

FT-FM-lite: tunable-entanglement refinement and empirical benchmarking on financial-cybersecurity datasets

Jean Marie Vianney Sindayigaya

Abstract—The financial-transaction feature map FT-FM, introduced in the first paper of this series, is empirically more expressive than generic quantum encodings but operates in a near-maximally entangled regime in which quantum-kernel values concentrate and kernel-based classification fails. This paper introduces FT-FM-lite, a refinement of FT-FM with a single entanglement-strength hyperparameter $\beta \in [0, 1]$ that preserves the domain-informed encoding axioms while allowing the feature map to operate in the intermediate expressibility–trainability regime; $\beta = 1$ recovers FT-FM exactly. A validation-based protocol selects β per training split. In a ten-seed synthetic diagnostic against exactly implemented matched controls, validation-selected FT-FM-lite attains a mean held-out F1 of 0.620 (95% bootstrap CI [0.530, 0.709]), exceeding exact amplitude encoding (0.362) and Qiskit’s ZZFeatureMap (0.033) with Holm–Bonferroni-corrected significance, while its advantage over full-strength FT-FM (0.490) is not significant after correction. We further specify a leakage-controlled, case-control benchmarking protocol with paired Wilcoxon significance testing and apply it to the Kaggle Credit-Card Fraud dataset over ten repeated stratified case-control splits. In contrast to the synthetic diagnostic, validation selection chose full strength ($\beta = 1$) in every seed, and FT-FM-lite attained a mean case-control test F1 of 0.891 (95% bootstrap CI [0.870, 0.907]; ROC-AUC 0.971; PR-AUC 0.976) — significantly above exact amplitude encoding and significantly below a matched RBF-SVM after Holm–Bonferroni correction, and statistically indistinguishable from Qiskit’s ZZFeatureMap and the remaining classical baselines. Balanced-cohort metrics are benchmark values, not deployment-prevalence estimates.

Keywords—Quantum Machine Learning; Quantum Feature Map; Entanglement; Expressibility; Financial Fraud Detection; Network Intrusion Detection

I. INTRODUCTION

FINANCIAL institutions and the network infrastructures they operate face an escalating landscape of transaction-level and network-level attacks — payment fraud, money laundering, intrusion, and adversarially adapted abuse — under extreme class imbalance, concept drift, and heterogeneous feature types. Quantum Machine Learning (QML), and in particular quantum-kernel methods built on parameterized quantum feature maps [1], has been proposed as a source of richer inductive biases for such data; yet generic encodings carry no explicit structure for financial-transaction records.

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

The first paper of this series introduced FT-FM, a financial-transaction-specific quantum feature map whose gate structure is derived axiomatically from a five-class taxonomy of transaction data (monetary, cyclic, categorical, velocity, relational) [2]. Its published empirical characterisation confirmed the predicted expressibility advantage over generic feature maps (Kullback–Leibler divergence to the Haar distribution of 4.01 ± 1.16 , versus 21.24 ± 0.79 and 19.53 ± 0.43 for ZZ-like and amplitude-like baselines at matched qubit count), but also exposed a kernel-concentration phenomenon: FT-FM states are near-maximally entangled (Meyer–Wallach $Q = 0.948 \pm 0.004$ [3]), fidelity kernel values concentrate regardless of input similarity, and kernel-SVM classification on the synthetic benchmark collapses [2]. This is an instance of the general expressibility–trainability tension of quantum models [4]–[6].

This paper closes the loop opened by that finding. We introduce *FT-FM-lite*, a refinement of FT-FM with a single entanglement-strength hyperparameter $\beta \in [0, 1]$ that continuously modulates the domain-informed entangling layer while leaving the per-class encoding axes untouched. At $\beta = 1$ the construction is statevector-equivalent to the published FT-FM; at $\beta = 0$ it degenerates to a structured product-state encoding. Because the informative regime is not knowable a priori, β is treated as a hyperparameter and selected on validation data only, under a protocol predeclared before any test evaluation.

The contributions of this paper are as follows. 1) We define the FT-FM-lite construction and analyse which of the six FT-FM axioms are preserved, relaxed, or modulated as functions of β (Section III). 2) We specify a leakage-controlled evaluation methodology — validation-only β selection, exactly implemented matched quantum controls, case-control cohorts, and paired nonparametric significance testing with Holm–Bonferroni correction [7], [8] (Section IV). 3) We report a ten-seed synthetic diagnostic in which validation-selected FT-FM-lite significantly exceeds exact amplitude encoding and Qiskit’s ZZFeatureMap on held-out data, and we report honestly that its advantage over full-strength FT-FM is not significant after correction and that reduced entanglement alone does not eliminate kernel concentration on this generator (Section V). 4) We report the real-data case-control benchmark on the Kaggle Credit-Card Fraud dataset promised in [2], with exactly implemented quantum controls, four classical



baselines, and Holm-corrected paired significance tests, and we find that the informative entanglement regime reverses between the synthetic generator and the real data (Section VI); the CICIDS-2017 and UNSW-NB15 benchmarks await predeclared column contracts and are deferred to future work in this series.

The remainder of the paper is organised as follows. Section II reviews background and related work. Section III defines FT-FM-lite. Section IV describes the experimental methodology. Section V reports the synthetic diagnostic, Section VI the real-data benchmark. Section VII discusses findings and limitations, and Section VIII concludes.

II. BACKGROUND AND RELATED WORK

A. FT-FM: an axiomatic financial-transaction feature map

FT-FM [2] allocates distinct encoding strategies — rotation-axis choice, scaling, and re-uploading depth — to the five classes of the financial-transaction taxonomy, and couples qubits according to a domain-informed entanglement graph rather than a fixed topology. Its construction is derived from six axioms: cyclic faithfulness (A1), monetary log-scale monotonicity (A2), categorical orthogonality (A3), domain-informed entanglement (A4), a NISQ qubit/depth budget (A5), and trainability (A6).

We fix notation, identical to [2]. A transaction is a tuple $x = (m, c, k, v, r)$, where m collects monetary features, c cyclic features (each with period T), k categorical indices, v velocity counters, and r relational features. Each non-categorical feature occupies one qubit; a categorical feature with K values occupies $\lceil \log_2 K \rceil$ qubits under binary encoding. Writing $U_{\text{FT}}(x)$ for the FT-FM circuit on n qubits, the encoded state is $|\psi(x)\rangle = U_{\text{FT}}(x)|0\rangle^{\otimes n}$, and the induced fidelity kernel is

$$K(x, x') = |\langle \psi(x) | \psi(x') \rangle|^2, \quad (1)$$

which is used by a support-vector classifier over the precomputed Gram matrix [1], [9]. The normalised encoding values are denoted $\mu_j \in [0, 1]$ (log-scaled monetary), $\theta_j = 2\pi c_j/T_j$ (cyclic angle), $\nu_j \in [0, 1]$ (power-transformed velocity), and $\rho_j \in [0, 1]$ (min-max relational).

B. Expressibility, trainability, and kernel concentration

The expressibility of a parameterized circuit — how closely its output distribution approximates the Haar measure [10] — is inversely related to gradient magnitude: highly expressive circuits generically exhibit barren plateaus [4], [5]. For quantum-kernel methods the analogous failure mode is exponential concentration of kernel values, which renders the Gram matrix uninformative regardless of the downstream classifier [6]. The published FT-FM characterisation exhibits precisely this signature [2]. FT-FM-lite is designed to let the practitioner move along the expressibility–trainability curve with a single interpretable parameter, in the spirit of the kernel-bandwidth controls studied for generic quantum kernels, where a global rescaling of encoding angles governs the transition from an overly expressive, concentrating regime to a generalising one [11], [12]. FT-FM-lite differs from bandwidth scaling

in what it rescales: the parameter β acts on the *entangling* interactions of the domain-informed graph only, leaving the per-class single-qubit encoding axes — and hence axioms A1–A3 — exactly intact.

C. QML for financial fraud and network intrusion detection

Applied quantum-kernel studies exist for both target domains of this series. Grossi et al. report an end-to-end QSVM pipeline for card-payment fraud with quantum-assisted feature selection, evaluated on an industrial payment dataset [13]; Kalinin and Krundyshev evaluate QSVM and quantum convolutional networks for network intrusion detection on flow data [14]. These studies establish feasibility, but their comparisons are typically *unmatched*: quantum and classical models are tuned separately, encodings are generic, per-split pairing is not enforced, and significance testing with multiplicity control is rare. The protocol of Section IV is designed to close exactly these gaps — exactly implemented matched quantum controls sharing every split, cohort, and selected feature with the treatment; predeclared classical baselines; leakage-controlled fitting; and Holm-corrected paired Wilcoxon tests [7], [8] — so that any reported difference is attributable to the feature map rather than to protocol asymmetry.

III. THE FT-FM-LITE CONSTRUCTION

A. Definition

FT-FM-lite is identical to FT-FM in its per-class encoding layer and differs only in the entangling interactions: every entangling angle of the domain-informed entanglement graph is scaled by a global factor $\beta \in [0, 1]$,

$$U_{\text{lite}}(x; \beta) = U_{\text{ent}}(\beta \cdot \Theta(x)) U_{\text{enc}}(x), \quad (2)$$

where U_{enc} is the FT-FM encoding layer and $U_{\text{ent}}(\Theta(x))$ its entangling layer with data-dependent angles $\Theta(x)$. At gate level, using the notation of Section II and exactly as realised in the tested implementation, the encoding layer applies to each feature’s qubit(s):

$$\begin{aligned} \text{monetary: } & R_y(\pi\mu_j) R_z(2\pi\mu_j^2), \\ \text{cyclic: } & R_x(\cos\theta_j) R_y(\sin\theta_j), \\ \text{categorical: } & X^{b_{q-1}} \otimes \dots \otimes X^{b_0}, \\ & \text{then } H \text{ on the least-significant qubit,} \\ \text{velocity: } & R_y(\pi\nu_j), \\ \text{relational: } & R_y(\pi\rho_j), \end{aligned} \quad (3)$$

none of which depends on β . The entangling interactions, all scaled by β , are:

$$\begin{aligned} \text{velocity–cyclic: } & \text{CRZ}(\beta\pi\nu_j) \text{ from each velocity qubit} \\ & \text{to the first cyclic qubit,} \\ \text{relational–monetary: } & e^{-i(\beta\pi\rho_j/2)Z \otimes Z} \text{ between each relational} \\ & \text{qubit and the first monetary qubit,} \\ & \text{compiled as } CX - R_z(\beta\pi\rho_j) - CX, \\ \text{ring: } & (\mathbb{I} \otimes H) \text{CP}(\beta\pi) (\mathbb{I} \otimes H) \\ & \text{on each pair } (q, q+1 \bmod n). \end{aligned} \quad (4)$$

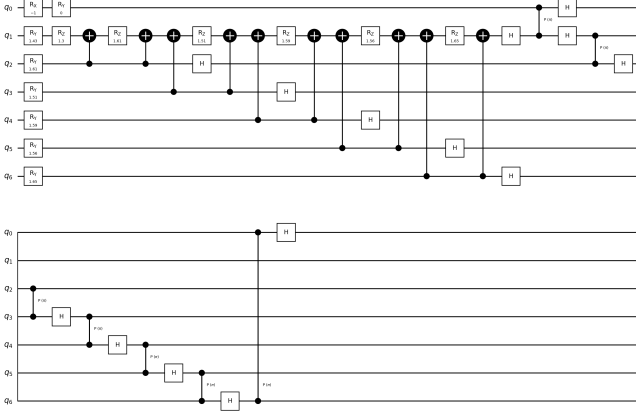


Fig. 1. FT-FM-lite circuit for the seven-qubit Kaggle CCF input contract (cyclic Time, monetary Amount, five relational V components) at a representative input, drawn from the tested implementation. Single-qubit encoding gates are β -independent; the CX- R_z -CX relational-monetary couplings and the Hadamard-conjugated controlled-phase ring carry the β -scaled angles.

Fig. 1 shows the resulting circuit for the seven-qubit Kaggle CCF input contract of Section IV, drawn directly from the tested implementation.

Two endpoints anchor the family, and their properties, together with axiom preservation, can be stated precisely.

- P1. *Endpoint equivalence.* At $\beta = 1$, FT-FM-lite is statevector-equivalent to the published FT-FM. *Sketch:* $CP(\pi) = CZ$ and $H CZ H$ (conjugation on the target) equals CNOT, so the ring of Eq. (4) reduces to FT-FM’s CNOT ring, while the CRZ and ZZ angles reduce to their published values. The equivalence is verified by unit test to statevector precision.
- P2. *Product-state limit.* At $\beta = 0$, every entangling gate is the identity ($CRZ(0) = \mathbb{I}$, $CX R_z(0) CX = \mathbb{I}$, $H CP(0) H = \mathbb{I}$), so $|\psi(x; 0)\rangle$ is a product of single-qubit encodings: the Meyer–Wallach measure [3] satisfies $Q = 0$ and the fidelity kernel factorises into a product of single-qubit fidelities.
- P3. *A1–A3 hold for every β .* All dependence on a cyclic feature enters through $(\cos \theta_j, \sin \theta_j)$, which is T_j -periodic, and no entangling angle depends on a cyclic value, so A1 holds. Monetary values enter only through the log-scaled μ_j , so A2 holds. For A3, two inputs differing only in a categorical index produce states whose categorical subregisters are orthogonal after the encoding layer; since both inputs then experience the *same* entangling unitary (its angles depend only on the unchanged ν_j and ρ_j), and unitaries preserve inner products, the encoded states remain orthogonal for every β .
- P4. *A4 modulated, A5 unchanged.* The entanglement-graph topology is β -independent and every coupling angle is linear in β , so A4 is preserved in structure and modulated in strength; gate count and depth do not depend on β , so the A5 NISQ budget is exactly that of FT-FM.

Axiom A6 (trainability) is the motivation for the construction: decreasing β moves the map continuously from the near-Haar regime in which the published FT-FM operates toward

the product regime of P2, in which kernel concentration driven by global entanglement cannot occur [6]. Where on this curve the informative regime lies is an empirical question, answered per dataset in Sections V and VI.

B. Validation-based selection of β

Because the informative regime depends on the data distribution, β is selected from the grid $\{0, 0.1, 0.25, 0.5, 0.75, 1.0\}$ by maximum validation F1, with ties broken by validation PR-AUC and then by lower β . The selected map is refit on the union of training and validation data and evaluated exactly once on the untouched test split. This makes FT-FM-lite an *operational* refinement: the practitioner does not need to know the correct entanglement regime in advance.

IV. EXPERIMENTAL METHODOLOGY

A. Common protocol

All experiments use a class-balanced support-vector classifier ($C = 1.0$) over a precomputed fidelity kernel $K_{ij} = |\langle \psi(x_i) | \psi(x_j) \rangle|^2$ obtained from statevector simulation [15]. For each seed, data are split into stratified train/validation/test partitions before any transform is fitted; feature scaling, feature selection, and imbalance handling are fitted on training rows only. The β -selection protocol of Section III-C is applied per seed. Alongside FT-FM-lite, three exactly implemented quantum controls are retrained and tested on the same per-seed splits: (i) full-strength FT-FM, realised as the $\beta = 1$ endpoint that is statevector-equivalent to [2]; (ii) Qiskit’s ZZFeatureMap (linear entanglement, one repetition) [15]; and (iii) exact amplitude encoding via state initialisation with normalisation and zero padding. Four predeclared classical baselines are trained on the identical per-seed cohorts and selected features: an L_2 -regularised logistic regression ($C = 1$, balanced class weights); a multilayer perceptron with hidden layers (64, 32), ReLU activations, Adam optimisation, and fixed early stopping; an RBF-kernel SVM ($C = 1$, $\gamma = \text{scale}$, balanced class weights); and a random forest with 100 trees and balanced class weights.

Statistical comparisons use two-sided Wilcoxon signed-rank tests [7] on paired per-seed test scores, with Holm–Bonferroni correction across the comparison family [8]. Repeated seeds are repeated stratified splits of one dataset; the resulting intervals and tests quantify stability of the predeclared protocol and are not inference from independent populations.

B. Synthetic diagnostic

The synthetic generator of [2] produces labelled transactions matching the five-class taxonomy. For each of ten seeds, 300 transactions are generated and split 60/20/20. The generic controls use a deterministic, no-fit vectorisation (log-scaled amount, periodic time of day, three merchant-category bits, bounded velocity, three fixed-bound relational values), so that no test statistic can inform any modelling choice.

TABLE I
FT-FM-LITE STRENGTH SWEEP: VALIDATION METRICS OVER TEN SEEDS (SYNTHETIC DIAGNOSTIC). CI IS THE 95% BOOTSTRAP INTERVAL OF MEAN VALIDATION F1.

β	Val. F1	95% CI	KL to Haar	Mean Q	Eff. rank
0.00	0.612	[0.469, 0.758]	58.06	0.000	23.22
0.10	0.585	[0.439, 0.735]	57.81	0.026	23.32
0.25	0.570	[0.421, 0.718]	56.63	0.151	23.84
0.50	0.551	[0.410, 0.695]	52.87	0.514	25.74
0.75	0.546	[0.375, 0.715]	47.42	0.855	28.95
1.00	0.459	[0.287, 0.619]	41.35	0.950	33.56

C. Real-data benchmark

Of the three datasets named in [2], this paper benchmarks Kaggle Credit-Card Fraud [16]; CICIDS-2017 [17] and UNSW-NB15 [18] require predeclared source-column contracts that have not yet been frozen and are deferred to future work in this series (Section VII).

1) *Kaggle Credit-Card Fraud*: A predeclared compact-input contract controls leakage: raw rows are split 60/20/20 (stratified) before any fitting; Time and Amount are always retained, and five of the PCA components V1–V28 are ranked by the squared difference of training-class means divided by the sum of training-class variances (computed on training rows only; zero variances ϵ -regularised; ties broken by source-column order), giving a seven-feature input within the A5 qubit budget. Deterministic 1:1 case-control cohorts of 240/80/80 rows are drawn from the splits; β is chosen on the validation cohort; the selected model is refit on train-plus-validation cohorts and evaluated once on the test cohort. Every quantum and classical control uses the same selected features, cohorts, and seed. The runner records source-row indices and the SHA-256 digest of the input file; the canonical run uses ten seeds over the full dataset (284,807 rows, 492 positive). Because test cohorts are balanced by design, all reported F1, precision, and PR-AUC values are case-control benchmark metrics, not deployment-prevalence estimates.

V. SYNTHETIC DIAGNOSTIC RESULTS

All values in this section derive from a single frozen diagnostic run of ten seeds (Python 3.12.3, Qiskit 0.46.3, statevector simulation), archived with its raw log in the repository; the real-data benchmark of Section VI was subsequently executed under the canonical Python 3.11 environment with the same Qiskit stack.

A. Strength sweep

Table I reports the validation-split behaviour of FT-FM-lite across the β grid, aggregated over ten seeds. Increasing β increases entanglement capability Q and effective kernel rank and decreases the KL divergence to Haar, but mean validation F1 declines monotonically; validation selection chose $\beta = 0$ for four seeds, $\beta = 0.1$ for three, and $\beta = 0.75$ for three, and never selected full strength. Fig. 2 visualises the trade-off. Section VI shows that this monotone decline is generator-specific and reverses on real data.

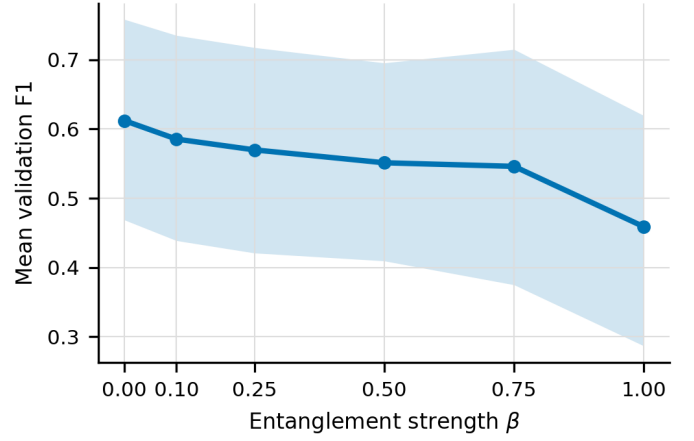


Fig. 2. Validation F1 versus entanglement strength β (ten seeds, synthetic diagnostic; shading is the 95% bootstrap interval of the mean).

TABLE II
HELD-OUT TEST F1 OVER TEN SEEDS (SYNTHETIC DIAGNOSTIC).

Model	Mean test F1	95% CI
FT-FM-lite (validation-selected)	0.620	[0.530, 0.709]
FT-FM full strength ($\beta = 1$)	0.490	[0.310, 0.650]
Exact amplitude encoding	0.362	[0.326, 0.398]
Qiskit ZZFeatureMap	0.033	[0.000, 0.100]

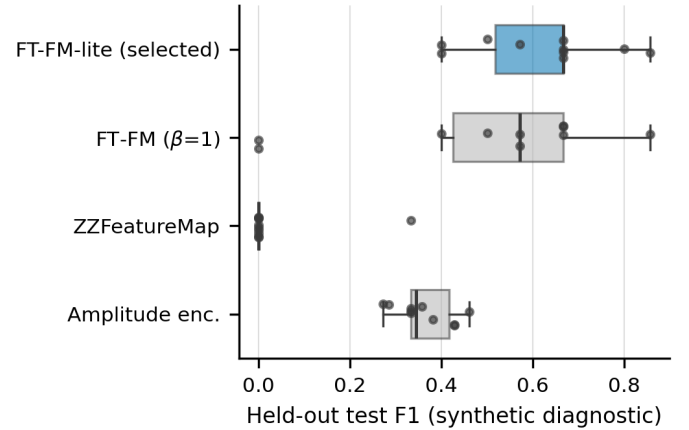


Fig. 3. Held-out test F1 of validation-selected FT-FM-lite against exact matched controls (ten seeds, synthetic diagnostic).

B. Held-out comparison against exact controls

Table II reports single-shot test evaluations. Validation-selected FT-FM-lite attains mean test F1 0.620 (95% bootstrap CI [0.530, 0.709]), mean test ROC-AUC 0.932 ([0.904, 0.958]), and mean test PR-AUC 0.688 ([0.601, 0.786]). Fig. 3 shows the per-model distributions.

C. Significance

Table III reports two-sided Wilcoxon signed-rank tests on the ten paired seed results with Holm–Bonferroni correction across the three F1 comparisons. FT-FM-lite exceeds both exact generic controls after correction; its advantage over full-strength FT-FM is positive in median but not significant.

TABLE III
PAIRED WILCOXON SIGNED-RANK TESTS ON TEST F1 (TEN PAIRS;
HOLM-BONFERRONI FAMILYWISE $\alpha = 0.05$).

Control	Median $\Delta F1$	Raw p	Holm sig.
Exact amplitude encoding	0.222	0.00195	Yes
FT-FM full strength ($\beta = 1$)	0.048	0.0625	No
Qiskit ZZFeatureMap	0.619	0.00195	Yes

TABLE IV
FT-FM-LITE STRENGTH SWEEP ON KAGGLE CCF: VALIDATION METRICS
OVER TEN SEEDS (CASE-CONTROL COHORTS). CI IS THE 95% BOOTSTRAP
INTERVAL OF MEAN VALIDATION F1. THE KL ESTIMATOR (75 BINS) IS
NOT COMPARABLE TO THE SYNTHETIC-DIAGNOSTIC OR PAPER 1
ESTIMATORS.

β	Val. F1	95% CI	KL to Haar	Mean Q	Eff. rank
0.00	0.856	[0.826, 0.885]	101.51	0.000	3.34
0.10	0.858	[0.828, 0.887]	101.49	0.027	3.34
0.25	0.859	[0.830, 0.886]	101.18	0.159	3.35
0.50	0.868	[0.842, 0.892]	100.01	0.538	3.38
0.75	0.885	[0.865, 0.906]	98.20	0.876	3.43
1.00	0.899	[0.884, 0.916]	95.98	0.968	3.50

D. Kernel concentration is not eliminated

The off-diagonal kernel median remained numerically near zero at every tested strength. This run therefore does not support the simple claim that weaker entanglement eliminates kernel concentration; the mechanism by which lower β improves classification on this generator appears to operate through the informative kernel entries rather than through bulk deconcentration. The logged kernel statistics quantify this: across the β grid, mean centred kernel-target alignment rises only from 0.051 to 0.055 and mean effective rank from 23.2 to 33.6, both moving *monotonically toward* the full-strength endpoint even though validation F1 moves monotonically away from it. On this generator, neither bulk statistic tracks task performance — a caution against selecting β by kernel diagnostics instead of validation performance.

VI. REAL-DATA BENCHMARK RESULTS

This section reports the Kaggle Credit-Card Fraud benchmark under the protocol of Section IV. All values derive from a single archived ten-seed run under the canonical environment (Python 3.11, Qiskit 0.46.3, statevector simulation), with the repository commit, the dataset’s SHA-256 digest, and per-seed source-row indices recorded in the run artifact. Because all cohorts are balanced 1:1 case-control samples, every F1 and PR-AUC value in this section is a case-control benchmark metric, not a deployment-prevalence estimate; the raw positive rate of the dataset is $492/284,807 \approx 0.17\%$.

A. Strength sweep and selection

Table IV reports validation-cohort behaviour across the β grid, and Fig. 4 visualises it. In contrast to the synthetic diagnostic of Section V, mean validation F1 increases monotonically with β , and the selection rule chose $\beta = 1$ in all ten seeds. Feature selection was stable: every seed retained $\{\text{Time}, \text{Amount}, V_{14}, V_4, V_{11}, V_{12}, V_{10}\}$, with only the relative rank of V_{11} and V_{12} varying.

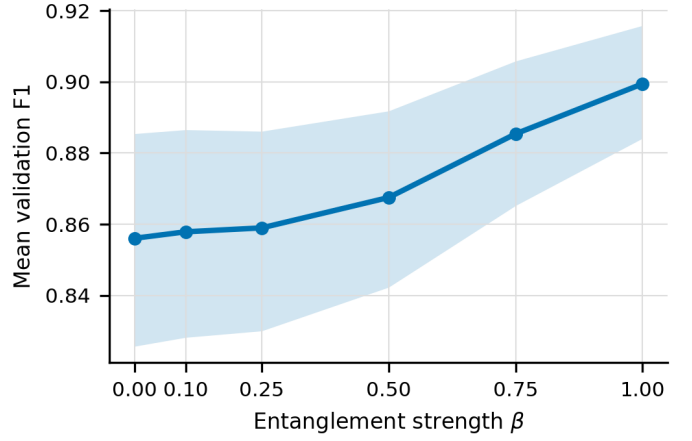


Fig. 4. Validation F1 versus entanglement strength β on Kaggle CCF (ten seeds, case-control cohorts; shading is the 95% bootstrap interval of the mean). The trend is the reverse of the synthetic diagnostic (Fig. 2).

TABLE V
KAGGLE CCF CASE-CONTROL BENCHMARK: HELD-OUT TEST METRICS
OVER TEN REPEATED STRATIFIED SPLITS (MEAN $\pm$ STD; BALANCED TEST
COHORTS — VALUES ARE BENCHMARK METRICS, NOT
DEPLOYMENT-PREVALENCE ESTIMATES).

Model	F1	PR-AUC	ROC-AUC
FT-FM-lite (val.-selected)	0.891 ± 0.031	0.976 ± 0.013	0.971 ± 0.015
FT-FM full strength ($\beta = 1$)	0.891 ± 0.031	0.976 ± 0.013	0.971 ± 0.015
Qiskit ZZFeatureMap	0.870 ± 0.060	0.961 ± 0.021	0.950 ± 0.027
Exact amplitude encoding	0.767 ± 0.048	0.974 ± 0.014	0.966 ± 0.018
Classical LogisticRegression	0.915 ± 0.022	0.976 ± 0.009	0.967 ± 0.014
Classical MLP	0.840 ± 0.065	0.945 ± 0.050	0.933 ± 0.049
Classical SVM (RBF)	0.918 ± 0.032	0.980 ± 0.008	0.971 ± 0.013
Classical RandomForest	0.923 ± 0.026	0.976 ± 0.009	0.967 ± 0.012

B. Held-out comparison

Table V reports single-shot test-cohort evaluations, and Fig. 5 the per-model distributions. Validation-selected FT-FM-lite attains mean test F1 0.891 (95% bootstrap CI [0.870, 0.907]), mean test ROC-AUC 0.971 ([0.961, 0.979]), and mean test PR-AUC 0.976 ([0.968, 0.982]). Because every seed selected $\beta = 1$, the validation-selected model coincides with the full-strength FT-FM control on this dataset, and their rows are identical by construction.

C. Significance

Table VI reports two-sided Wilcoxon signed-rank tests on the ten paired seed F1 scores, Holm-Bonferroni-corrected across the seven-comparison family. Two comparisons survive correction, in opposite directions: FT-FM-lite significantly exceeds exact amplitude encoding (median $\Delta F1 = +0.115$) and is significantly below the RBF-SVM (median $\Delta F1 = -0.015$). Its differences against Qiskit’s ZZFeatureMap and the remaining classical baselines are not significant after correction. We therefore make no claim of quantum advantage on this dataset; the supported claim is that the domain-informed quantum-kernel pipeline is competitive with strong classical baselines under an identical leakage-controlled contract, while significantly exceeding the generic amplitude-encoding quantum control.

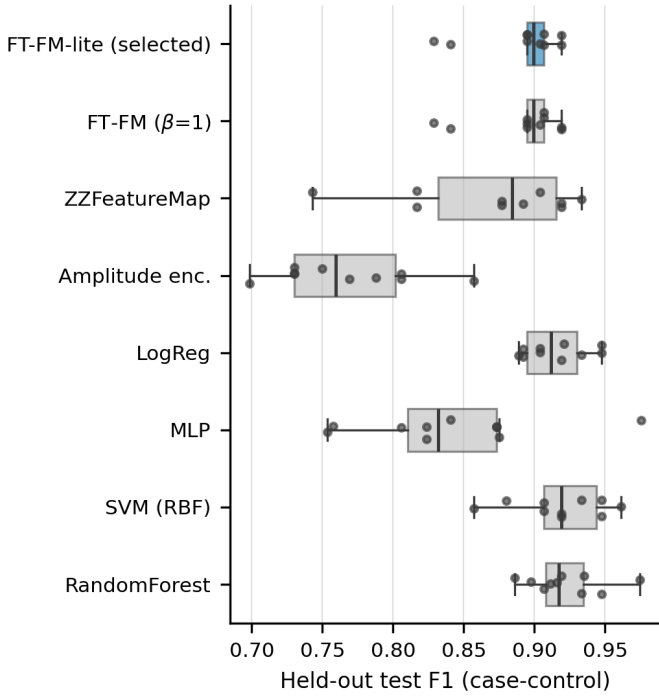


Fig. 5. Held-out case-control test F1 on Kaggle CCF for validation-selected FT-FM-lite, the exact quantum controls, and the classical baselines (ten seeds each). The FT-FM full-strength distribution is identical to the selected model because $\beta = 1$ was selected in every seed.

TABLE VI
KAGGLE CCF: PAIRED WILCOXON SIGNED-RANK TESTS ON CASE-CONTROL TEST F1 (TEN PAIRS; HOLM-BONFERRONI FAMILYWISE $\alpha = 0.05$ ACROSS SEVEN COMPARISONS). $\Delta F1$ IS FT-FM-LITE MINUS CONTROL.

Control	Median $\Delta F1$	Raw p	Holm sig.
Exact amplitude encoding	+0.115	0.00195	Yes
Classical SVM (RBF)	-0.015	0.00391	Yes
Classical RandomForest	-0.025	0.0195	No
Classical MLP	+0.045	0.0371	No
Classical LogisticRegression	-0.020	0.0391	No
Qiskit ZZFeatureMap	+0.013	0.496	No
FT-FM full strength ($\beta = 1$)	0.000	1.0	No

D. Reversal of the informative entanglement regime

The real-data sweep inverts the synthetic finding: on the synthetic generator, validation F1 declined with β and full strength was never selected; on Kaggle CCF, validation F1 rises monotonically with β and full strength is selected in every seed. The regime that was pathological on the synthetic generator is the best-performing regime on this real dataset. This is the central empirical argument for treating entanglement strength as a validated hyperparameter rather than a fixed design choice: the informative regime is dataset-dependent and not knowable a priori, and the predeclared selection protocol adapted correctly in both cases.

Kernel structure sharpens the point. The FT-FM-family kernel remains heavily concentrated on real data — mean effective rank rises only from 3.34 to 3.50 (of a possible 80) across the grid, versus 24.75 for ZZFeatureMap — yet FT-FM-lite matches or exceeds ZZFeatureMap’s F1 with roughly half its across-seed standard deviation (0.031 vs. 0.060).

Centred kernel–target alignment increases monotonically with β (0.068 to 0.095), tracking the validation-F1 gain, where on the synthetic generator alignment and task performance moved in opposite directions. As in the synthetic diagnostic, task performance is governed by the informative minority of kernel entries rather than by bulk deconcentration.

VII. DISCUSSION

A. What the entanglement-strength parameter buys

The synthetic diagnostic supports the operational claim of FT-FM-lite: validation-based selection reliably avoids the pathological full-strength regime and recovers classification performance that significantly exceeds exactly implemented generic controls. It does not support the stronger claim that FT-FM-lite significantly exceeds full-strength FT-FM — the median advantage is positive but not significant on ten paired seeds. The Kaggle CCF benchmark of Section VI refines this picture: on real case-control data the informative regime reverses, full strength is selected in every seed, and the value of the β parameter lies in *adaptivity* — the same predeclared selection protocol avoided full strength where it was pathological (synthetic) and embraced it where it was optimal (real data). With respect to hypothesis H2-FM-lite, the real-data evidence is mixed: FT-FM-lite significantly exceeds exact amplitude encoding, is statistically indistinguishable from Qiskit’s ZZFeatureMap, and is significantly below a matched RBF-SVM after Holm correction. The hypothesis is therefore supported against generic amplitude encoding but not against tuned classical baselines on this dataset.

B. On the mechanism

Reduced entanglement strength did not deconcentrate the bulk of the kernel on the synthetic generator: the off-diagonal kernel median stayed near zero at all β . The observed F1 gains must therefore be attributed to changes in the informative minority of kernel entries. This nuance matters for the theory of domain-specific feature maps: the practical benefit of moving along the expressibility–trainability curve may be visible in task metrics before it is visible in bulk kernel statistics [4], [6]. The Kaggle CCF run reinforces this: kernel effective rank stays in the 3.3–3.5 range across the entire β grid (versus 24.75 for ZZFeatureMap), while centred kernel–target alignment rises monotonically with β from 0.068 to 0.095 in step with validation F1. On both data sources, classification quality tracks the alignment of the informative kernel entries, not bulk kernel rank — and on the synthetic generator not even alignment, cautioning against selecting β by kernel diagnostics instead of validated task performance.

C. Limitations and threats to validity

Six limitations bound the claims of this paper. First, repeated seeds are repeated stratified splits of a single dataset; the reported intervals quantify protocol stability, not inference over independent populations. Second, all cohorts are balanced by design, so reported metrics are case-control benchmark values and must not be read as deployment-prevalence estimates.

Third, all quantum evaluations use noiseless statevector simulation; shot-budget and hardware-noise effects are out of scope and deferred to the Q-RAD paper of this series. Fourth, the synthetic diagnostic KL estimator differs from the Paper 1 estimator, so KL values are not comparable across the two papers. Fifth, the Kaggle CCF V-features are PCA components of an undisclosed original feature set [16], so the domain-typed encoding operates on anonymised projections rather than raw transaction semantics; only Time and Amount receive their taxonomy types on direct domain grounds. Sixth, results on a single real dataset do not establish generality across financial-cybersecurity domains: the CICIDS-2017 [17] and UNSW-NB15 [18] benchmarks promised in [2] require predeclared source-column contracts that have not yet been frozen, and remain future work in this series.

VIII. CONCLUSION

This paper introduced FT-FM-lite, a tunable-entanglement refinement of the financial-transaction feature map FT-FM, together with a leakage-controlled evaluation methodology built on validation-only hyperparameter selection, exactly implemented matched controls, and multiplicity-corrected paired statistics. On a ten-seed synthetic diagnostic, validation-selected FT-FM-lite attained mean held-out F1 0.620 and significantly exceeded exact amplitude encoding and Qiskit's ZZFeatureMap, while never selecting the pathological full-strength regime identified in the first paper of this series. On the Kaggle Credit-Card Fraud case-control benchmark, the informative regime reversed: validation selection chose full strength in every seed, and FT-FM-lite attained mean case-control test F1 0.891 (ROC-AUC 0.971, PR-AUC 0.976), significantly exceeding exact amplitude encoding while remaining significantly below a matched RBF-SVM after Holm correction. Together, the two studies establish entanglement strength as a dataset-dependent, validation-selectable hyperparameter rather than a fixed design choice, and they set an honest baseline: the domain-informed quantum kernel is competitive with, but does not surpass, strong classical baselines on this dataset.

The next paper in this series will embed FT-FM-lite as the quantum branch of Q-RAD, a hybrid residual architecture in which a learned gating function routes only ambiguous examples to the quantum module under realistic shot-budget constraints.

REFERENCES

- [1] 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>
- [2] 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>
- [3] D. A. Meyer and N. R. Wallach, "Global entanglement in multiparticle systems," *Journal of Mathematical Physics*, vol. 43, no. 9, pp. 4273–4278, 2002, <https://doi.org/10.1063/1.1497700>. [Online]. Available: <https://doi.org/10.1063/1.1497700>
- [4] Z. Holmes, K. Sharma, M. Cerezo, and P. J. Coles, "Connecting ansatz expressibility to gradient magnitudes and barren plateaus," *PRX Quantum*, vol. 3, no. 1, p. 010313, 2022, <https://doi.org/10.1103/PRXQuantum.3.010313>. [Online]. Available: <https://doi.org/10.1103/PRXQuantum.3.010313>
- [5] J. R. McClean, S. Boixo, V. N. Smelyanskiy, R. Babbush, and H. Neven, "Barren plateaus in quantum neural network training landscapes," *Nature Communications*, vol. 9, p. 4812, 2018, <https://doi.org/10.1038/s41467-018-07090-4>. [Online]. Available: <https://doi.org/10.1038/s41467-018-07090-4>
- [6] 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>
- [7] 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>
- [8] 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>
- [9] 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>
- [10] S. Sim, P. D. Johnson, and A. Aspuru-Guzik, "Expressibility and entangling capability of parameterized quantum circuits for hybrid quantum-classical algorithms," *Advanced Quantum Technologies*, vol. 2, no. 12, p. 1900070, 2019, <https://doi.org/10.1002/qute.201900070>. [Online]. Available: <https://doi.org/10.1002/qute.201900070>
- [11] R. Shaydulin and S. M. Wild, "Importance of kernel bandwidth in quantum machine learning," *Physical Review A*, vol. 106, no. 4, p. 042407, 2022, <https://doi.org/10.1103/PhysRevA.106.042407>. [Online]. Available: <https://doi.org/10.1103/PhysRevA.106.042407>
- [12] A. Canatar, E. Peters, C. Pehlevan, S. M. Wild, and R. Shaydulin, "Bandwidth enables generalization in quantum kernel models," 2023, transactions on Machine Learning Research. <https://doi.org/10.48550/arXiv.2206.06686>. [Online]. Available: <https://doi.org/10.48550/arXiv.2206.06686>
- [13] M. Grossi, N. Ibrahim, V. Radescu, R. Lored, K. Voigt, C. Von Altroch, and A. Rudnik, "Mixed quantum-classical method for fraud detection with quantum feature selection," *IEEE Transactions on Quantum Engineering*, vol. 3, pp. 1–12, 2022, <https://doi.org/10.1109/TQE.2022.3213474>. [Online]. Available: <https://doi.org/10.1109/TQE.2022.3213474>
- [14] M. Kalinin and V. Krundyshev, "Security intrusion detection using quantum machine learning techniques," *Journal of Computer Virology and Hacking Techniques*, vol. 19, no. 1, pp. 125–136, 2023, <https://doi.org/10.1007/s11416-022-00435-0>. [Online]. Available: <https://doi.org/10.1007/s11416-022-00435-0>
- [15] 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>
- [16] 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>
- [17] 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>
- [18] 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>