logistic Regression and `RF` denotes Random Forest.

This comparison space is deliberately structured. Classical baselines define the reference frontier, hybrid pipelines test the representational contribution of `QPCA` under matched classical learners, the `QPCA` $\rightarrow$ `VQC` branch probes a deeper quantum role at higher cost and variability, and the swap-selection pipelines provide auxiliary ablation evidence without changing the main narrative.

### 3.3. NISQ Constraints and Design Choices

NISQ constraints are treated as a core methodological component rather than a post-hoc limitation. All quantum executions are simulator-based; no claims are made about real quantum hardware execution. The term NISQ denotes a deliberately bounded design: small quantum-feature dimensionality, shallow circuits, limited sample budgets, and explicit cost accounting. Unless otherwise noted, the reported runs do not inject calibrated hardware noise models, so hardware noise and device calibration remain outside the empirical scope.

The QPCA stage maps the reduced classical vector to a fixed quantum representation that is then passed to the downstream classifier. In the reported benchmark, this representation is constrained to four features/qubits, with angle encoding, shallow nearest-neighbor entanglement, and bounded depth. The VQC branch uses the same QPCA representation followed by a hardware-efficient variational classifier with 24 or 32 trainable parameters, depending on the regime.

The `nisq_preset` is the primary strict setting, while `scaled_preset` increases the data budget without changing the pipeline. Thus, scaled-regime results are matched PCA-vs.-QPCA evidence within the same enlarged budget, not proof that the gain comes from quantum processing alone. The objective is practical utility under controlled NISQ-compatible assumptions, not hardware-level quantum advantage.

**Tabela 1. Pipeline families and roles.**

| Family | Pipelines | Purpose |
|---|---|---|
| Classical baseline | PCA→LR/SVM/RF | Reference frontier under matched dimensionality reduction. |
| Hybrid quantum-classical | QPCA→LR/SVM/RF | Test whether QPCA improves representations for the same classifiers. |
| Comparative quantum arm | QPCA→VQC | Estimate cost, variability, and possible gain of a variational decision stage. |
| Auxiliary ablation | swap-selection→LR/SVM | Diagnostic support signal, not central to the main claim. |

## 4. Experimental Protocol

### 4.1. Datasets and Operating Regimes

We evaluate the framework on two IDS benchmarks under two operating regimes with different computational budgets. All samples are mapped to binary labels (*benign* vs. *attack*), ensuring a consistent decision setting across pipelines. Table 2 summarizes their roles.

**Tabela 2. Summary of datasets and operating regimes.**

| Element | Option | Purpose | Constraint / Role |
|---|---|---|---|
| Dataset | `CICIDS2017` | Evaluate robustness on a modern IDS benchmark | Richer traffic structure and contemporary intrusion patterns |
| Dataset | `NSL-KDD` | Assess cross-dataset consistency | Established IDS reference for comparative evaluation |
| Regime | `nisq_preset` | Primary controlled benchmark | Reduced sample budget and fixed quantum features |
| Regime | `scaled_preset` | Test stability at larger scale | Increased data budget with unchanged pipeline design |

### 4.2. Evaluation Protocol

All models use explicit train/validation/test splits; validation supports model and threshold selection, and test is reserved for final evaluation [Kohavi 1995, Cawley and Talbot 2010]. The base protocol uses five seeds, $[7, 21, 42, 84, 168]$. The v2 protocol adds $[101, 202, 303, 404, 505]$ only to central comparison pairs, increasing them to $n = 10$. All quantum components use the same simulator configuration. Scripts, processed results, seeds, and figure notebooks are intended for public release: `REPOSITORY-URL-TO-BE-INSERTED`.

We report accuracy, precision, recall, F1, ROC-AUC, PR-AUC, and balanced accuracy. F1 is primary, while ROC-AUC and PR-AUC capture ranking behavior under threshold and imbalance effects.

### 4.3. Statistical Testing and Cost Analysis

Priority comparisons use paired Wilcoxon signed-rank tests with Holm correction [Wilcoxon 1945, Holm 1979]. We also report $95\%$ bootstrap intervals for key metrics and paired differences. Cost is measured through mean training time, inference time, and model complexity, since variational quantum models incur overhead from repeated circuit evaluations and classical optimization.

## 5. Results

This section reports the final multi-seed benchmark and v2 hardening of the central comparison pairs, focusing on benchmark behavior, the representational effect of `QPCA`, the variational arm, and cost–performance trade-offs.

### 5.1. Main Multi-Seed Benchmark Results

Table 3 summarizes the benchmark. The four central v2 contrasts use $n = 10$ seeds; the remaining pipelines retain $n = 5$.

Table 3 shows no universal winner. On `CICIDS2017` [Sharafaldin et al. 2018], tree-based models achieve the highest F1, while `QPCA`→RF remains competitive and has the best NISQ-regime ROC-AUC. On `NSL-KDD`, `QPCA`→VQC has the highest mean F1 but much higher variance. Thus, `QPCA` is best interpreted as a selective representation intervention, not a universal performance amplifier [Preskill 2018, Cerezo et al. 2021].

### 5.2. Does QPCA Improve Representation?

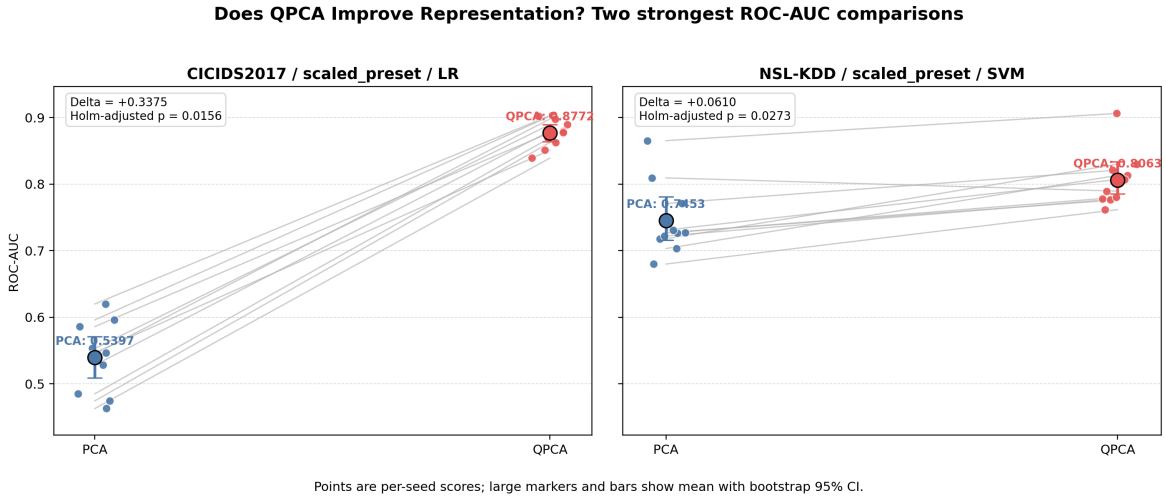
The main question is not whether `QPCA` improves all downstream metrics, but whether it can improve the representation given to the classifier under NISQ-compatible constraints [Preskill 2018, Cerezo et al. 2021]. The results suggest that it can, but mainly in ranking and separability metrics rather than in universal gains in thresholded F1 [Fawcett 2006].

The clearest evidence appears on `CICIDS2017` [Sharafaldin et al. 2018] under `scaled_preset` with logistic regression. Replacing `PCA` with `QPCA` increases mean ROC-AUC from $0.5397$ to $0.8772$, with paired Wilcoxon–Holm significance preserved after multiplicity correction ($p_{\text{Holm}} = 0.0156$). The paired mean difference is $0.3375$, with bootstrap $\text{CI}_{95}$ $[0.3146, 0.3601]$. Because both pipelines use the same enlarged data budget, the result supports a matched representational difference within `scaled_preset`, not an isolated proof that QPCA alone explains the gain. By contrast, the F1 increase from $0.3261$ to $0.3927$ does not remain significant after Holm correction. This suggests that `QPCA` improves ranking quality and class separability, while the final thresholded decision remains limited by the classifier and the fixed operating point [Fawcett 2006, Schuld and Killoran 2019].

A similar result appears on `NSL-KDD` under `scaled_preset` with SVM. Here, `QPCA`→`SVM` improves mean ROC-AUC from $0.7453$ to $0.8063$, with $p_{\text{Holm}} = 0.0273$, paired mean difference $0.0610$, and bootstrap $\text{CI}_{95}$ $[0.0376, 0.0822]$. Again, the gain is clearer in ranking than in thresholded F1, which changes only slightly from $0.6849$ to $0.6958$ and is not significant. This supports the view that `QPCA` can reveal a more informative ordering of attack and benign samples without guaranteeing an improvement at the default classification threshold [Fawcett 2006].

**Tabela 3. Main multi-seed benchmark results (mean ± standard deviation). Boldface marks the best result within each dataset/regime for each metric.**

| Data | Regime | Metric | PCA | | | QPCA | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | LR | SVM | RF | LR | SVM | RF | VQC |
| CIC17 | nisq | F1 | 0.262±0.122 | 0.568±0.098 | **0.865±0.097** | 0.241±0.235 | 0.572±0.100 | 0.809±0.128 | 0.349±0.110 |
| | | ROC | 0.598±0.073 | 0.887±0.021 | 0.952±0.066 | 0.882±0.029 | 0.884±0.015 | **0.955±0.059** | 0.879±0.021 |
| CIC17 | scaled | F1 | 0.326±0.108 | 0.585±0.042 | **0.898±0.051** | 0.393±0.075 | 0.553±0.070 | 0.890±0.059 | 0.307±0.054 |
| | | ROC | 0.540±0.053 | 0.885±0.012 | **0.984±0.015** | 0.877±0.022 | 0.872±0.036 | 0.983±0.014 | 0.829±0.058 |
| NSL | nisq | F1 | 0.666±0.018 | 0.673±0.033 | 0.697±0.018 | 0.662±0.171 | 0.629±0.165 | 0.694±0.078 | **0.712±0.206** |
| | | ROC | 0.850±0.016 | 0.739±0.025 | 0.859±0.011 | **0.879±0.020** | 0.791±0.044 | 0.847±0.027 | 0.810±0.062 |
| NSL | scaled | F1 | 0.666±0.036 | 0.685±0.023 | 0.704±0.031 | 0.640±0.133 | 0.696±0.129 | 0.689±0.053 | **0.723±0.166** |
| | | ROC | 0.838±0.008 | 0.745±0.055 | 0.849±0.016 | **0.872±0.015** | 0.806±0.042 | 0.834±0.019 | 0.854±0.020 |




**(a) `NSL-KDD / scaled_preset`: PCA → SVM vs. QPCA → SVM.**

**Figura 1. Central representation-sensitive results. Replacing `PCA` with `QPCA` improves ROC-AUC from 0.5397 to 0.8772 in `CICIDS2017/scaled/LR` and from 0.7453 to 0.8063 in `NSL-KDD/scaled/SVM`, highlighting gains in ranking rather than universal F1 improvement.**

A similar but weaker trend appears in the remaining comparisons. On `NSL-KDD/nisq_preset` [Tavallaee et al. 2009], `QPCA→SVM` raises mean ROC-AUC from 0.7387 to 0.7913; the uncorrected Wilcoxon test is significant ($p = 0.0098$), but the effect narrowly misses Holm significance ($p_{\text{Holm}} = 0.0586$). On `CICIDS2017/nisq_preset` [Sharafaldin et al. 2018], `QPCA→Logistic Regression` yields a large ROC-AUC increase from 0.5982 to 0.8824, but the smaller paired sample in the original protocol prevents a strong corrected claim. These cases are therefore best viewed as supportive rather than primary evidence.

Overall, the results indicate that `QPCA` is most useful when the downstream classifier is relatively simple and the metric is sensitive to ranking quality [Fawcett 2006]. When the downstream learner already has high nonlinear capacity—especially random forests on `CICIDS2017`—the marginal benefit of `QPCA` becomes much smaller. The main claim is therefore not that `QPCA` universally improves IDS, but that it can improve separability in selected regimes, especially for low-capacity classical heads.

### 5.3. Hybrid Pipelines versus the Variational Quantum Arm

The QPCA→VQC branch serves as a fully quantum-informed comparative arm, allowing us to test whether quantum gains are better captured by end-to-end variational classification or by a hybrid decomposition. This is especially relevant under NISQ constraints, where variational quantum algorithms are promising but remain limited by trainability, noise, and optimization cost [Preskill 2018, Cerezo et al. 2021, Schuld et al. 2020, McClean et al. 2018, Wang et al. 2021]. The results favor the hybrid design as the more robust option.

On CICIDS2017, the variational arm is clearly outperformed by hybrid pipelines at the operating point. In nisq_preset, QPCA→VQC reaches only $0.3490 \pm 0.1101$ F1, below QPCA→Random Forest ($0.8092 \pm 0.1281$) and the SVM-based alternatives. The same holds in scaled_preset, where QPCA→VQC obtains $0.3073 \pm 0.0539$ F1 versus $0.8898 \pm 0.0594$ for QPCA→Random Forest. Even in ROC-AUC, the variational arm remains behind the strongest hybrid and classical baselines.

On NSL-KDD, the picture is more nuanced. The variational arm has the highest raw mean F1 in both regimes ($0.7115 \pm 0.2058$ in nisq_preset and $0.7230 \pm 0.1661$ in scaled_preset), so it is not merely a negative control. However, these gains come with much higher variance, reducing stability. In the scaled regime, for example, QPCA→VQC slightly exceeds QPCA→SVM in raw mean F1, but the hybrid SVM pipeline is cheaper and more conservative as a main-body choice. For this reason, the VQC arm is best seen as a comparative benchmark rather than the recommended deployment path.

Overall, the strongest practical synthesis is hybrid: use quantum processing where it may enrich representation, but retain classical decision layers when stability, repeatability, and efficiency matter most. This aligns with the NISQ literature, which treats hybrid quantum-classical learning as a pragmatic near-term strategy and evaluates quantum feature maps and variational models jointly with trainability and cost [Havlíček et al. 2019, Mitarai et al. 2018, Cerezo et al. 2021, Schuld et al. 2020]. The variational arm remains useful because it provides methodological completeness and shows that the strongest defensible gains in this study do not come from end-to-end quantum superiority.

### 5.4. Cost–Performance Trade-off

Figure 2 relates predictive performance to computational cost: the most defensible region is occupied by hybrid pipelines, not the variational arm.

The cost gap is substantial. On CICIDS2017/nisq_preset, QPCA→VQC requires $128.73$ seconds of training on average, while QPCA→SVM and QPCA→Random Forest require $0.0022$ and $0.1816$ seconds, respectively. Inference shows the same trend: $1.0439$ seconds for the VQC arm versus $0.0004$ seconds for QPCA→SVM and $0.0321$ seconds for QPCA→Random Forest. The scaled regime increases the gap further, with VQC training rising to $256.99$ seconds on CICIDS2017 and $269.49$ seconds on NSL-KDD. This matches the broader literature on variational quantum algorithms, where repeated circuit evaluations, shot noise, and classical optimization dominate the runtime [Cerezo et al. 2021, Menickelly et al. 2023, Preskill 2018].

This asymmetry is not explained by model size: despite having only $24$ or

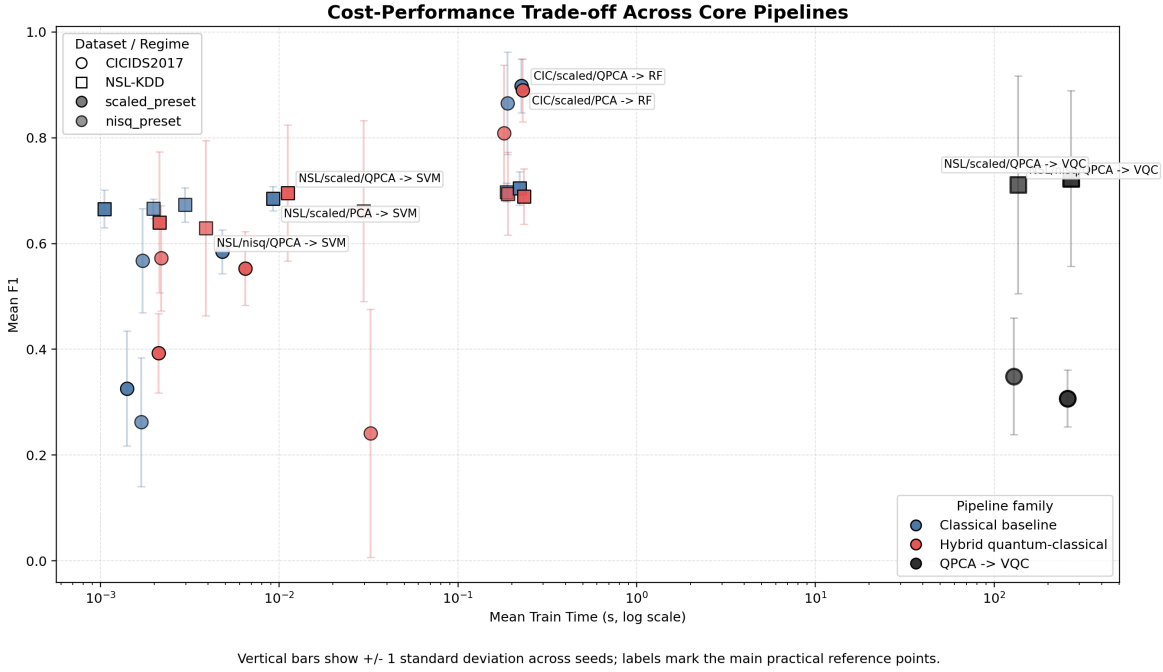



**Figura 2. Cost–performance trade-off across the main-body pipelines, using mean training time as the cost axis and aggregate predictive performance as the utility axis. The figure highlights the strong separation between the hybrid pipelines and the variational quantum arm.**

32 trainable parameters, VQC is orders of magnitude slower because runtime is driven by repeated circuit evaluations and optimization overhead [Cerezo et al. 2021, Menickelly et al. 2023, Benedetti et al. 2019]. The practical conclusion is therefore hybrid: use QPCA where it enriches representation, but retain classical decision layers for stability and cost [Preskill 2018, Cerezo et al. 2021, Benedetti et al. 2019, Havlíček et al. 2019].

## 6. Discussion

### 6.1. What the Results Support

The results support a limited but meaningful claim: *QPCA* can improve representation quality in IDS, mainly in ranking and separability metrics. The clearest evidence appears in the v2 central pairs. In *CICIDS2017 / scaled_preset*, *QPCA* $\rightarrow$ *Logistic Regression* raises ROC-AUC from $0.5397$ to $0.8772$ ($\Delta = 0.3375$, $p_{\text{Holm}} = 0.0156$). In *NSL-KDD / scaled_preset*, *QPCA* $\rightarrow$ *SVM* improves ROC-AUC from $0.7453$ to $0.8063$ ($p_{\text{Holm}} = 0.0273$). In both cases, the gains are stronger in ROC-AUC than in F1, suggesting an effect on representation geometry rather than on thresholded decisions [Fawcett 2006, Hernández-Orallo et al. 2012].

Hybrid models provide the best balance between performance, seed stability, and cost. The contribution is therefore not end-to-end quantum superiority, but evidence that a quantum feature stage can improve selected hybrid IDS pipelines under NISQ-compatible assumptions [Preskill 2018, Havlíček et al. 2019, Benedetti et al. 2019, Cerezo et al. 2021, LaRose et al. 2019].

A further implication is that QPCA should be treated as a conditional design option. It is most attractive when the downstream classifier is simple, interpretability or

ranking is important, and the extra representation cost remains small compared with the deployment budget. Conversely, when a nonlinear ensemble already separates the classes well, the additional quantum stage may add complexity without improving the operational decision. This interpretation is intentionally conservative and responds to the concern that selective gains could be overread as a general quantum advantage.

### 6.2. What the Results Do Not Support

The results do not support a general advantage of *QPCA* over *PCA*. The effect depends on dataset, regime, and downstream model. For example, in *CICIDS2017 / scaled_preset*, *QPCA* $\rightarrow$ *Random Forest* is competitive but not better than *PCA* $\rightarrow$ *Random Forest*, and *QPCA* $\rightarrow$ *SVM* is slightly worse. In *NSL-KDD / nisq_preset*, increasing to $n = 10$ seeds does not improve F1, and ROC-AUC gains do not survive Holm correction. Therefore, claims of consistent superiority would be overstated.

The results also do not support universal *quantum advantage*. *QPCA* $\rightarrow$ *VQC* can achieve high raw F1, but is more variable and expensive; the QPCA effect is contextual and clearer when representation quality matters more than downstream model capacity [Huang et al. 2021a, Huang et al. 2021b, Schreiber et al. 2023, Preskill 2018, Cerezo et al. 2021, McClean et al. 2018].

### 6.3. Limitations

This study has five main limitations. First, quantum executions are simulator-based and do not include real hardware runs or calibrated noise models. Second, controlled regimes improve reproducibility but reduce scalability. Third, simulation cost is high, especially for *QPCA* $\rightarrow$ *VQC*. Fourth, VQC training remains sensitive to seeds, thresholds, and optimization. Fifth, results depend on encoding choices, quantum-feature count, and classifier tuning. Quantum autoencoders and stronger classical reducers beyond PCA remain important future baselines. Another limitation is that repository release is essential for independent validation: although the protocol specifies seeds, metrics, and generated artifacts, external reproduction still depends on publishing preprocessing scripts, environment files, and the exact simulator configuration used in the runs.

## 7. Conclusion

This work evaluated quantum representations for IDS under a multi-seed statistical protocol. QPCA improves representation quality in selected settings, especially ranking and separability in *CICIDS2017/scaled/LR* and *NSL-KDD/scaled/SVM*, but it does not yield broad quantum superiority. The most robust behavior comes from hybrid quantum-classical pipelines rather than the fully variational arm. Future work should validate the findings on real hardware or calibrated fake-backend/noise simulations, explore alternative reducers and encodings, and study when ROC-AUC gains translate into stable thresholded decisions.

## Acknowledgments and Transparency

Generative AI tools supported language revision only. The authors reviewed and take responsibility for the technical content, experiments, and claims. Code and generated artifacts will be released through the repository indicated in the protocol.

## Referências

Abdallah, E. E., Eleisah, W., and Otoom, A. F. (2022). Intrusion detection systems using supervised machine learning techniques: A survey. *Procedia Computer Science*, 201:205–212.

Arcos-Argudo, D., García, P., et al. (2025). Quantum approaches for anomaly detection: A survey. *ACM Computing Surveys*. to appear.

Benedetti, M., Lloyd, E., Sack, S., and Fiorentini, M. (2019). Parameterized quantum circuits as machine learning models. *Quantum Science and Technology*, 4(4):043001.

Biamonte, J., Wittek, P., Pancotti, N., Rebentrost, P., Wiebe, N., and Lloyd, S. (2017). Quantum machine learning. *Nature*, 549(7671):195–202.

Cawley, G. C. and Talbot, N. L. C. (2010). On over-fitting in model selection and subsequent selection bias in performance evaluation. *Journal of Machine Learning Research*, 11(70):2079–2107.

Cerezo, M., Arrasmith, A., Babbush, R., Benjamin, S. C., Endo, S., Fujii, K., McClean, J. R., Mitarai, K., Yuan, X., Cincio, L., and Coles, P. J. (2021). Variational quantum algorithms. *Nature Reviews Physics*, 3(9):625–644.

Fawcett, T. (2006). An introduction to roc analysis. *Pattern Recognition Letters*, 27(8):861–874.

Havlíček, V., Córcoles, A. D., Temme, K., Harrow, A. W., Kandala, A., Chow, J. M., and Gambetta, J. M. (2019). Supervised learning with quantum-enhanced feature spaces. *Nature*, 567(7747):209–212.

Hernández-Orallo, J., Ferri, C., Lachiche, N., and Flach, P. A. (2012). A unified view of performance metrics: Translating threshold choice into expected classification loss. *Journal of Machine Learning Research*, 13:2813–2869.

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, 6(2):65–70.

Hosmer, D. W., Lemeshow, S., and Sturdivant, R. X. (2013). *Applied Logistic Regression*. Wiley, 3 edition.

Huang, H.-Y., Broughton, M., Mohseni, M., Babbush, R., Boixo, S., Neven, H., and McClean, J. R. (2021a). Power of data in quantum machine learning. *Nature Communications*, 12:2631.

Huang, H.-Y., Kueng, R., and Preskill, J. (2021b). Information-theoretic bounds on quantum advantage in machine learning. *Physical Review Letters*, 126(19):190505.

Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. In *Proceedings of the 14th International Joint Conference on Artificial Intelligence (IJCAI)*, volume 2, pages 1137–1143.

LaRose, R., Tikku, A., O'Neel-Judy, E., Cincio, L., and Coles, P. J. (2019). Variational quantum state diagonalization. *npj Quantum Information*, 5:57.

McClean, J. R., Boixo, S., Smelyanskiy, V. N., Babbush, R., and Neven, H. (2018). Barren plateaus in quantum neural network training landscapes. *Nature Communications*, 9:4812.

Menickelly, M., Ha, Y., and Otten, M. (2023). Latency considerations for stochastic optimizers in variational quantum algorithms. *Quantum*, 7:949.

Mitarai, K., Negoro, M., Kitagawa, M., and Fujii, K. (2018). Quantum circuit learning. *Physical Review A*, 98(3):032309.

Peral-García, M. et al. (2024). On the use of precision-recall curves in imbalanced learning. *Information Sciences*, 640:119000.

Preskill, J. (2018). Quantum computing in the nisq era and beyond. *Quantum*, 2:79.

Sai, K. et al. (2025). Evaluation metrics for imbalanced classification problems. *Pattern Recognition Letters*. to appear.

Schreiber, F. J., Eisert, J., and Meyer, J. J. (2023). Classical surrogates for quantum learning models. *Physical Review Letters*, 131(10):100803.

Schuld, M., Bocharov, A., Svore, K. M., and Wiebe, N. (2020). Circuit-centric quantum classifiers. *Physical Review A*, 101(3):032308.

Schuld, M. and Killoran, N. (2019). Quantum machine learning in feature hilbert spaces. *Physical Review Letters*, 122(4):040504.

Shaikhanova, A. et al. (2025). Hybrid quantum-classical models for anomaly detection. *Quantum Information Processing*. to appear.

Sharafaldin, I., Lashkari, A. H., and Ghorbani, A. A. (2018). Toward generating a new intrusion detection dataset and intrusion traffic characterization. *Proceedings of the 4th International Conference on Information Systems Security and Privacy (ICISSP)*, pages 108–116.

Sowmya, T. and Mary Anita, E. A. (2023). A comprehensive review of ai based intrusion detection system. *Measurement: Sensors*, 28:100827.

Tavallaee, M., Bagheri, E., Lu, W., and Ghorbani, A. A. (2009). A detailed analysis of the kdd cup 99 data set. In *Proceedings of the 2009 IEEE Symposium on Computational Intelligence for Security and Defense Applications*, pages 1–6.

Wang, H. and Liu, X. (2024). Auc optimization in machine learning: A review. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. early access.

Wang, S., Fontana, E., Cerezo, M., Sharma, K., Sone, A., Cincio, L., and Coles, P. J. (2021). Noise-induced barren plateaus in variational quantum algorithms. *Nature Communications*, 12(1):6961.

Wilcoxon, F. (1945). Individual comparisons by ranking methods. *Biometrics Bulletin*, 1(6):80–83.