

Where a domain-specific quantum feature map pays: a cross-domain replication on three cybersecurity datasets

Jean Marie Vianney Sindayigaya

Abstract—Claims about quantum machine learning are routinely established on a single dataset and generalised informally. This paper tests one such claim directly. Earlier work in this series found, under a leakage-controlled predeclared protocol, that a financial-transaction-specific quantum feature map does not outperform matched classical baselines on the Kaggle Credit-Card Fraud dataset. We freeze source-column contracts and an experimental protocol in advance, inherit every predicted direction from that published result, and replicate on CICIDS-2017 (2.8 million flows) and UNSW-NB15 (257,673 records) — six paired comparisons under one Holm–Bonferroni correction. The replication is partial, and the failure is informative. The map’s advantage over naive amplitude encoding replicates on both datasets and is now established across three datasets and two domains. But the prediction that a matched RBF-SVM would win is significantly *reversed* on UNSW-NB15, where the quantum kernel wins by a median 0.021 F1 — identifying a regime in which the approach pays. That result is bounded: a random forest on identical cohorts outperforms the quantum kernel on both datasets, so no general claim of quantum advantage is supportable. Two further findings weaken the premise of domain-specific encoding itself. The advantage over generic quantum encodings differs in *sign* between the two network datasets, and the unanimous entanglement strength selected on financial data does not reproduce on either. We conclude that quantum kernel results reported on one dataset should not be assumed to transfer, in either direction, and that predeclared cross-domain replication is cheap enough to be routine. Balanced-cohort metrics are benchmark values, not deployment-prevalence estimates.

Keywords—Quantum Machine Learning; Quantum Kernel Methods; Cross-Domain Replication; Network Intrusion Detection; Financial Fraud Detection; Predeclared Protocols; Reproducibility

I. INTRODUCTION

QUANTUM machine-learning results are usually established on one dataset. A feature map is proposed, benchmarked, and its behaviour described as though it were a property of the map rather than of the pairing between that map and that data. The field has few replications, and almost no predeclared ones, so it is rarely possible to say whether a reported outcome would survive a change of domain — in either direction.

This paper tests one such result directly, and the test is unusually clean because the original was itself predeclared.

Author is with Warsaw University of Technology, Poland (e-mail: jean_marie_vianney.sindayigaya.dokt@pw.edu.pl).

Earlier work in this series introduced FT-FM, a quantum feature map whose gate structure is derived axiomatically from a five-class taxonomy of financial-transaction data [1], and FT-FM-lite, which adds a single entanglement-strength parameter β [2]. Under a leakage-controlled protocol on the Kaggle Credit-Card Fraud dataset [3], that work found the quantum kernel significantly *below* a matched RBF-SVM [4] and significantly *above* exact amplitude encoding, with $\beta = 1$ selected unanimously across ten seeds. Two companion studies then found that a gated hybrid built on the map contributed no net corrections [5] and that its entanglement structure does not explain its decisions [6].

Every one of those findings rests on a single dataset in a single domain. We ask whether they transfer, and we ask a second question that bears on the premise of the map itself. FT-FM’s axioms are derived from financial structure: log-scaled monetary magnitudes, periodic time, velocity counters. Network flows have different semantics. If the domain-specific derivation does real work, the map’s advantage over *generic* quantum encodings should be larger inside its design domain than outside it. If that advantage is unchanged, the derivation never bought anything.

Our design is a replication, not an exploration. Source-column contracts for CICIDS-2017 [7] and UNSW-NB15 [8] were frozen before any experiment, together with a protocol whose six paired comparisons take their predicted directions from the published Kaggle result. Nothing about the analysis was chosen after seeing an outcome; the reference classical baseline was named in advance for exactly that reason.

The contributions are as follows. 1) We publish frozen source-column contracts mapping two network-flow datasets onto the FT-FM taxonomy, including three data hazards that silently corrupt naive pipelines (Section III). 2) We report a partial replication in which the map’s advantage over amplitude encoding holds on both datasets, now across three datasets and two domains. 3) We report a significant *reversal*: on UNSW-NB15 the quantum kernel beats the matched RBF-SVM, identifying a regime where the approach pays — while showing that a random forest beats it on both datasets, so no general advantage claim survives (Section IV). 4) We show that the advantage over generic encodings differs in *sign* between the two network datasets, and that the unanimous β of



the financial domain does not reproduce in either (Sections IV–V).

Section II reviews quantum kernels and the replication problem, Section III gives the frozen contracts and protocol, Section IV the results, Section V discusses and bounds them, and Section VI concludes.

II. BACKGROUND AND RELATED WORK

A. Quantum kernels and where their claims come from

A quantum feature map induces the fidelity kernel $K(x, x') = |\langle \psi(x) | \psi(x') \rangle|^2$, which a support-vector classifier consumes as a precomputed Gram matrix [9], [10]; supervised quantum models are kernel methods in a precise sense [11]. The dominant failure mode is exponential concentration, in which kernel values cease to depend on input similarity and the Gram matrix becomes uninformative [12]. Applied studies in both of this paper’s domains — card-payment fraud and network intrusion — report quantum-kernel pipelines, but almost always on a single dataset, with separately tuned classical comparators and without multiplicity control. That combination makes it impossible to tell whether a reported difference reflects the encoding or the protocol.

B. The replication gap

Replication is standard practice in adjacent fields and rare here. The specific hazard for quantum feature maps is that a map’s behaviour is a property of the map–data pairing, not of the map: concentration depends on the empirical input distribution, and an encoding tuned to one feature taxonomy has no guarantee of behaving the same way on another. A published result therefore licenses a claim about one dataset unless replication says otherwise — and, as this paper shows, the claim can fail in the direction that favours the quantum model as easily as against it.

We treat the earlier Kaggle result [2] as the object of replication and inherit its predicted directions verbatim, so that the present analysis has no free parameters at the point where results are interpreted.

C. The datasets and their known hazards

CICIDS-2017 comprises labelled flow records generated over five days of emulated traffic [7]; the machine-learning distribution supplies 79 flow statistics per record. UNSW-NB15 supplies 49 features generated by the IXIA PerfectStorm tool with an authors’ feature dictionary declaring each feature’s type [8], which we use as the authority for taxonomy typing rather than inferring types ourselves.

Both distributions carry defects that a naive pipeline absorbs silently, and that we fix in the contracts of Section III: non-UTF-8 bytes in CICIDS-2017’s attack labels, partition filenames in UNSW-NB15 that contradict their contents, and eight groups of duplicated columns in CICIDS-2017 — some bit-identical, one pair separated only by rounding at the sixth decimal place. Documenting these is a secondary contribution: any study using these datasets faces them.

III. FROZEN CONTRACTS AND PREDECLARED PROTOCOL

Both the data handling and the experiment were fixed in written specifications before any experiment was run on either dataset, and neither has been changed in response to an outcome.

A. Source-column contracts

Each contract fixes six things: the source distribution and its per-file digest, the label definition, the taxonomy typing, the fixed features, the selection rule for the remainder, and the cohort construction. Both mirror the structure of the Kaggle contract they replicate, so that differences in result cannot be attributed to differences in protocol.

Taxonomy typing. Seven qubits each, one feature per qubit. CICIDS-2017 fixes `Flow Duration` as monetary — a strictly positive heavy-tailed magnitude for which log-scaling is the correct treatment, which is what the monetary block applies — and `Total Fwd Packets` as velocity, with five relational features selected. UNSW-NB15 fixes `proto` as categorical (binary: most frequent protocol versus other), `sbytes` as monetary and `spkts` as velocity, with four relational selected; the nominal typing follows the authors’ own feature dictionary. Neither dataset supplies a periodic feature, so axiom A1 is not exercised; UNSW-NB15 supplies the first categorical feature in the series, so A3 is exercised for the first time.

Selection. Free features are chosen by training-only Fisher score, identical in rule to the Kaggle contract: squared difference of class means over summed class variances, zero variances ϵ -regularised, ties broken by source column order.

Three data hazards, fixed in advance. CICIDS-2017’s attack labels contain non-UTF-8 bytes, so the contract fixes `latin-1` decoding and whitespace normalisation before any class is defined. UNSW-NB15’s partition files are named contrary to their sizes — `training-set.csv` holds 82,332 records and `testing-set.csv` holds 175,341 — so the contract trusts neither name, pools both, and applies its own split; results are therefore internally comparable but not directly comparable to work using the published partition. CICIDS-2017 contains eight groups of duplicated columns, one of them ten columns wide, and one further pair separated only by rounding at 10^{-6} absolute (a relative 8.2×10^{-10} , below the precision the file stores). Candidates identical at source precision are dropped, keeping source order; the rule compares column values only and never consults labels or scores. Without it the selector spends several of its seven qubits re-encoding one signal.

B. Cohorts and arms

Per seed: stratified 60/20/20 split before any transform is fitted; selection fitted on training rows only; deterministic 1:1 case-control cohorts of 240/80/80. Ten seeds. Because every cohort is balanced by design, all reported F1 values are case-control benchmark metrics, not deployment-prevalence estimates — and the three datasets’ natural positive rates differ by orders of magnitude (0.0017, 0.197, 0.639), so cross-dataset comparisons are not comparisons of deployment difficulty.

Treatment: FT-FM-lite with β selected per seed on validation from $\{0, 0.1, 0.25, 0.5, 0.75, 1\}$ by maximum F1, ties broken by PR-AUC then lower β , refit on train-plus-validation and evaluated once on test. Quantum controls: FT-FM at full strength, Qiskit’s ZZFeatureMap [13], and exact amplitude encoding. Classical baselines: L_2 logistic regression, MLP, RBF-SVM [4], and random forest [14], on identical cohorts and features. All evaluation is exact statevector simulation; no hardware claim is made.

Two amendments were required before the run and are dated in the specification. UNSW-NB15’s categorical was reduced from four buckets to two because the matched quantum controls take one qubit per source feature, so a wider categorical would have run the controls on six qubits against the treatment’s seven, destroying the matching the comparison depends on. The controls were extended to encode a categorical at all, mapping index k of K to $k/(K-1)$: a generic encoding has no notion of categorical orthogonality and treats the index as a magnitude, which is precisely the deficiency axiom A3 was formulated against, so this is the correct control behaviour rather than a convenience. Both were discovered by executing the controls, not by observing a result, and neither affects the published Kaggle results, whose schemas contain no categorical.

C. Confirmatory family

One family of six two-sided Wilcoxon signed-rank comparisons [15] on paired per-seed test F1 under a single Holm–Bonferroni correction at familywise $\alpha = 0.05$ [16] — three per dataset. Directionality lives in the claim rule: a *win* requires Holm significance *and* a median difference in the predicted direction, so a significant result in the wrong direction is reported as such with equal prominence. Predictions are inherited from the Kaggle result: R1, versus the RBF-SVM, predicts the quantum kernel lower; R2, versus amplitude encoding, predicts it higher; R3, versus ZZFeatureMap, predicts no difference.

The RBF-SVM is named in advance as the confirmatory comparator because the earlier work established it as the arm that significantly beat the quantum kernel. Fixing it beforehand prevents a “best classical baseline” being chosen after the fact; Section IV reports what that choice costs in interpretation.

Effective pair counts, dropped-zero counts, and exact attainable p -floors are reported per comparison. With ten pairs the exact two-sided floor is $2^{-9} \approx 0.00195$ against a rank-one Holm bar of $0.05/6 \approx 0.00833$, so every comparison is attainable provided at most one seed ties.

D. Secondary analysis

The *domain-specificity premium* is $P = F1(\text{FT-FM-lite}) - \max(F1(\text{ZZFeatureMap}), F1(\text{amplitude}))$, the margin over the better generic encoding. It is paired within a dataset but not across datasets, since seeds index different data. Cross-dataset comparison of the premium is reported descriptively and carries *no causal claim about domain*: prevalence, dimensionality and difficulty vary together, so a difference cannot be attributed to domain alone.

TABLE I
CONFIRMATORY FAMILY: TWO-SIDED WILCOXON SIGNED-RANK TESTS ON PAIRED PER-SEED TEST F1, ONE HOLM–BONFERRONI CORRECTION ACROSS BOTH DATASETS, FAMILYWISE $\alpha = 0.05$. DIRECTIONS ARE INHERITED FROM THE PUBLISHED KAGGLE RESULT; A WIN REQUIRES SIGNIFICANCE *and* THE PREDICTED DIRECTION.

	Control	Median Δ	Raw p	Verdict
<i>CICIDS-2017</i>				
R1-C	RBF-SVM	−0.0224	0.203	as predicted, n.s.
R2-C	Amplitude	+0.0588	0.0039	replicates
R3-C	ZZFeatureMap	−0.0639	0.131	null, as on Kaggle
<i>UNSW-NB15</i>				
R1-U	RBF-SVM	+0.0213	0.0059	REVERSED
R2-U	Amplitude	+0.0064	0.0156	replicates
R3-U	ZZFeatureMap	+0.0648	0.0020	beats generic

IV. RESULTS

All values derive from one archived run of ten seeds per dataset under the canonical Python 3.11 environment with a clean working tree, carrying per-file dataset digests. Balanced cohorts make every F1 a case-control benchmark metric.

A. Confirmatory family

Table I reports the six predeclared comparisons. Four are Holm-significant, and the replication is partial in an informative way.

R2 replicates on both datasets. The domain-specific map beats exact amplitude encoding wherever it has been tested — now three datasets across two domains. This is the most robust finding of the series.

R1-U is a significant reversal. On UNSW-NB15 the quantum kernel does not fall below the matched RBF-SVM as predicted; it rises above it, by a median +0.0213 F1, Holm-significant. Under the claim rule this is recorded as a loss — a significant result contrary to prediction — but its substantive content is that a regime exists in which the approach pays. On CICIDS-2017 the same comparison runs in the predicted direction and does not reach significance.

B. The bound on that result

Fig. 1 places every arm side by side, and the reversal must be read against it. **A random forest outperforms FT-FM-lite on both datasets**, decisively on CICIDS-2017 (0.9557 versus 0.8122) and clearly on UNSW-NB15 (0.8659 versus 0.8305).

The RBF-SVM was predeclared as the confirmatory comparator because the earlier work had established it as the arm that beat the quantum kernel. Fixing it in advance protected the analysis from post-hoc baseline selection, but it is not the strongest classical model on either network dataset. **No claim that the quantum kernel beats classical methods in general is supportable from these data.** The supported claim is exact and narrow: against the predeclared comparator, on UNSW-NB15, the quantum kernel significantly wins.

C. The advantage over generic encodings changes sign

The domain-specificity premium is +0.0081 on UNSW-NB15, with a 95% bootstrap interval of [+0.0026, +0.0149]

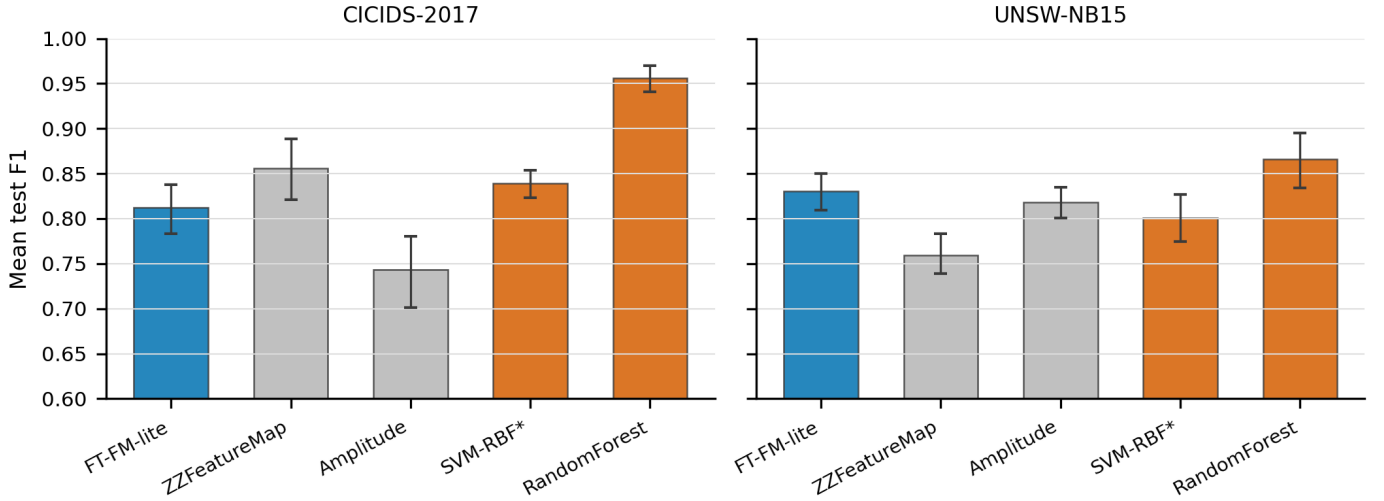


Fig. 1. Mean test F1 by arm on both datasets, with 95% bootstrap intervals over ten seeds. Blue: the domain-specific quantum map. Grey: generic quantum encodings. Orange: classical baselines. The asterisk marks the predeclared confirmatory comparator. FT-FM-lite exceeds the RBF-SVM on UNSW-NB15, and a random forest exceeds FT-FM-lite on both datasets.

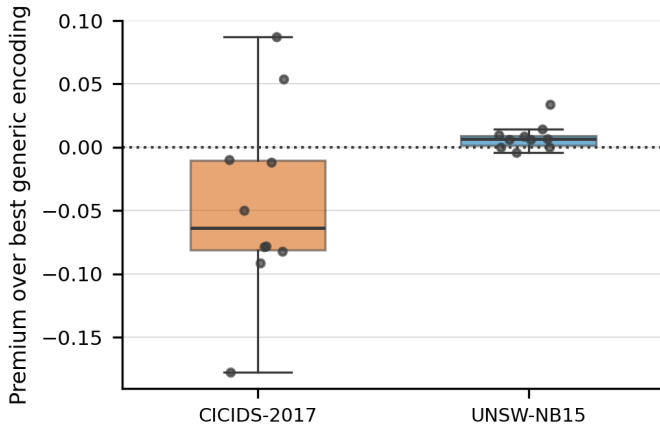


Fig. 2. Per-seed premium of FT-FM-lite over the better generic quantum encoding. Positive on UNSW-NB15, negative on CICIDS-2017. Dotted line marks parity.

excluding zero, and -0.0439 on CICIDS-2017, interval $[-0.0888, +0.0019]$. Fig. 2 shows the per-seed distributions.

On CICIDS-2017 the *generic* ZZFeatureMap outperforms the domain-specific map (0.8561 versus 0.8122). The advantage that motivates a domain-specific construction is therefore not a property of the map alone: it is present in one network domain and reversed in the other. As predeclared, no causal attribution to domain is made — these datasets differ in prevalence, dimensionality and difficulty simultaneously, and the comparison is unpaired.

D. The selected entanglement regime does not transfer

Fig. 3 shows selected β across all three datasets. The Kaggle result selected $\beta = 1$ in every one of ten seeds. That unanimity does not reproduce: CICIDS-2017 spreads across all six grid values, and UNSW-NB15 is bimodal at the endpoints ($\beta = 1$ five times, $\beta = 0$ three times, $\beta = 0.25$ twice).

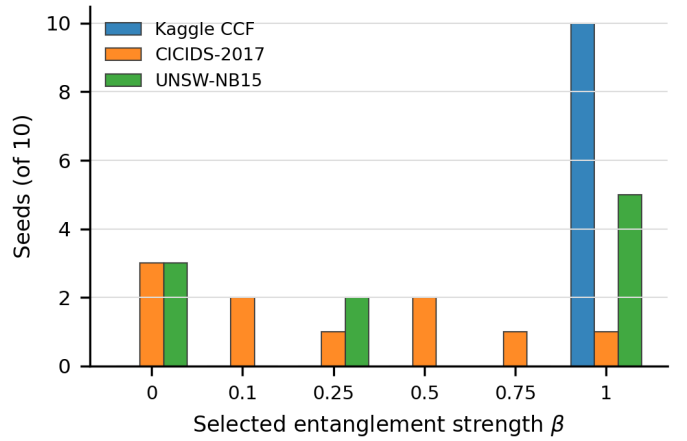


Fig. 3. Entanglement strength selected by the validation protocol, over ten seeds per dataset. The unanimous $\beta = 1$ of the financial domain does not reproduce on either network dataset.

This is direct cross-domain support for the architectural argument that motivated the tunable parameter in the first place — that entanglement strength must be validated per dataset rather than fixed by design — and it is stronger evidence than a single dataset could provide.

Two comparisons leave pairs to exact ties — R1-C retained nine effective pairs and R2-U eight, the rest all ten. At those counts the exact attainable two-sided p -floors are $2^{-8} = 0.0039$ and $2^{-7} = 0.0078$, both still below the rank-one Holm bar of $0.05/6 \approx 0.0083$, so every comparison in Table I remained capable of reaching significance. No null reported here is an artefact of the test’s resolution.

V. DISCUSSION

A. What replicated, what reversed, and why it matters

One prediction replicated cleanly on both datasets: the domain-specific map beats naive amplitude encoding. Across

three datasets and two domains that is now the best-supported claim in this series, and it is the modest claim — that a structured encoding beats an unstructured one.

The prediction that failed is the interesting one. On UNSW-NB15 the quantum kernel significantly *beat* the matched RBF-SVM, reversing the direction inherited from the financial dataset. Had this study been run on Kaggle CCF alone, as the earlier work was, the conclusion would have been that the approach does not compete with classical baselines. Had it been run on UNSW-NB15 alone, the conclusion would have been the opposite. Neither single-dataset conclusion would have been wrong about its own data, and both would have been wrong as general statements. That is the argument for cross-domain replication in a compact form.

We resist the obvious over-reading. A random forest beats the quantum kernel on both datasets, so the reversal identifies a regime relative to a predeclared comparator, not a quantum advantage. The honest summary is that the domain-specific quantum kernel is competitive with some classical baselines on some datasets, reliably better than unstructured quantum encodings, and not better than a well-chosen classical model anywhere we have looked.

B. The premise of domain-specific encoding is weakened, not refuted

FT-FM’s construction rests on the claim that deriving gate structure from a data-type taxonomy buys something a generic encoding cannot supply. The premium measured here changes sign between two datasets from the *same* broad domain: positive on UNSW-NB15, negative on CICIDS-2017, where a generic ZZFeatureMap outperforms the domain-specific map outright.

Two readings are consistent with this and are not separated by our design. The taxonomy mapping may be doing real work where the flow statistics genuinely carry the typed structure the axioms assume, and failing where they do not — CICIDS-2017’s selected features are dominated by correlated backward-packet-length statistics, which are relational in name but carry little of the monetary or velocity structure the map is built around. Or the premium may reflect nothing about domain at all, and simply track dataset difficulty. Distinguishing these requires holding difficulty fixed while varying taxonomy fit, which no pair of public datasets provides. We report the observation and decline the causal claim.

C. Implications for practice

Three transfer beyond this series. First, entanglement strength must be validated per dataset: the unanimous selection on financial data reproduces on neither network dataset, so any published value is a property of a pairing and not of a map. Second, a predeclared comparator protects against post-hoc baseline selection but does not substitute for reporting the full baseline set — our own headline would read very differently without the random forest beside it, and we would regard a paper that omitted it as having overclaimed. Third, cross-domain replication under a frozen protocol cost a fraction of the effort of the original studies, because the protocol already existed; the barrier is discipline rather than compute.

D. Limitations

Six bound these claims. First, ten seeds are repeated splits of one pool per dataset, so intervals quantify protocol stability rather than inference over populations. Second, every cohort is balanced by design and the three datasets’ natural positive rates span 0.0017 to 0.639, so no metric here is a deployment-prevalence estimate and cross-dataset comparisons are not comparisons of deployment difficulty. Third, the premium comparison across datasets is unpaired and confounded, as stated. Fourth, UNSW-NB15 cohorts come from a pooled re-split rather than the authors’ published partition, so results are internally comparable but not directly comparable to work using that partition. Fifth, all evaluation is exact statevector simulation; finite-shot behaviour was characterised separately in this series and shot noise would affect the quantum arms only. Sixth, two datasets in one broad domain do not establish behaviour across cybersecurity data generally, and the seven-qubit budget bounds what any of these encodings can represent.

E. Reproducibility

The archived artifact carries the repository commit, a clean working-tree flag, pinned package versions, per-file SHA-256 digests for both datasets, per-seed selected features and Fisher scores, the duplicate-column groups that were dropped, and every arm’s per-seed result. Both frozen specifications are versioned alongside it, with their two dated amendments and the reason each was required.

VI. CONCLUSION

We asked whether a published, predeclared result about a financial-transaction-specific quantum feature map transfers to network intrusion detection, and we answered under contracts and a protocol frozen before any experiment, inheriting every predicted direction from the original.

The replication is partial, and both halves are informative. The map’s advantage over naive amplitude encoding holds on both new datasets, making it the best-supported claim of the series across three datasets and two domains. But the prediction that a matched RBF-SVM would win is significantly reversed on UNSW-NB15, where the quantum kernel wins by a median 0.021 F1 — the first regime in this programme where the approach pays. That result is bounded rather than celebrated: a random forest on identical cohorts beats the quantum kernel on both datasets, so no general claim of quantum advantage follows.

Two further results bear on the premise of domain-specific encoding itself. The advantage over generic quantum encodings changes *sign* between two datasets from the same broad domain, with a generic ZZFeatureMap winning outright on CICIDS-2017. And the entanglement strength selected unanimously on financial data reproduces on neither network dataset, confirming across domains that this parameter is a property of a map–data pairing rather than of a map.

The methodological conclusion is the one we would press hardest. A study of this map on Kaggle CCF alone concludes that it does not compete with classical baselines; a study on

UNSW-NB15 alone concludes the opposite. Neither is wrong about its own data and both are wrong as general claims. Quantum-kernel results reported on a single dataset should not be assumed to transfer in either direction, and predeclared cross-domain replication — which cost a small fraction of the original studies here, because the protocol already existed — is cheap enough to become routine.

REFERENCES

- [1] J. M. V. Sindayigaya, “FT-FM: A financial-transaction-specific quantum feature map for variational quantum classifiers in fraud detection,” *International Journal of Electronics and Telecommunications*, vol. 72, no. 3, pp. 1–8, 2026, <https://doi.org/10.24425/ijet.2026.1729>. [Online]. Available: <https://doi.org/10.24425/ijet.2026.1729>
- [2] —, “FT-FM-lite: Tunable-entanglement refinement and empirical benchmarking on financial-cybersecurity datasets,” 2026, manuscript submitted for publication; no DOI assigned.
- [3] A. Dal Pozzolo, O. Caelen, R. A. Johnson, and G. Bontempi, “Calibrating probability with undersampling for unbalanced classification,” in *Proc. IEEE Symposium Series on Computational Intelligence*, 2015, pp. 159–166, <https://doi.org/10.1109/SSCI.2015.33>. [Online]. Available: <https://doi.org/10.1109/SSCI.2015.33>
- [4] C. Cortes and V. Vapnik, “Support-vector networks,” *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995, <https://doi.org/10.1007/BF00994018>. [Online]. Available: <https://doi.org/10.1007/BF00994018>
- [5] J. M. V. Sindayigaya, “Q-RAD: A gated residual architecture for quantum-assisted fraud detection,” 2026, manuscript in preparation; no DOI assigned.
- [6] —, “Entanglement entropy is not a faithful explanation: Auditing a financial quantum feature map,” 2026, manuscript in preparation; no DOI assigned.
- [7] I. Sharafaldin, A. H. Lashkari, and A. A. Ghorbani, “Toward generating a new intrusion detection dataset and intrusion traffic characterization,” in *Proc. 4th International Conference on Information Systems Security and Privacy*, 2018, pp. 108–116, <https://doi.org/10.5220/0006639801080116>. [Online]. Available: <https://doi.org/10.5220/0006639801080116>
- [8] N. Moustafa and J. Slay, “UNSW-NB15: A comprehensive data set for network intrusion detection systems,” in *Proc. Military Communications and Information Systems Conference*, 2015, pp. 1–6, <https://doi.org/10.1109/MilCIS.2015.7348942>. [Online]. Available: <https://doi.org/10.1109/MilCIS.2015.7348942>
- [9] V. Havlíček, A. D. Córcoles, K. Temme, A. W. Harrow, A. Kandala, J. M. Chow, and J. M. Gambetta, “Supervised learning with quantum-enhanced feature spaces,” *Nature*, vol. 567, pp. 209–212, 2019, <https://doi.org/10.1038/s41586-019-0980-2>. [Online]. Available: <https://doi.org/10.1038/s41586-019-0980-2>
- [10] M. Schuld and N. Killoran, “Quantum machine learning in feature hilbert spaces,” *Physical Review Letters*, vol. 122, no. 4, p. 040504, 2019, <https://doi.org/10.1103/PhysRevLett.122.040504>. [Online]. Available: <https://doi.org/10.1103/PhysRevLett.122.040504>
- [11] M. Schuld, “Supervised quantum machine learning models are kernel methods,” 2021, <https://doi.org/10.48550/arXiv.2101.11020>. [Online]. Available: <https://doi.org/10.48550/arXiv.2101.11020>
- [12] S. Thanasilp, S. Wang, M. Cerezo, and Z. Holmes, “Exponential concentration in quantum kernel methods,” *Nature Communications*, vol. 15, p. 5200, 2024, <https://doi.org/10.1038/s41467-024-49287-w>. [Online]. Available: <https://doi.org/10.1038/s41467-024-49287-w>
- [13] Qiskit contributors, “Qiskit: An open-source framework for quantum computing,” 2023, <https://doi.org/10.5281/zenodo.2573505>. [Online]. Available: <https://doi.org/10.5281/zenodo.2573505>
- [14] L. Breiman, “Random forests,” *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001, <https://doi.org/10.1023/A:1010933404324>. [Online]. Available: <https://doi.org/10.1023/A:1010933404324>
- [15] F. Wilcoxon, “Individual comparisons by ranking methods,” *Biometrics Bulletin*, vol. 1, no. 6, pp. 80–83, 1945, <https://doi.org/10.2307/3001968>. [Online]. Available: <https://doi.org/10.2307/3001968>
- [16] S. Holm, “A simple sequentially rejective multiple test procedure,” *Scandinavian Journal of Statistics*, vol. 6, no. 2, pp. 65–70, 1979, <https://doi.org/10.2307/4615733>. [Online]. Available: <https://doi.org/10.2307/4615733>