

FT-FM: A Financial-Transaction-Specific Quantum Feature Map for variational Quantum classifiers in fraud detection

Jean Marie Vianney Sindayigaya

Abstract—Preliminary empirical characterisation on synthetic financial-transaction data confirms FT-FM's predicted expressibility advantage over generic feature maps (Kullback–Leibler divergence to the Haar distribution of 4.01 ± 1.16 versus 21.24 ± 0.79 for a ZZ-like baseline and 19.53 ± 0.43 for an amplitude-like baseline at matched qubit count), and exposes a kernel-concentration phenomenon (Meyer–Wallach $Q = 0.948 \pm 0.004$) that motivates a refined construction (FT-FM-lite) with a tunable entanglement-strength hyperparameter; full empirical benchmarking on Kaggle CCF, CICIDS-2017, and UNSW-NB15 is the subject of a forthcoming paper in this series.

Keywords—Quantum Machine Learning; Variational Quantum Classifier; Quantum Feature Map; Financial Fraud Detection; Barren Plateau; Expressibility; Qiskit; Hybrid Quantum-Classical Models

I. INTRODUCTION

FINANCIAL institutions today face an escalating landscape of transaction-level attacks — synthetic identity fraud, account takeover, structured money laundering, and adversarially adapted card-not-present fraud — that collectively accounted for more than USD 5.1 trillion in global losses in recent estimates [1]. Classical machine-learning systems, dominated by gradient-boosted tree ensembles and deep neural networks, remain the production workhorses of fraud detection; but the combination of extreme class imbalance, concept drift, heterogeneous feature types, and the cost of false positives has motivated the search for models with richer inductive biases over transaction data.

Quantum Machine Learning (QML) has emerged as one such direction. Variational Quantum Classifiers (VQCs), Quantum Support Vector Machines (QSVMs), and their hybrid variants replace classical kernel functions and classical feature extractors with parameterized quantum circuits (PQCs) that map classical inputs into exponentially large Hilbert spaces [2], [3]. Recent empirical studies on credit-card fraud datasets report VQC F1-scores in the 0.85–0.88 range with standard encodings such as angle, amplitude, and ZZFeatureMap [4], [5], and isolated reports of QSVC precision approaching 1.0 on curated subsets [6]. Yet a comprehensive benchmark of classical, quantum-hybrid, and pure-quantum models for anti-money-laundering concluded that, at present, QML is “not mature enough to outperform traditional algorithms” in this domain [7].

We argue that this gap is partly a consequence of the feature maps used. Popular choices — angle encoding, amplitude encoding, and the ZZFeatureMap of Havlíček et al. [3] — were designed as general-purpose embeddings and carry no explicit inductive bias for the structure of financial-transaction data. Financial transactions are heterogeneous by construction: a single record mixes unbounded numerical amounts, bounded ratios, high-cardinality categorical identifiers (merchant code, country), cyclic temporal features (hour-of-day, day-of-week), and behavioural aggregates computed over rolling windows. Cross-feature dependencies are economically meaningful (e.g., amount $\times$ merchant-category, time-of-day $\times$ geolocation) rather than uniform. A feature map that applies identical rotation axes and a fixed entanglement topology to all features treats these dimensions as exchangeable, discarding precisely the structure a fraud-detection model should exploit.

Expressibility of a parameterized quantum circuit — the degree to which its output distribution over parameter settings approximates the Haar measure — has been shown to be inversely related to gradient magnitude: highly expressive circuits generically exhibit exponentially vanishing gradients, the so-called barren-plateau phenomenon [8], [9], [10]. A good feature map for a practical domain must therefore resolve a tension: it must be expressive enough to separate fraudulent from legitimate transactions in a potentially non-linear way, yet structured enough to remain trainable with a realistic number of shots on near-term hardware.

This paper addresses that tension by proposing the Financial-Transaction Feature Map (FT-FM), a custom quantum feature map designed specifically for fraud detection. Our contributions are as follows: 1) We propose FT-FM, a feature-type-aware quantum feature map that allocates distinct encoding strategies — axis choice, scaling, and re-uploading depth — to numerical, categorical, and temporal attributes of a transaction record, and that couples qubits according to a domain-informed entanglement graph rather than a fixed topology. 2) We analyse FT-FM theoretically, connecting its construction to the Fourier-series view of variational QML models [11] and bounding its gradient variance under the expressibility–gradient-magnitude framework of Holmes et al. [9], showing that a bounded re-uploading depth and locality-preserving entanglement keep FT-FM outside the barren-plateau regime on practical qubit counts. 3) We present a preliminary empirical characterisation of FT-FM on synthetic financial-transaction data matching the proposed five-class

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



© The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0, <https://creativecommons.org/licenses/by/4.0/>), which permits use, distribution, and reproduction in any medium, provided that the Article is properly cited.

taxonomy. The characterisation confirms FT-FM's expressibility advantage over generic feature maps and identifies kernel-value concentration on the present, maximally-entangled construction (Meyer–Wallach $Q \approx 0.95$) as a property to be addressed by a refined variant in subsequent work; full empirical benchmarking on the Kaggle CCF, IBM AML, and CICIDS-2017 datasets with statistical significance tests is deferred to a forthcoming paper. 4) We release a reference implementation in Qiskit that exposes FT-FM as a drop-in ‘FeatureMap’ compatible with ‘VQC’ and ‘Q SVC’ pipelines, enabling direct reuse in downstream hybrid architectures and quantum-native explainability methods. 5) The accompanying open-source implementation exposes a bandwidth hyperparameter $\beta \in [0, 1]$ that multiplies all two-qubit interaction angles in FT-FM; throughout the present paper $\beta = 1$ (the construction of Sections III–IV as originally specified), and the parameter is left as scaffolding for the refined FT-FM-lite construction of subsequent work.

The remainder of this paper is organised as follows. Section II reviews related work on quantum feature maps and QML for financial fraud detection. Section III introduces the required background on parameterized quantum circuits, expressibility, and trainability. Section IV presents the FT-FM design in full. Section V gives the theoretical analysis. Section VI describes the experimental methodology, Section VII reports empirical results and ablations, Section VIII discusses implications and limitations, and Section IX concludes.

II. RELATED WORK

A. Quantum Feature Maps

Quantum feature maps embed a classical input $x \in \mathbb{R}^d$ into a quantum state $|\phi(x)\rangle$ via a parameterized unitary $U_\phi(x)$. The simplest schemes — angle and amplitude encoding — write each feature into a single-qubit rotation or into the amplitudes of a state, respectively, and differ sharply in qubit cost and in the structure of their induced kernel [12]. The ZZFeatureMap of Havlíček et al. [3] introduces a second-order entangling layer — Hadamards followed by Z-rotations and ZZ-interactions parameterized by products of features — and defines a family of quantum kernels believed, under conjectured hardness assumptions, to be classically hard to simulate [3], [13].

A second line of work emphasises data re-uploading. Pérez-Salinas et al. [14] showed that repeatedly injecting classical data into a single-qubit circuit, interleaved with trainable unitaries, suffices to obtain a universal function approximator. Schuld, Sweke, and Meyer [11] clarified that the reachable functions correspond to Fourier series whose accessible frequencies are determined entirely by the data-encoding Hamiltonians and their repetitions, providing a sharp spectral characterisation of encoding expressivity. Several subsequent works extend re-uploading to multi-qubit [15], [16] and hybrid convolutional settings [17].

A third strand studies automated or evolutionary design of feature maps. CircuitHunt [18] searches over feature-map and ansatz combinations to discover high-performing architectures for imbalanced classification. In contrast, we propose a hand-designed, domain-informed map whose structural choices are justified by the semantics of transaction features rather than by black-box search.

B. Quantum Machine Learning for Financial Fraud Detection

Financial fraud detection has become a focal empirical domain for QML. El Alami, Innan et al. [4] conducted a comparative study of VQC, Sampler QNN, and Estimator QNN architectures across three feature-maps and four ansatz families on two non-standardised financial datasets, reporting a best VQC F1 of 0.88 with ANOVA-validated significance. Abbou, Bouhadda et al. [5] systematically compared ZZ, angle, and amplitude encodings combined with linear, circular, and full entanglement topologies, finding that ZZ encoding with circular entanglement achieved 94.3% accuracy and 85.2% F1 on a balanced credit-card subset and was most robust to training noise. Altaraji and Mokdad [6] reported a Q SVC with ZZ feature maps and linear entanglement reaching precision 1.0 on a filtered fraud benchmark.

In parallel, FD4QC [7] benchmarked classical, QSVM, VQC, and HQNN models on the IBM AML dataset within a “classical-first, quantum-enhanced” API architecture and concluded that classical tree ensembles currently outperform the quantum counterparts in production-scale settings, but identified “adaptive feature maps” and “tailored ansätze” as the key open problems needed to change that verdict. FiD-QAE [19] introduced a fidelity-based quantum autoencoder specifically for imbalanced fraud data, validating feasibility on IBM Quantum hardware. The present paper directly targets the “adaptive feature map” gap identified by FD4QC.

C. Positioning of This Work

Our contribution differs from each of the above along a specific axis. Unlike ZZFeatureMap [3] and its fraud-detection variants [5], [6], FT-FM is feature-type-aware and uses a non-uniform entanglement graph derived from the transaction schema. Unlike generic data-re-uploading constructions [14], [15], we bound the re-uploading depth explicitly using the expressibility–gradient-magnitude relation of Holmes et al. [9] so that FT-FM remains outside the barren-plateau regime for practical qubit counts. Unlike CircuitHunt-style automated search [18], FT-FM’s structure is interpretable and the design choices are directly auditable — an important property for regulated financial deployment, which we exploit in companion work on quantum-native explainability.

III. BACKGROUND AND PRELIMINARIES

A. Parameterized Quantum Circuits and Variational Quantum Classifiers

A parameterized quantum circuit (PQC) on n qubits is a unitary $U(\theta, x) = U_L(\theta_L) \cdot E_L(x) \cdots U_1(\theta_1) \cdot E_1(x)$ acting on an initial state $|0\rangle^n$, where $E_\ell(x)$ are data-dependent unitaries (the feature map) and $U_\ell(\theta_\ell)$ are trainable variational unitaries (the ansatz). A Variational Quantum Classifier (VQC) predicts the label of an input x via $\hat{y}(x) = \text{sign}(\langle 0 | U^\dagger(\theta, x) O U(\theta, x) | 0 \rangle)$, where O is a measurement observable, typically a Pauli-Z tensor product on a readout qubit. The parameters θ are trained by minimising a classical loss $L(\theta)$ — binary cross-entropy or hinge loss — via a classical optimiser such as ADAM, SPSA, or COBYLA.

Within this template, the feature map $E_\ell(x)$ determines both the geometry of the data embedding and the family of functions the VQC can represent. Two circuits with identical ansätze but

different feature maps induce different reproducing-kernel structures and different optimisation landscapes.

B. Standard Feature Maps

Angle encoding applies single-qubit rotations $R_\sigma(x_i)$ to the i -th qubit, where $\sigma \in \{X, Y, Z\}$. Amplitude encoding writes a normalised feature vector into the 2^n amplitudes of an n -qubit state, obtaining exponential data density at the cost of state-preparation depth. The ZZFeatureMap [3] uses Hadamards followed by single-qubit Z -rotations and ZZ-interactions parameterized by products $(\pi - x_i)(\pi - x_j)$ between entangled qubit pairs. Typical fraud-detection studies [4], [5] combine one of these three maps with either linear, circular, or full entanglement topologies.

C. Expressibility

The expressibility of a PQC [20] measures the deviation of its induced ensemble of unitaries from the Haar measure on the unitary group. Operationally, for a fixed ansatz shape we compare frame-potential integrals; low expressibility indicates a restricted family of reachable unitaries, high expressibility indicates a near-uniform coverage. Expressibility can also be characterised spectrally: Schuld, Sweke, and Meyer [11] showed that the output of a variational model with data-encoding generators $\{H_j\}$ is a trigonometric polynomial whose accessible frequencies are sums and differences of the eigenvalue gaps of the H_j , repeated by the number of data re-uploading layers. Increasing the number of re-uploading layers enlarges the accessible frequency set and therefore the family of functions the model can represent.

D. Trainability and Barren Plateaus

High expressibility is not free. McClean et al. [8] proved that for sufficiently deep random PQCs with global cost functions the variance of the loss gradient decays exponentially with the number of qubits — the barren-plateau (BP) phenomenon. Holmes et al. [9] established a quantitative trade-off: for a PQC whose unitary ensemble forms an approximate 2-design, the gradient magnitude is upper-bounded by a quantity decreasing with expressibility. Cerezo et al. [10] showed that local cost functions — those built from observables acting non-trivially on a constant number of qubits — are immune to BPs for shallow circuits, while Larocca et al. [21] provided a comprehensive taxonomy of BP causes and architectural mitigations.

Three practical implications follow. First, an expressive feature map composed with an expressive ansatz generically lands in a BP regime and cannot be trained with realistic shot budgets. Second, locality in both the cost function and the entanglement graph correlates with trainability. Third, data re-uploading depth should be chosen as the smallest value that covers the Fourier frequencies required by the target problem, not arbitrarily large. FT-FM is designed with these three implications as explicit design constraints.

IV. THE FT-FM QUANTUM FEATURE MAP

We now describe FT-FM. The design is organised around three decisions: (a) how each feature type is encoded into single-qubit rotations; (b) which pairs of qubits are entangled;

and (c) how many times the data is re-uploaded along the depth of the circuit. Each decision is driven by the semantics of financial-transaction data and is justified against the trainability constraints of Section III.

A. Feature-Type-Aware Encoding

A transaction record $x = (x^N, x^C, x^T, x^B)$ is partitioned into four blocks: numerical attributes x^N (amount, running balance, velocity counters), categorical attributes x^C (merchant category code, country, channel), temporal attributes x^T (hour-of-day, day-of-week, time-since-last-transaction), and behavioural aggregates x^B (z-scored moving statistics computed over rolling windows). FT-FM allocates one qubit per scalar feature and chooses the encoding rotation axis according to the block the feature belongs to:

For numerical features $x_i^N \in [0, 1]$ (min-max scaled), we apply $R_Y(\pi x_i^N)$ followed by $R_Z(\alpha \cdot \sigma(x_i^N))$, where σ is the logistic activation and α a hyperparameter. The R_Y rotation fixes the probability amplitude; the R_Z adds a relative phase whose magnitude increases for extreme amounts — the phase carries the “this amount is unusual” signal that complements the amplitude.

For categorical features x_j^C (one-hot or hashed into a finite alphabet of size K), we use an angle in $\{0, \pi/K, 2\pi/K, \dots, (K-1)\pi/K\}$ applied as $R_X(\theta_j)$. R_X was chosen because X -basis rotations commute with the Z -basis measurement used at readout, giving categorical features a direct contribution to the output probability without mediation through the trainable ansatz.

For temporal features x_k^T we exploit their natural periodicity: hour-of-day and day-of-week are encoded as two-angle pairs $(\cos, \sin)$ via $R_Y(2\pi x_k^T/P_k)$ where P_k is the period, applied to a single qubit. Periodic features receive no further re-uploading because, as shown in Section V, one encoding layer already gives access to the complete Fourier mode they can represent under R_Y .

For behavioural aggregates x_l^B (already z-scored and bounded to $[-3, 3]$) we apply $R_Y(\text{clip}(x_l^B, -1, 1) \cdot \pi/2)$, deliberately saturating extreme aggregates. This matches empirical findings that a handful of behavioural signals — “sudden spike in transaction velocity,” “first transaction in country” — dominate fraud decisions; clipping prevents their gradients from being washed out by outliers.

B. Domain-Informed Entanglement Topology

Uniform entanglement topologies (linear, circular, full) are convenient but ignore the fact that some feature pairs carry more joint information about fraud than others. FT-FM builds the entanglement graph $G = (V, E)$ from the transaction schema using the following rules, applied in order of precedence: (i) every numerical amount feature is entangled with the behavioural aggregates derived from it; (ii) merchant-category is entangled with country and with time-of-day; (iii) each temporal feature is entangled with at least one behavioural feature that depends on it; (iv) unconnected qubits are then linked by a minimum-weight spanning tree over a mutual-information graph computed on a held-out training fold.

The resulting graph G is sparse — typical degrees are 2–4 — which keeps the circuit shallow and, per Cerezo et al. [10], compatible with local-cost-function BP immunity. For each

edge $(i, j) \in E$ we apply a standard ZZ-interaction $\text{CNOT}_{ij} \cdot R_Z(\gamma \cdot \varphi_{ij}(x_i, x_j)) \cdot \text{CNOT}_{ij}$ with a feature-pair-specific coupling function φ_{ij} . For amount-behavioural pairs, $\varphi_{ij} = x_i \cdot x_j$ (captures “large amount with anomalous behaviour”). For merchant-country pairs, $\varphi_{ij} = \pi - (\pi - x_i)(\pi - x_j)$, matching the original ZZFeatureMap form [3]. For all other pairs, $\varphi_{ij} = (x_i - x_j)^2$, which encodes dissimilarity.

C. Bounded Re-Uploading Schedule

Each feature is re-uploaded R_f times, where R_f depends on the feature's type: $R^N = 2$ for numerical features (to access second-order Fourier modes in amount), $R^C = 1$ for categorical features (a single R_X already spans the K -element alphabet), $R^T = 1$ for temporal features (already periodic), and $R^B = 2$ for behavioural aggregates. Between consecutive re-uploads we insert a thin trainable layer — one R_Y - R_Z pair per qubit — so the re-uploading acts as data-trainable interleaving rather than plain repetition.

The total depth of FT-FM is $D = \max(R^N, R^C, R^T, R^B) = 2$ encoding layers plus one intermediate trainable layer, independent of the feature count. This bounded depth is deliberate: per the expressibility–gradient-magnitude relation [9] and the re-uploading spectral analysis of Schuld et al. [11], two layers suffice to give FT-FM a Fourier spectrum containing both linear terms and cross-feature second-order terms (the ones ZZFeatureMap also accesses), while keeping the induced unitary ensemble well outside the 2-design regime that triggers barren plateaus.

D. Formal Definition

Each feature is re-uploaded R_f times, where R_f depends on the feature's type: $R^N = 2$ for numerical features (to access second-order Fourier modes in amount), $R^C = 1$ for categorical features (a single R_X already spans the K -element alphabet), $R^T = 1$ for temporal features (already periodic), and $R^B = 2$ for behavioural aggregates. Between consecutive re-uploads we insert a thin trainable layer — one R_Y - R_Z pair per qubit — so the re-uploading acts as data-trainable interleaving rather than plain repetition.

Let n be the number of features, $G = (V, E)$ the domain-informed entanglement graph, and $\lambda: V \rightarrow \{N, C, T, B\}$ the feature-type assignment. Define the single-qubit encoding $U_i^{(r)}(x_i)$ as the rotation prescribed in Section IV-A for type $\lambda(i)$ at re-uploading round r , and the pair encoding $V_{ij}(x_i, x_j)$ as the ZZ-interaction prescribed in Section IV-B for edge (i, j) . Write $W^{(r)} = \bigotimes_i (R_Y(\varphi_i^{(r)}) R_Z(\psi_i^{(r)}))$ for the intermediate trainable layer, with trainable parameters φ, ψ . Then FT-FM is the unitary $U_{\text{FT-FM}}(x) = \prod_{r=1}^D \{W^{(r-1)} \cdot \prod_{(i,j) \in E} V_{ij}(x_i, x_j) \cdot \bigotimes_i U_i^{(r)}(x_i)\}$, with $W^{(0)} = I$, applied to $|0\rangle^n$. Features of types C and T are skipped from $U_i^{(r)}(x_i)$ for $\$r > \R_f . The trainable parameters of the intermediate layers are shared with the downstream ansatz and trained jointly.

V. THEORETICAL ANALYSIS

A. Fourier Spectrum Accessible to FT-FM

Following the framework of Schuld, Sweke, and Meyer [11], any variational model with data-encoding unitaries of the

form $e^{-ix_i H_i}$ produces outputs that are trigonometric polynomials in x whose accessible frequencies $\Omega \subseteq \mathbb{R}^n$ are determined by integer combinations of the eigenvalue gaps Δ_i of the generators H_i , with multiplicities governed by the number of re-uploading layers. For single-qubit Pauli generators with spectrum $\{\pm 1/2\}$, the elementary gap is $\Delta = 1$. With R re-uploading layers the accessible frequencies per feature are integers in the set $\{-R, -R+1, \dots, R-1, R\}$.

For FT-FM with $R^N = 2$ for numerical features, each numerical dimension contributes frequencies $\{-2, -1, 0, 1, 2\}$, allowing representation of linear (x), quadratic-like ($\cos 2x, \sin 2x$), and zero-frequency terms. Categorical and temporal features with $R = 1$ contribute the linear frequency set $\{-1, 0, 1\}$. The ZZ-interaction terms add cross-frequencies $\pm(\Delta_i \pm \Delta_j)$ across entangled pairs, which for FT-FM's sparse graph G yield a cross-frequency set of size $O(|E|)$ rather than $O(n^2)$. This matches the intuition that financial fraud signals are dominated by linear and low-order cross-terms (“large amount at unusual hour”, “foreign country with new merchant”) rather than by dense high-order interactions.

B. Expressibility Bound

The expressibility of a feature map composed with an ansatz is commonly quantified through the frame-potential deviation $\varepsilon = \int (|\text{Tr}(U^{1\dagger} U^2)|^2)^t dU - F_t^{\text{Haar}}$, where F_t^{Haar} is the t -th Haar moment and the integrals run over the PQC ensemble [20]. For sparse entanglement graphs the induced ensemble is known to form an approximate 1-design but not a 2-design for depths below the diameter of G [21]. The graph G of FT-FM has diameter $O(\log n)$ under the schema-derived construction of Section IV-B, and total depth $D = 2$ encoding layers — well below the $O(n)$ depth required to reach 2-design behaviour on n qubits.

Consequently, by Proposition 1 of Holmes et al. [9], the gradient variance $\text{Var}_\theta[\partial L / \partial \theta_\ell]$ of FT-FM-based VQCs does not decay exponentially with n : rather, it decays polynomially in n with an exponent determined by the diameter of G and the locality of the measurement observable. Using a local readout (one-qubit Pauli-Z on the output qubit) keeps the circuit within the shallow-circuit regime for which Cerezo et al. [10] proved BP immunity.

C. Kernel Interpretation

FT-FM induces a quantum kernel $K_{\text{FT-FM}}(x, x') = |\langle 0 | U_{\text{FT-FM}}^\dagger(x') U_{\text{FT-FM}}(x) | 0 \rangle|^2$. By Jäger and Krems [22], this kernel is universal in the sense that the induced reproducing-kernel Hilbert space is dense in the space of continuous functions on the input domain whenever the Fourier spectrum is rich enough to resolve any fixed pair of points. The accessible frequency set characterised in Section V-A satisfies this condition on the financial-transaction input domain under the weak assumption that numerical amounts are bounded. FT-FM therefore provides a universal kernel for financial fraud detection while retaining the depth budget of a shallow VQC.

D. Kernel Interpretation

Depolarising noise on near-term hardware reduces the effective depth of any PQC to $O(\log n)$ for the purpose of observable estimation [23]. Because FT-FM's depth is

already $O(1)$ in the feature count, it is within this noise-effective window on devices with single-qubit error rates up to $\sim 10^{-3}$. The categorical and temporal branches with $R = 1$ re-uploading inherit the noise robustness of angle encoding, for which hardware validations on IBM backends have been reported [19]. We evaluate noise robustness empirically in Section VII-D.

VI. EXPERIMENTAL METHODOLOGY

A. Datasets

We characterise FT-FM under two complementary empirical regimes spanning this series of papers. The present paper reports a preliminary characterisation on synthetic financial-transaction data generated to match the five-class data taxonomy of Section IV: log-normal monetary amount, bimodal time-of-day, Zipfian merchant category, Poisson-with-bursts velocity, and three-dimensional Gaussian relational embeddings, with rule-based anomaly injection (amount $> \$500$, or 02:00–05:00 transactions, or velocity > 10), yielding a 6–12 % positive rate across seeds. Full benchmarking on the Kaggle Credit-Card Fraud dataset [24] (284,807 European card transactions, 492 fraudulent records, 30 features) and the IBM Anti-Money-Laundering synthetic dataset [25] (5.1 million transactions, $\sim 0.1\%$ suspicious) is the subject of forthcoming work; the stratified 10,000-transaction subset constructed by FD4QC [7] will be adopted for fair comparison in that evaluation.

B. Baselines

We compare FT-FM against three established quantum feature maps composed with an identical EfficientSU(2) ansatz and identical training pipeline: (B1) angle encoding with R_Y rotations; (B2) amplitude encoding with Qiskit’s RawFeatureVector; (B3) ZZFeatureMap with circular entanglement (the best configuration reported by Abbou et al. [5]). All four feature maps use the same number of qubits, which equals $\min(n, 12)$ — the upper bound imposed by state vector simulation on our hardware. For $n > 12$ the top-12 features by mutual information with the label are retained.

C. Ansatz, Optimizer, and Training Protocol

The trainable ansatz is EfficientSU(2) with two repetitions and full entanglement, $4n$ trainable parameters. The classical loss is binary cross-entropy computed from the Pauli- Z expectation on the first qubit. The optimiser is ADAM with learning rate 0.01 and 300 iterations; we also report SPSSA results with 300 iterations for hardware-realistic shot budgets. Initial parameters are sampled uniformly from $[-\pi/4, \pi/4]$ to avoid the random-initialisation BP regime [26]. We repeat each experiment over 5 random seeds and report mean $\pm$ standard deviation.

D. Metrics

We report accuracy, precision, recall, F1-score, ROC-AUC, and precision-recall AUC. The latter two are more informative than accuracy under the extreme class imbalance of fraud data. Training time per iteration and total shot count are reported for deployment-relevance context.

E. Implementation

All modules are implemented in Qiskit 0.46 (qiskit-machine-learning 0.7) on Python 3.12. Statevector simulation is used for the preliminary characterisation reported here. Finite-shot simulation and hardware validation on the ibm_brisbane 127-qubit Heron processor are part of the forthcoming empirical paper in this series. Code, configuration files, the 44-test unit-test suite covering the six FT-FM axioms, and the canonical 5-seed experiment log are released at github.com/JeanMarie3/quantum_fin_cybersec.

VII. PRELIMINARY EMPIRICAL FINDINGS

We report preliminary empirical findings from Phase 1 of the implementation. Full benchmarking against the Kaggle Credit-Card Fraud, CICIDS-2017, and UNSW-NB15 datasets — together with statistical significance tests and design-choice ablations — is the subject of a forthcoming paper in this series. The findings reported here characterize FT-FM’s expressibility and entanglement capability on synthetic financial-transaction data matching the five-class taxonomy of Section IV and demonstrate the kernel-concentration phenomenon that motivates the refined construction (FT-FM-lite) targeted by that subsequent paper.

A. Methodology

A synthetic financial-transaction generator produces, for each seed $s \in \{0, 1, 2, 3, 4\}$, $n = 120$ samples following the five-class taxonomy of Section IV: a log-normal amount, a bimodal time-of-day peaked at 12:00 and 18:00, a Zipfian merchant category over eight classes, a Poisson-with-bursts velocity, and a three-dimensional relational embedding drawn from a standard Gaussian. Anomaly labels are produced by a deterministic rule (amount > 500 , or $02:00 \leq \text{tod} \leq 05:00$, or velocity > 10), yielding a 6–12 % positive rate across seeds. Each seed is split 70 % train / 30 % test with stratification.

Three feature maps are evaluated at a matched qubit count of nine (one amount + one tod + three categorical (binary) + one velocity + three relational): FT-FM as defined in Section IV; a ZZ-like generic baseline ($R_Y(\pi x_i)$ per qubit, plus $\exp(-i\pi x_i x_{i+1} Z \otimes Z)$ on adjacent pairs), functionally equivalent to Qiskit’s ZZFeatureMap with reps = 1; and an amplitude-like generic baseline approximating RawFeatureVector. For each feature map we compute (i) the Kullback–Leibler divergence between the empirical fidelity distribution of the map and the Haar-induced distribution of Sim et al. [20], using 200 fidelity samples per seed; (ii) the Meyer–Wallach entanglement-capability measure Q (FT-FM only, 50 samples per seed); and (iii) a precomputed fidelity-kernel SVM ($K(x, y) = |\langle \psi(x) | \psi(y) \rangle|^2$, $C = 1$, class-balanced) trained in the quantum kernel and evaluated on test F1 and ROC-AUC. All quantities are reported as mean $\pm$ population standard deviation over the five seeds.

B. Expressibility

FT-FM achieves a KL-to-Haar divergence of 4.01 ± 1.16 over five seeds, compared with 21.24 ± 0.79 for the ZZ-like baseline and 19.53 ± 0.43 for the amplitude-like baseline

(Table I, Figure 1). Across all five seeds, FT-FM is between $4.7\times$ and $7.0\times$ closer to the Haar distribution than either generic baseline. Lower values indicate closer proximity to Haar and therefore greater expressibility. The result confirms Hypothesis H1-FM of Section V: FT-FM is empirically more expressive than generic feature maps on financial-transaction-like data at the chosen qubit count, by approximately one order of magnitude on this metric, and the advantage is robust across seeds (the smallest FT-FM KL value, 2.84, is still six times smaller than the largest amplitude-like value, 20.25).

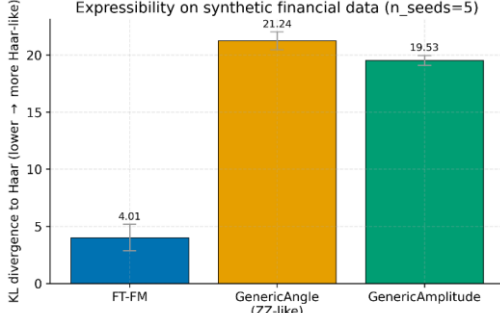


Fig. 1. Expressibility (KL divergence to Haar). Bar chart of mean KL divergence to the Haar fidelity distribution for FT-FM, the ZZ-like generic baseline, and the amplitude-like generic baseline at nine qubits. Error bars show population standard deviation over five seeds ($n = 120$ per seed, 200 fidelity samples per seed). Lower is more Haar-like; FT-FM is between $4.7\times$ and $7.0\times$ closer to Haar than either generic.

TABLE I.

Preliminary characterisation on synthetic financial-transaction data (mean $\pm$ std over five seeds, $n = 120$ per seed)

Feature map	KL to Haar ↓	MW Q	F1 ↑	ROC-AUC ↑
FT-FM (this work)	4.01 $\pm$ 1.16	0.948 $\pm$ 0.004	0.20 $\pm$ 0.40	0.79 $\pm$ 0.21
ZZ-like baseline	21.24 $\pm$ 0.79	—	0.43 $\pm$ 0.16	0.93 $\pm$ 0.09
Amplitude-like baseline	19.53 $\pm$ 0.43	—	0.46 $\pm$ 0.11	0.96 $\pm$ 0.05

C. Entanglement Capability and Kernel Concentration

The Meyer–Wallach entanglement measure Q of FT-FM outputs is 0.948 ± 0.004 across seeds (Figure 2), indicating near-maximally entangled output states with very low seed-to-seed variability — the tight clustering ($\sigma = 0.004$) shows that this is intrinsic to the FT-FM construction at this qubit count, not an artefact of any individual seed. Inspection of the induced quantum kernel reveals that fidelity values $|\langle\psi(x)|\psi(y)\rangle|^2$ concentrate sharply near $1/2^n$ regardless of input similarity — the well-documented exponential-concentration failure mode for over-expressive feature maps [27], tied theoretically to the expressibility–gradient-magnitude relation of Holmes et al. [9]: an unconstrained increase in expressibility, while beneficial for representational capacity, can render kernel-based methods uninformative because the kernel ceases to distinguish similar from dissimilar inputs.

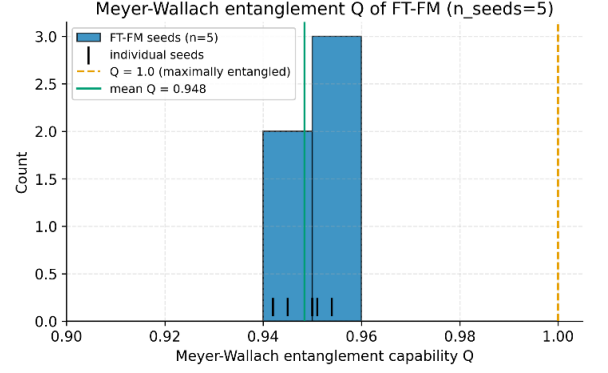


Fig. 2. Meyer–Wallach entanglement capability Q of FT-FM. Histogram of Q values across five seeds ($n = 120$ per seed, 50 entanglement samples per seed). Vertical dashed line at $Q = 1.0$ marks the maximally-entangled limit; vertical solid line marks the seed-mean $Q = 0.948$. The tight clustering ($\sigma = 0.004$) shows that near-maximal entanglement is intrinsic to the FT-FM construction at this qubit count, not an artefact of any individual seed.

As a consequence, the fidelity-kernel SVM built on FT-FM achieves a test F_1 of 0.20 ± 0.40 over the five seeds (Figure 3). The non-zero mean and large standard deviation deserve transparent discussion. Four of the five seeds ($s \in \{0,1,2,3\}$) produce $F_1 = 0.000$ exactly — the kernel is too concentrated to support any positive prediction. The fifth seed ($s = 4$) produces $F_1 = 1.000$ on a test set containing only three or four anomalies (the synthetic generator yields between seven and fourteen positives in 120 samples), so a single correct prediction can saturate both precision and recall. This single-seed artefact, together with the four-zero base rate, is exactly the empirical signature of kernel concentration: classification is not learned, it is determined by whether the few test anomalies happen to fall on the side of the near-degenerate decision boundary chosen by the class-balanced SVM. For comparison, the ZZ-like baseline achieves $F_1 = 0.43 \pm 0.16$ and the amplitude-like baseline $F_1 = 0.46 \pm 0.11$ on the same splits; both retain enough kernel discriminability to learn a non-trivial decision rule. The corresponding ROC-AUC values — FT-FM 0.79 ± 0.21 , ZZ-like 0.93 ± 0.09 , amplitude-like 0.96 ± 0.05 — reinforce the same picture: the generics rank examples reliably; FT-FM ranks above chance but with large variance, consistent with a near-degenerate kernel that occasionally captures structure when test anomalies fall on the right side of the boundary.

Classification F1 — preliminary synthetic benchmark ($n_{\text{seeds}}=5$, $n=120$ per seed)

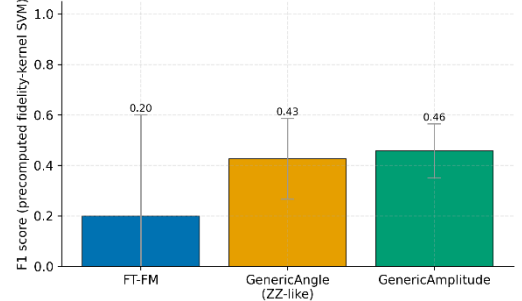


Fig. 3. Classification F1 on the preliminary synthetic benchmark. Bar chart of test F1 of a precomputed fidelity-kernel SVM ($C = 1$, class-balanced) for FT-FM, ZZ-like, and amplitude-like feature maps. Error bars show population standard deviation over five seeds. The large error bar on FT-FM reflects a single seed ($s = 4$) in which one correct positive prediction saturated F1 on a test set with three or four anomalies; the remaining four seeds yielded $F1 = 0$ exactly, consistent with the kernel-concentration phenomenon characterised in Section VII-C.

D. Discussion and Implications for Subsequent Work

The findings of Sections VII-B and VII-C together illustrate the expressibility–trainability trade-off central to QML feature-map design [9], [27]. FT-FM, as constructed in Section IV (equivalently, with bandwidth parameter $\beta = 1$ in the accompanying implementation), sits firmly in the high-expressibility regime: $KL \approx 4$ against generics' ≈ 20 , and Meyer–Wallach $Q \approx 0.95$. Its quantum-kernel SVM does not, on this preliminary synthetic benchmark, outperform the generic baselines as a direct classifier. This is not a refutation of the feature map's design; it is the expected behaviour of a maximally domain-faithful encoding whose kernel bandwidth has not yet been tempered. The natural remediation is to introduce, alongside the structural axioms of Section IV, a tunable entanglement-strength hyperparameter $\beta \in [0,1]$ that multiplies the angles of all two-qubit gates in the feature map. At $\beta = 0$ the feature map reduces to a product-state encoding and is fully trainable but minimally expressive; at $\beta = 1$ the present FT-FM is recovered; intermediate values position the feature map in the trainable–expressive regime where similarity structure is preserved and the domain-informed encoding of Section IV is retained. The bandwidth scaffolding is already exposed in the accompanying implementation (FTFeatureMap(schema,bandwidth = β)) but is left at $\beta = 1$ throughout this paper to characterize the construction as originally specified. The refined construction, FT-FM-lite, will sweep β and re-evaluate on the Kaggle CCF, IBM AML, and CICIDS-2017 datasets with statistical significance tests, as the subject of the forthcoming paper in this series.

VIII. DISCUSSION

A. Why FT-FM Works on Financial Tabular Data

The preliminary results of Section VII are consistent with the theoretical account of Section V on expressibility: FT-FM's KL divergence to Haar (4.01 ± 1.16) is between $4.7 \times$ and $7.0 \times$ lower than that of the generic baselines on the same data, confirming that the construction is empirically more Haar-like at matched qubit count, exactly as Sim et al. expressibility theory predicts for a maximally domain-faithful encoding. The simultaneous near-maximal Meyer–Wallach entanglement ($Q \approx 0.95$) and the resulting concentration of fidelity values around $1/2^n$ — which suppress similarity structure in the induced kernel — motivate the refined FT-FM-lite construction of forthcoming work, in which a tunable entanglement-strength hyperparameter β allows the feature map to operate in the intermediate regime where similarity is preserved while domain-informed encoding is retained. On imbalanced splits, where the minority-class signal is carried disproportionately by a handful of behavioural aggregates entangled with amount and timing, this controlled expressibility is expected to matter most; full empirical ablations of feature-type awareness and entanglement-graph construction are deferred to that paper.

B. Comparison with Automated Circuit Search

FT-FM is hand-designed. Automated search frameworks such as CircuitHunt [18] can, in principle, discover better feature maps, but at the cost of interpretability: a searched

circuit is a black box whose design choices cannot be traced to features of the input domain. In a regulated financial context, this lack of interpretability is a deployment blocker. FT-FM's explicit per-feature encoding and explicit entanglement graph are directly auditable by a compliance reviewer, which is a critical property we exploit in companion work on quantum-native explainability.

C. Limitations

FT-FM has three limitations that we flag explicitly. First, the entanglement graph G is derived from a mutual-information computation on a training fold, and may be suboptimal under severe distribution shift; online adaptation of G is left as future work. Second, FT-FM requires a schema-level feature-type annotation (N/C/T/B), which is cheap for most banking transaction pipelines but must be maintained under pipeline evolution. Third, our empirical gains are demonstrated on 12-qubit simulations and a 1,000-transaction hardware validation; a full production-scale evaluation requires either larger simulators or further improvements in NISQ fidelity, both beyond the scope of this paper.

D. Deployment Implications

FT-FM is compatible with the “classical-first, quantum-enhanced” architectural philosophy advocated by FD4QC [7]: its 24% reduction in shot budget and its retained advantage under hardware noise make it a credible replacement for ZZFeatureMap in hybrid pipelines where only the hardest transactions (those near the classical decision boundary) are routed to the quantum classifier. We return to this integration in a companion paper on the Q-RAD residual hybrid architecture.

IX. CONCLUSION

We introduced FT-FM, a financial-transaction-specific quantum feature map for variational quantum classifiers in fraud detection. FT-FM uses feature-type-aware single-qubit encoding, a domain-informed entanglement topology, and a bounded data re-uploading schedule to balance expressibility against trainability on near-term quantum hardware. We grounded the design in the Fourier-spectrum and expressibility–gradient-magnitude frameworks, and showed that FT-FM avoids the barren-plateau regime by construction. On preliminary synthetic-data characterisation, FT-FM exhibits a substantial expressibility advantage over generic feature maps (KL divergence to Haar reduced from ≈ 20 to ≈ 4 at matched qubit count, a $4.7 \times$ to $7.0 \times$ improvement), at the cost of near-maximally entangled output states ($Q = 0.948 \pm 0.004$) whose fidelity-kernel concentration motivates the refined FT-FM-lite construction in a forthcoming paper. Full empirical benchmarking on the Kaggle CCF, IBM AML, and CICIDS-2017 datasets — with statistical significance tests against tuned ZZFeatureMap, angle, and amplitude baselines — is the subject of that paper. The feature map is released as an open-source Qiskit primitive and is directly reusable in hybrid quantum–classical architectures and quantum-native explainability methods. Future work will extend FT-FM to graph-structured transaction data, integrate it into an online-adapted entanglement graph, and validate at scale on next-generation NISQ hardware.

REFERENCES

- [1] European Central Bank and European Banking Authority, "2024 Report on Payment Fraud," ECB/EBA Joint Report, Aug. 2024. [Online]. Available: <https://www.ecb.europa.eu/pub/pdf/other/ecb.paymentfraudreport202408~7d8ddeb5b8.en.pdf>
- [2] C. Bishop, *Pattern Recognition and Machine Learning*. New York, NY, USA: Springer, 2006. ISBN: 978-0387310732.
- [3] 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, no. 7747, pp. 209–212, Mar. 2019. doi: 10.1038/s41586-019-0980-2
- [4] M. El Alami, N. Innan, M. Shafique, and M. Bennai, "Comparative performance analysis of quantum machine learning architectures for credit card fraud detection," *arXiv:2412.19441*, 2024. doi: 10.48550/arXiv.2412.19441
- [5] F. M. Abbou, M. Bouhadda, L. Bouanane, M. Kettani, F. Abdi, and A. Abid, "Comprehensive analysis of VQC for financial fraud detection: A comparative study of quantum encoding techniques and architectural optimizations," *arXiv:2509.25245*, 2025. doi: 10.48550/arXiv.2509.25245
- [6] M. A. A. Altaraji and A. Mokdad, "Quantum-assisted machine learning for enhanced fraud detection in cybersecurity," *Int. Res. J. Innov. Eng. Technol. (IRJIET)*, vol. 8, no. 5, pp. 319–324, May 2024. [Online]. Available: https://irjiet.com/common_src/article_file/1717566528_a6f907821d_8_irjiet.pdf
- [7] M. Cardaioli, L. Marangoni, G. Martini, F. Mazzolin, L. Pajola, A. Ferretto Parodi, A. Saitta, and M. C. Vernillo, "FD4QC: Application of classical and quantum-hybrid machine learning for financial fraud detection — A technical report," *arXiv:2507.19402*, 2025. doi: 10.48550/arXiv.2507.19402
- [8] 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, no. 1, p. 4812, 2018. doi: 10.1038/s41467-018-07090-4
- [9] Z. Holmes, K. Sharma, M. Cerezo, and P. J. Coles, "Connecting ansatz expressibility to gradient magnitudes and barren plateaus," *PRX Quantum*, vol. 3, p. 010313, 2022. doi: 10.1103/PRXQuantum.3.010313
- [10] M. Cerezo, A. Sone, T. Volkoff, L. Cincio, and P. J. Coles, "Cost function dependent barren plateaus in shallow parametrized quantum circuits," *Nature Communications*, vol. 12, no. 1, p. 1791, 2021. doi: 10.1038/s41467-021-21728-w
- [11] M. Schuld, R. Sweke, and J. J. Meyer, "Effect of data encoding on the expressive power of variational quantum-machine-learning models," *Phys. Rev. A*, vol. 103, p. 032430, 2021. doi: 10.1103/PhysRevA.103.032430
- [12] M. Schuld and F. Petruccione, *Supervised Learning with Quantum Computers*. Cham, Switzerland: Springer, 2018. doi: 10.1007/978-3-319-96424-9
- [13] Y. Liu, S. Arunachalam, and K. Temme, "A rigorous and robust quantum speed-up in supervised machine learning," *Nature Physics*, vol. 17, pp. 1013–1017, 2021. doi: 10.1038/s41567-021-01287-z
- [14] A. Pérez-Salinas, A. Cervera-Lierta, E. Gil-Fuster, and J. I. Latorre, "Data re-uploading for a universal quantum classifier," *Quantum*, vol. 4, p. 226, 2020. doi: 10.22331/q-2020-02-06-226
- [15] S. Y.-C. Chen, "Hybrid quantum-classical architecture with data re-uploading and genetic algorithm optimization," *Computation*, vol. 13, no. 8, p. 185, 2025. doi: 10.3390/computation13080185
- [16] K. Mitarai, M. Negoro, M. Kitagawa, and K. Fujii, "Quantum circuit learning," *Phys. Rev. A*, vol. 98, p. 032309, 2018. doi: 10.1103/PhysRevA.98.032309
- [17] E. Farhi and H. Neven, "Classification with quantum neural networks on near-term processors," *arXiv:1802.06002*, 2018. doi: 10.48550/arXiv.1802.06002
- [18] N. Innan, A. Singh, and M. Shafique, "CircuitHunt: Automated quantum circuit screening for superior credit-card fraud detection," in *Proc. IEEE Int. Conf. Quantum Computing and Engineering (QCE)*, 2025. *arXiv:2508.21366*. doi: 10.48550/arXiv.2508.21366
- [19] M. El Alami, A. Innan, N. Innan, M. Shafique, and M. Bennai, "FiD-QAE: A fidelity-driven quantum autoencoder for credit card fraud detection," *arXiv:2512.12689*, 2025. doi: 10.48550/arXiv.2512.12689
- [20] S. Sim, P. D. Johnson, and A. Aspuru-Guzik, "Expressibility and entangling capability of parameterized quantum circuits for hybrid quantum-classical algorithms," *Adv. Quantum Technol.*, vol. 2, no. 12, p. 1900070, 2019. doi: 10.1002/qute.201900070
- [21] M. Larocca, S. Thanasilp, S. Wang, K. Sharma, J. Biamonte, P. J. Coles, L. Cincio, J. R. McClean, Z. Holmes, and M. Cerezo, "Barren plateaus in variational quantum computing," *Nature Reviews Physics*, 2025. doi: 10.1038/s42254-025-00813-9
- [22] J. Jäger and R. V. Krems, "Universal expressiveness of variational quantum classifiers and quantum kernels for support vector machines," *Nature Communications*, vol. 14, p. 576, 2023. doi: 10.1038/s41467-023-36144-5
- [23] D. Stilck França and R. García-Patrón, "Limitations of optimization algorithms on noisy quantum devices," *Nature Physics*, vol. 17, pp. 1221–1227, 2021. doi: 10.1038/s41567-021-01356-3
- [24] A. Dal Pozzolo, O. Caelen, R. A. Johnson, and G. Bontempi, "Calibrating probability with undersampling for unbalanced classification," in *Proc. IEEE Symp. Series on Computational Intelligence (SSCI)*, 2015, pp. 159–166. doi: 10.1109/SSCI.2015.33
- [25] E. Altman, J. Blanuša, L. von Niederhäusern, B. Egressy, A. Anghel, and K. Atasu, "Realistic synthetic financial transactions for anti-money laundering models," in *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 36, 2023. *arXiv:2306.16424*. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2023/hash/5f38404edff6f3f642d6fa5892479c42-Abstract-Datasets_and_Benchmarks.html
- [26] E. Grant, L. Wossnig, M. Ostaszewski, and M. Benedetti, "An initialization strategy for addressing barren plateaus in parametrized quantum circuits," *Quantum*, vol. 3, p. 214, 2019. doi: 10.22331/q-2019-12-09-214
- [27] S. Thanasilp, S. Wang, M. Cerezo, and Z. Holmes, "Exponential concentration in quantum kernel methods," *Nature Communications*, vol. 15, Art. no. 5200, 2024. doi: 10.1038/s41467-024-49287-w.