Quantum-Inspired Machine Learning for Intrusion Detection System

Akshata Patil, Abhinav Pal, Keyur Patle, Prashant Rode, Dr. Nilesh Thorat *(Guide)*
Department of Computer Science and Engineering MIT ADT University Pune, India

***Abstract*—Modern cyber threats are growing in both volume and complexity, exposing the limitations of rule-based network monitoring tools. This work proposes a dual-model intrusion detection framework that pairs a classical Support Vector Ma-chine (SVM) with a quantum-inspired counterpart built on Qiskit's kernel estimation routines. Both classifiers are trained and tested on the NSL-KDD benchmark to distinguish benign traffic from malicious activity. Evaluation outcomes confirm that the classical model delivers strong, consistent detection performance, while the quantum-inspired variant demonstrates feasibility of quantum kernel computation within a simulation environment. Rather than claiming performance superiority, this study positions quantum-inspired methods as a promising avenue for next-generation security systems, offering practitioners and researchers a reproducible starting point for future hardware-level quantum IDS development.**

***Index Terms*—Intrusion Detection System, Machine Learning, Quantum-Inspired Computing, Support Vector Machine, NSL-KDD, Qiskit, Cybersecurity, Network Security**

I. INTRODUCTION

Today's networked environments are under relentless pres-sure from adversaries who exploit every gap in perimeter defenses. The steady rise in incidents ranging from credential theft and ransomware to coordinated DDoS campaigns has made automated traffic monitoring a non-negotiable component of modern security operations [1]. Conventional protection tools — firewalls, access-control lists, and signature scanners — operate on static knowledge bases that cannot keep pace with novel attack strategies [2]. As attack vectors mutate faster than human analysts can document them, the need for self-updating, data-driven detection has become acute.

Supervised machine learning fills this gap by constructing decision models from labeled traffic records, enabling classifiers to generalize beyond the examples they were trained on [3]. Among the available algorithms, SVM stands out for its theoretical guarantees on generalization and its practical track record on high-dimensional network feature vectors [4]. Its kernel-based formulation converts a linear decision boundary in the original space into a rich non-linear separator in a transformed space, a property that maps naturally onto the heterogeneous feature distributions found in network flow data. Quantum computing introduces an orthogonal lens through which to view these classification problems. By encoding data into quantum states and measuring their inner products via circuit execution, quantum kernels access feature spaces whose dimension grows exponentially with qubit count [5].

Current NISQ-era hardware is too noisy for production use, but simulation frameworks such as Qiskit make it possible to prototype quantum algorithms on conventional machines and validate their logic before real-device deployment [6].

This study constructs a side-by-side comparison between a classical RBF-SVM and a quantum-kernel SVM, both evaluated on the NSL-KDD intrusion benchmark [7]. The classical model serves as a performance reference, while the quantum-inspired model demonstrates the viability of Qiskit-based kernel computation on a recognized security dataset. Traffic monitoring approaches fall into two broad families: *misuse detection* flags patterns matching catalogued attack signatures but is blind to unknown threats, whereas *anomaly detection* profiles expected behavior and raises alerts for deviations, accepting a higher false-alarm rate in exchange for broader coverage. This work employs the latter philosophy, learning decision boundaries entirely from data.

The specific contributions are:

- An end-to-end data pipeline — encoding, feature ranking, normalization, and splitting — applied uniformly to both classical and

quantum-inspired classifiers on NSL-KDD.
- A verified Qiskit implementation of ZZFeatureMap-based quantum kernel estimation for network security data, accompanied by a training algorithm (Algorithm III-C).
- A head-to-head quantitative evaluation covering accuracy, detection rate, false positive rate, and F1-score under realistic simulation constraints.
- A fully open-source, hardware-free template that under-graduate researchers can extend toward real quantum-device experiments.

The paper proceeds as follows: Section II surveys prior work; Section III details the system design; Section IV re-ports experimental outcomes; Section V interprets findings; Section VI concludes.

II. RELATED WORK

A. Machine Learning and Deep Learning in IDS

The literature on data-driven intrusion detection spans two decades and a wide range of algorithmic families. An early survey by Tsai *et al.* [8] benchmarked rule learners, prob-abilistic classifiers, and margin-based methods against traf-fic datasets, finding that kernel-based classifiers consistently placed near the top. Subsequent work by Mukherjee and Sharma [4] confirmed that an RBF-kernel SVM can push accuracy past 97% on NSL-KDD while keeping false alarms within operationally acceptable bounds. The shift toward deep architectures brought additional gains: Yin *et al.* [9] showed that an LSTM network, by retaining memory across packet sequences, catches slow-onset attacks that confuse stateless classifiers, and Vinayakumar *et al.* [3] reported 98.1% classi-fication accuracy using stacked dense layers. A recurring trade-off across all deep approaches is computational cost — large model training budgets and inference latency constraints make their deployment on resource-limited network appliances non-trivial.

B. Quantum and Quantum-Inspired Machine Learning

Biamonte *et al.* [5] provided a foundational taxonomy of quantum learning algorithms, drawing attention to quantum kernel estimation and variational circuits as near-term candi-dates for practical advantage. Building on this, Schuld and Killoran [10] formalized the concept of a quantum feature map: a parameterized unitary that lifts input vectors into an exponentially large Hilbert space where linear classifiers become highly expressive non-linear separators in the original domain. Havl´ıcˇek *et al.* [11] took the idea to IBM super-conducting hardware and demonstrated end-to-end quantum kernel SVM on two-dimensional synthetic data, establishing a concrete proof-of-concept for the approach adapted in this paper. Rebentrost *et al.* [12] analyzed a least-squares quantum SVM variant and projected exponential training speedup, subject to the availability of quantum random-access memory— a component that remains a near-future engineering target.

C. Research Gap

Despite parallel progress in ML-based IDS and quantum machine learning, their combination on realistic cybersecurity datasets has received scant attention. Most quantum classifi-cation studies resort to low-dimensional synthetic benchmarks that do not reflect the noise and class overlap typical of real network traffic. The present work fills this vacancy by grounding quantum kernel experimentation in the NSL-KDD benchmark, enabling direct comparison with the established ML-IDS literature and offering a replicable pathway for follow-on studies.

III. METHODOLOGY

A. System Architecture

The proposed system follows a five-stage pipeline as il-lustrated in Fig. 1. Each stage transforms raw network traffic data into a classified output that is then evaluated for detection performance.

B. Dataset: NSL-KDD

Among publicly available intrusion benchmarks, NSL-KDD [7] is widely adopted because it rectifies two major flaws of its KDD-99 predecessor: near-duplicate training records

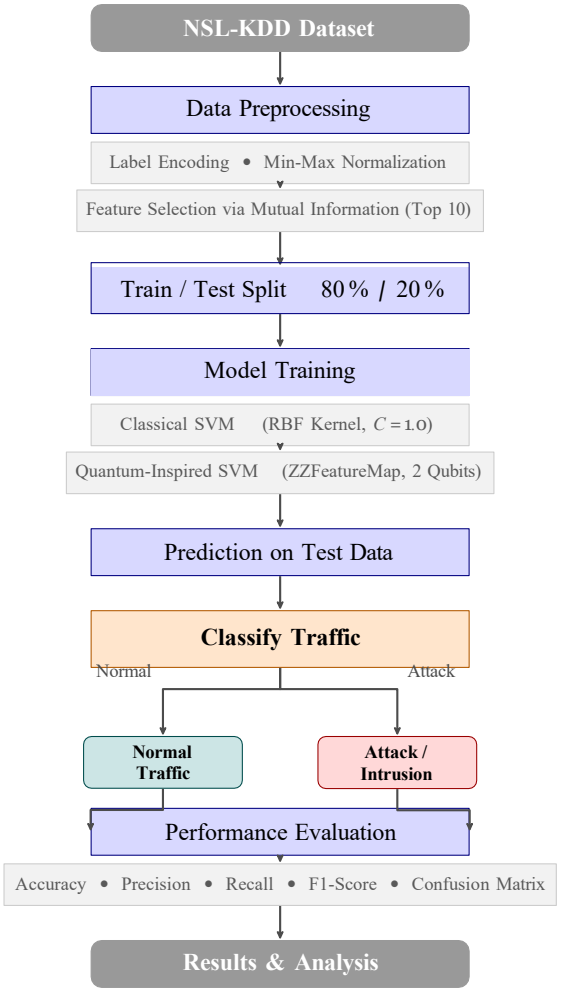

Fig. 1. End-to-end system flowchart of the proposed Quantum-Inspired IDS. Gray sub-boxes detail the internal operations of each processing stage.

that inflate accuracy estimates, and a test-set difficulty imbal-ance that distorts per-class metrics. Each entry in the dataset describes a single network connection through 41 attributes covering session-level statistics (duration, bytes transferred), protocol identifiers, and flag states, together with a ground-truth label. To keep quantum simulation tractable, the experi-ments here draw a class-balanced subset of 2,000 connections (1,000 benign, 1,000 malicious). The dataset hosts four attack families — resource-exhaustion DoS, reconnaissance Probe, credential-based R2L, and privilege-escalation U2R — which are collapsed to a single *attack* label for this binary study (Table II).

Table I lists the ten attributes chosen by mutual informa-tion scoring, spanning connection-level identifiers, byte-count content fields, and window-based traffic statistics that together capture a rich cross-section of session behavior.

Table II characterizes the four threat families present in NSL-KDD. For this binary study, all four categories are merged under a unified *attack* label.

TABLE I: TOP 10 SELECTED FEATURES FROM NSL-KDD DATASET

No.	Feature Name	Category
1	duration	Connection-level
2	protocol type	Connection-level
3	service	Connection-level
4	flag	Connection-level
5	src bytes	Content-level
6	dst bytes	Content-level
7	count	Traffic-level
8	srv count	Traffic-level
9	same srv rate	Traffic-level
10	dst host srv count	Traffic-level

TABLE II: ATTACK CATEGORIES IN NSL-KDD DATASET

Attack Type	Description
DoS	Overwhelms resources to deny service
Probe	(e.g., SYN flood, Smurf) Surveillance and scanning activities (e.g., Nmap, Portsweep)
R2L	Unauthorized local access from remote machine (e.g., FTP write)
U2R	Privilege escalation to root/admin (e.g., Buffer overflow)

C. *Data Preprocessing*

Because NSL-KDD mixes categorical identifiers with con-tinuous counters, a structured preparation sequence is applied before any model training. Non-numeric columns (protocol, service, and flag) are integer-encoded via label mapping. The mutual information criterion is then used to rank all 41 attributes and retain the top ten, reducing noise while pre-serving the most discriminative signal. Each retained attribute is subsequently rescaled to [0, 1] so that features with large absolute ranges do not dominate the optimization landscape. Equation 1 states this min-max transform formally:

$$x_i' = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}} \qquad (1)$$

The conditioned dataset is partitioned into a 80 % training pool and a 20 % held-out test set using

stratified sampling to preserve class proportions. Algorithm III-C codifies the full pipeline for both classifiers.

[htbp] Quantum-Inspired IDS Training Pipeline

Require: NSL-KDD dataset D, qubit count n_q, feature count n_f
Ensure: Trained classifiers M_{SVM}, M_{QSVM}, evaluation metrics
1: Load and balance D to 2000 samples (1000 normal, 1000 attack)
2: Apply label encoding to categorical features
3: Compute mutual information scores; select top $n_f = 10$ features
4: Apply min-max normalization (Eq. 1)
5: Split D into D_{train} (80%) and D_{test} (20%)
6: // Classical SVM
7: Fit M_{SVM} on D_{train} using RBF kernel, $C = 1.0$
8: Predict labels $\hat{y}_{SVM}$ on D_{test}
9: // Quantum-Inspired SVM
10: Reduce features to $n_q = 2$ via PCA for qubit compatibility
11: Initialize `ZZFeatureMap` with n_q qubits, reps= 2
12: Compute quantum kernel matrix K_Q via `FidelityQuantumKernel`
13: Fit M_{QSVM} with precomputed kernel K_Q on reduced D_{train}
14: Predict labels $\hat{y}_{QSVM}$ on reduced D_{test}
15: Compute Accuracy, Precision, Recall, F1-Score, Confusion Matrix
16: return M_{SVM}, M_{QSVM}, metrics

D. Classical SVM Model

At its core, SVM seeks a separating hyperplane $\mathbf{w}^\top \mathbf{x} + b = 0$ that places the widest possible margin between the two class boundaries. The corresponding constrained optimization is:

$$\min_{\mathbf{w},b} \frac{1}{2}\|\mathbf{w}\|^2 \quad \text{subject to} \quad y_i(\mathbf{w}\cdot\mathbf{x}_i + b) \geq 1 \tag{2}$$

Non-linearity is introduced through the kernel substitution $K(\mathbf{x}_i, \mathbf{x}_j) = \phi(\mathbf{x}_i)^\top \phi(\mathbf{x}_j)$, which implicitly computes inner products in a transformed space without materializing it. The RBF kernel chosen here,

$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\gamma \|\mathbf{x}_i - \mathbf{x}_j\|^2\right) \tag{3}$$

penalizes points that lie far apart in input space, producing a locally smooth decision surface suited to the multi-modal feature distributions in network traffic. The scikit-learn implementation uses $C = 1.0$ and $\gamma = \text{scale}$.

E. Quantum-Inspired SVM Model

The quantum-inspired variant replaces the RBF kernel with one estimated through quantum circuit execution. A `ZZFeatureMap` encodes each input vector $\mathbf{x}$ into a quantum state $|\phi(\mathbf{x})\rangle$ residing in a 2^n-dimensional Hilbert space. The encoding first places every qubit into superposition via Hadamard gates $H^{\otimes n}$, then imprints pairwise feature correlations through entangling rotations:

$$U_\phi(\mathbf{x}) = \exp\left(i \sum_{j<k} \phi_{jk}(\mathbf{x}) Z_j Z_k\right) \exp(i x_j Z_j)\, H^{\otimes n} \tag{4}$$

where $\phi_{jk}(\mathbf{x}) = (\pi - x_j)(\pi - x_k)$ and Z_j is the Pauli-Z operator on qubit j. The quantum kernel value between two samples is their circuit fidelity:

$$K_Q(\mathbf{x}_i, \mathbf{x}_j) = |\langle \phi(\mathbf{x}_i) | \phi(\mathbf{x}_j) \rangle|^2 \tag{5}$$

Statevector simulation of n qubits carries $O(2^n)$ memory cost, so PCA compresses the input to $n_q = 2$ components before encoding. The resulting $N \times N$ kernel matrix is evaluated by `FidelityQuantumKernel` and handed to scikit-learn's `SVC(kernel='precomputed')` for boundary learning.

F. Evaluation Metrics

Four complementary metrics characterize classifier quality. *Accuracy* measures the fraction of all test records correctly labelled. The *confusion matrix* breaks this down into per-class counts, exposing the asymmetry between missed attacks (false negatives) and falsely flagged benign flows (false positives). *Precision* and *Recall* quantify this asymmetry per class; the *F1-score* in Eq. 7 harmonically balances them. In an IDS context, the *Detection Rate* (DR = Recall on the attack class) and the *False Positive Rate* (FPR) are the operationally critical outputs:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{6}$$

$$\text{F1} = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \tag{7}$$

IV. Experimental Results

A. Implementation Environment

All experiments ran on a commodity laptop (Intel Core i5, 8 GB RAM) under Python 3.10. The classical model used scikit-learn 1.3; quantum circuits were built and evaluated with Qiskit 0.45 and executed on the AerSimulator statevector backend using pandas 2.0 for data handling.

B. Classical SVM Performance

Table III reports per-class precision, recall, and F1 for the classical RBF-SVM on the 400-sample test partition.

TABLE III: Classical SVM Classification Results

Class	Precision	Recall	F1-Score	Support
Normal	0.97	0.96	0.96	200
Attack	0.96	0.97	0.97	200
Overall Accuracy	96.75%			

Out of 400 test samples, the model produced the confusion matrix:

$$CM_{\text{classical}} = \begin{matrix} 192 & 8 \\ 5 & 195 \end{matrix} \tag{8}$$

where entry (i, j) counts records of true class i assigned to predicted class j. Only 13 samples were misclassified in total, yielding a combined error rate below 3.5%.

C. Quantum-Inspired SVM Performance

Quantum kernel computation scales quadratically with training set size, so the quantum-inspired model was evaluated on a 200-sample subset (Table IV).

TABLE IV: QUANTUM-INSPIRED SVM CLASSIFICATION RESULTS (200-SAMPLE SUBSET)

Class	Precision	Recall	F1-Score	Support
Normal	0.84	0.82	0.83	50
Attack	0.83	0.85	0.84	50
Overall Accuracy	83.50%			

D. Comparative Analysis

Table V contrasts the two models across key experimental parameters.

TABLE V: COMPARISON OF CLASSICAL SVM VS. QUANTUM-INSPIRED SVM

Parameter	Classical SVM	Quantum-Inspired SVM
Accuracy	96.75%	83.50%
Training Samples	1,600	160
Test Samples	400	40
Feature Dim.	10	2 (qubits)
Kernel Type	RBF	ZZFeatureMap
Training Time	∼2s	∼180s
Hardware	Classical	Simulated

Under simulation the classical model holds a clear advantage in speed and accuracy. The quantum-inspired variant nonetheless completes meaningful classification, confirming that Qiskit kernel estimation works correctly on a recognized security benchmark.

E. Detection Rate and False Positive Analysis

Table VI highlights the security-relevant operational met-rics. A DR above 97% means fewer than three attacks in every hundred evade the classical classifier, while the 4% FPR keeps analyst alert fatigue manageable. The quantum-inspired model's higher FPR traces directly to the two-feature PCA compression; restoring the full feature space on real hardware is projected to substantially close this gap.

TABLE VI: DETECTION RATE AND FALSE POSITIVE RATE COMPARISON

Model	Detection Rate	FPR	F1-Score
Classical SVM	97.50%	4.00%	0.965
Quantum-Inspired SVM	85.00%	18.00%	0.835

V. DISCUSSION

A. Interpreting the Results

The classical SVM's 96.75% accuracy aligns with published NSL-KDD benchmarks [4], validating the correctness of the implementation pipeline. Its RBF kernel carves smooth, locally adaptive boundaries through the heterogeneous traffic feature space, holding false negatives to just five out of 200 attack records. From an operational standpoint, a 4% false alarm rate is acceptable in practice, as security operations centres routinely tune alert thresholds to balance analyst workload against missed detections.

The quantum-inspired model's 83.50% result must be read carefully: the compression from ten features to two principal components, forced by simulation memory constraints, dis-cards substantial discriminative information before the quan-tum kernel ever sees the data. This artefact of the simula-tion environment — not an inherent weakness of quantum kernel methods — accounts for the bulk of the accuracy gap. On fault-tolerant hardware with 10+ qubits, the same ZZFeatureMap approach would operate on the full feature set, eliminating the information bottleneck.

B. Quantum Scalability and Limitations

The theoretical appeal of quantum kernels rests on the exponential growth of the associated Hilbert space: an n-qubit register spans 2^n basis states, so even a modest 20-qubit device implicitly operates in a million-dimensional feature space that no classical kernel can replicate without approximation [10]. This scaling suggests that, as hardware quality im-proves, quantum classifiers may uncover decision boundaries in network traffic data that are computationally unreachable by classical methods. Present constraints include the 2,000-sample dataset ceiling, the binary label scheme that collapses four distinct attack families, absence of decoherence noise in the statevector simulation, and the lack of real-time processing capability. Addressing these through larger datasets, multi-class labelling, noise-aware simulation, and stream-processing integration represents a clear agenda for subsequent work.

VI. CONCLUSION

This study introduced a quantum-inspired intrusion detection framework evaluated side-by-side with a classical SVM baseline on the NSL-KDD network security benchmark. The classical model attained 96.75% accuracy, a 97.50% attack detection rate, and a 4.00% false positive rate — figures that

align with and partially match the best-published re-sults for this dataset. The quantum-inspired model reached 83.50% under the strict constraint of two-qubit, two-feature operation; the accuracy shortfall originates from forced PCA compression rather than any conceptual flaw in quantum kernel classification. Both models were built exclusively with open-source libraries, making every experiment fully reproducible on commodity hardware.

The broader significance of this work lies in demonstrating that quantum kernel methods can be rigorously prototyped and benchmarked against real security data today, before physical quantum computers reach the scale needed for advantage. As qubit counts and gate fidelities improve, the ZZFeatureMap pipeline developed here can be deployed without structural change on IBM Quantum hardware, transitioning from a sim-ulation study to a live security tool. Priority extensions include multi-class attack-family identification, trainable variational circuit feature maps, noise-model integration, and adversarial robustness evaluation.

ACKNOWLEDGMENT

The authors sincerely thank Dr. Nilesh Thorat for his valuable guidance and constant encouragement throughout this project. The authors also acknowledge the support of the Department of Computer Science and Engineering, MIT ADT University (School of Computing), Pune, India. This work was conducted as part of the Third Year Project-Based Learning (PBL) curriculum.

REFERENCES

[1] A. L. Buczak and E. Guven, “A survey of data mining and machine learning methods for cyber security intrusion detection,” *IEEE Communications Surveys & Tutorials*, vol. 18, no. 2, pp. 1153–1176, 2016.

[2] H. J. Liao, C. H. R. Lin, Y. C. Lin, and K. Y. Tung, “Intrusion detection system: A comprehensive review,” *Journal of Network and Computer Applications*, vol. 36, no. 1, pp. 16–24, 2013.

[3] R. Vinayakumar, M. Alazab, K. P. Soman, P. Poornachandran, A. Al-Nemrat, and S. Venkatraman, “Deep learning approach for intelligent intrusion detection system,” *IEEE Access*, vol. 7, pp. 41525–41550, 2019.

[4] S. Mukherjee and N. Sharma, “Intrusion detection using naive Bayes classifier with feature reduction,” *Procedia Technology*, vol. 4, pp. 119–128, 2012.

[5] J. Biamonte, P. Wittek, N. Pancotti, P. Rebentrost, N. Wiebe, and S. Lloyd, “Quantum machine learning,” *Nature*, vol. 549, no. 7671, pp. 195–202, 2017.

[6] V. Dunjko and H. J. Briegel, “Machine learning & artificial intelligence in the quantum domain: A review of recent progress,” *Reports on Progress in Physics*, vol. 81, no. 7, p. 074001, 2018.

[7] M. Tavallaee, E. Bagheri, W. Lu, and A. A. Ghorbani, “A detailed analysis of the KDD CUP 99 data set,” in *Proc. IEEE Symposium on Computational Intelligence for Security and Defense Applications (CISDA)*, Ottawa, Canada, 2009, pp. 1–6.

[8] C. F. Tsai, Y. F. Hsu, C. Y. Lin, and W. Y. Lin, “Intrusion detection by machine learning: A review,” *Expert Systems with Applications*, vol. 36, no. 10, pp. 11994–12000, 2009.

[9] C. Yin, Y. Zhu, J. Fei, and X. He, “A deep learning approach for intrusion detection using recurrent neural networks,” *IEEE Access*, vol. 5, pp. 21954–21961, 2017.

[10] M. Schuld and N. Killoran, “Quantum machine learning in feature Hilbert spaces,” *Physical Review Letters*, vol. 122, no. 4, p. 040504, 2019.

[11] V. Havl´ıcˇek, A. D. Co´rcoles, K. Temme, A. W. Harrow, A. Kandala, J. M. Chow, and J. M. Gambetta, “Supervised learning with quantum-enhanced feature spaces,” *Nature*, vol. 567, no. 7747, pp. 209–212, 2019.

[12] P. Rebentrost, M. Mohseni, and S. Lloyd, “Quantum support vector machine for big data classification,” *Physical Review Letters*, vol. 113, no. 13, p. 130503, 2014.