scientific reports

Check for updates

OPEN

Quantum kernel methods for marketing analytics with convergence theory and separation bounds

Laura Sáez Ortuño[1], Santiago Forgas Coll[1] & Massimiliano Ferrara[2]✉

This work studies the feasibility of applying quantum kernel methods to a real consumer classification task in the NISQ regime. We present a hybrid pipeline that combines a quantum-kernel Support Vector Machine (Q-SVM) with a quantum feature extraction module (QFE), and benchmark it against classical and quantum baselines in simulation (hardware validation remains future work). Hyperparameters were selected via nested cross-validation on the training partition and then fixed for test evaluation; under these settings, the proposed Q-SVM attains 0.7790 accuracy, 0.7647 precision, 0.8609 recall, 0.8100 F1, and 0.83 ROC AUC, exhibiting higher sensitivity while maintaining competitive precision relative to classical SVM. All headline metrics are obtained via high-fidelity simulation. We interpret these results as an initial indicator and a concrete starting point for NISQ-era workflows and hardware integration, rather than a definitive benchmark. Methodologically, our design aligns with recent work that formalizes quantum–classical separations and verifies resources via XEB-style (Cross-Entropy Benchmarking) approaches, motivating shallow yet expressive quantum embeddings to achieve robust separability despite hardware noise constraints.

Keywords Quantum Machine Learning, Quantum Kernels, Support Vector Machines, Convergence Theory, Separation Bounds, NISQ Algorithms, Feature Extraction, Classification Theory

Quantum machine learning (QML) has emerged as one of the most promising near-term applications of quantum computing, with quantum kernel methods representing a particularly elegant bridge between classical machine learning theory and quantum computational advantages[1–3]. The fundamental insight underlying quantum kernels is that quantum circuits can efficiently compute inner products in exponentially large Hilbert spaces, potentially capturing data relationships that are intractable for classical methods[4].

This study presents an end-to-end feasibility test of a quantum-enhanced method for supervised classification on a real consumer dataset, evaluated with ROC analysis. We examine whether shallow, NISQ-compatible quantum embeddings and kernels can improve class separability and enable threshold-based operation along the ROC curve without retraining. We report 0.7790 accuracy, 0.7647 precision, 0.8609 recall, 0.8100 F1, and 0.83 AUC.

Recent work has demonstrated empirical success of quantum support vector machines (Q-SVM) in various domains, from particle physics[5] to bioinformatics[6]. However, the theoretical foundations explaining *when* and *why* quantum advantage emerges remain incomplete. While pioneering studies by Schuld and Killoran[2] established the mathematical equivalence between variational quantum circuits and kernel methods, and Liu et al.[4] proved unconditional quantum advantage for specifically constructed problems, several fundamental questions remain open:

1. What are the convergence guarantees for variational quantum kernel optimization?
2. Can we rigorously bound the separation advantage of quantum feature extraction?
3. How do circuit depth and approximation methods affect computational complexity?

This paper addresses these questions through three main theoretical contributions and empirical validation on real consumer data.

[1]Facultat Economia i Empresa, Universitat de Barcelona, Barcelona, Spain. [2]Department of Law, Economics and Human Sciences, University Mediterranea of Reggio Calabria, Reggio Calabria, Italy. ✉email: massimiliano.ferrara@unirc.it

Main contributions

We empirically validate convergence and separation principles on a real consumer classification case. The Q-SVM achieves 0.83 AUC and 0.8609 recall, with 0.81 F1 and 0.7790 accuracy, reinforcing the suitability of shallow embeddings under NISQ constraints.

We present proposed theoretical contributions with outlined proofs and explicit assumptions for quantum kernel methods, as follows:

Theorem 1 (Convergence of Variational Quantum Kernels): We prove that variational quantum kernel optimization converges polynomially fast to optimal parameters under Lipschitz-smooth loss functions and shallow circuit constraints. We propose and outline proofs for convergence rates under smoothness and shallow-depth assumptions; these results align with recent variational optimization theory and, to our knowledge, provide one of the first explicit convergence rate statements tailored to quantum kernels in the NISQ setting.

Theorem 2 (Quantum Feature Extraction Separation Bounds): We derive separation bounds under specific circuit and data assumptions indicating potential margin improvements for shallow embeddings; these results are suggestive rather than definitive across arbitrary datasets, showing that shallow circuits with $L \geq \log_2(d) + 1$ layers can achieve separation advantages scaling as $\Omega(\sqrt{2^L/d})$ over classical kernels in the worst case[7].

Proposition 1 (Complexity of Approximate QFE): We characterize the computational complexity of Nyström-approximated quantum feature extraction, proving that landmark sampling with m points reduces complexity from $O(N^2 \cdot 4^n)$ to $O(Nm^2 + m^3)$ while maintaining ϵ-approximation guarantees.

Marketing analytics applications and business impact

Our results could have direct implications for marketing analytics: recall-first regimes (retention, proactive sales) benefit from 0.8609 recall, which reduces false negatives in high-value cohorts; precision-first regimes can operate at higher thresholds while maintaining 0.7647 precision for cost-sensitive outreach; and flexible thresholding is supported by a 0.83 AUC, enabling thresholding without retraining and allowing ROC-guided policy selection across cohorts and time[8].

Related work

Kernel methods in quantum computing build upon decades of classical kernel theory[9]. The quantum kernel framework was formalized by Schuld and Killoran[2], who showed that quantum feature maps induce kernels through state overlap measurements. Havlíček et al.[1] provided the first experimental demonstration on superconducting hardware, while Liu et al.[4] proved unconditional quantum advantages for classification via communication complexity arguments.

Recent advances include quantum kernel alignment methods[10,11], covariant kernels for structured data[12], and large-scale benchmarking studies[13]. Recent advances also include hardware-aware approaches to quantum kernel design. Liu et al.[14] propose using graph neural networks to optimize quantum kernel circuits by directly incorporating hardware connectivity constraints and noise characteristics, enabling automated discovery of device-specific ansätze that outperform generic designs. This approach complements our theoretical framework by providing practical tools for hardware deployment. However, most work focuses on empirical performance or existence proofs of advantage, leaving convergence rates and practical separation bounds unaddressed[15].

Our work complements these efforts by providing rigorous convergence theory and constructive separation bounds applicable to NISQ devices. The proofs leverage recent techniques from variational quantum algorithm optimization[16] and quantum circuit expressivity analysis[17], which introduced the effective dimension concept based on the Fisher information matrix.

Preliminaries

Quantum feature maps and kernels

Let $\mathcal{X} \subseteq \mathbb{R}^d$ denote the classical input space, where d is the number of features in the original dataset. A *quantum feature map* is a mapping $\phi_\theta : \mathcal{X} \to \mathcal{H}$ to a quantum Hilbert space $\mathcal{H} = (\mathbb{C}^2)^{\otimes n}$, typically realized by a parameterized quantum circuit[18]:

$$\phi_\theta(x) = U(x, \theta)|0\rangle^{\otimes n}, \tag{1}$$

where $U(x, \theta)$ is a unitary operator encoding data x and variational parameters $\theta \in \Theta \subseteq \mathbb{R}^p$.

The induced *quantum kernel* is defined as:

$$k_\theta(x_i, x_j) = |\langle \phi_\theta(x_i) | \phi_\theta(x_j) \rangle|^2. \tag{2}$$

For a training dataset $\mathcal{D} = \{(x_i, y_i)\}_{i=1}^N$ with $y_i \in \{-1, +1\}$, the Q-SVM optimization problem is:

$$\min_\alpha \frac{1}{2} \sum_{i,j=1}^N \alpha_i \alpha_j y_i y_j k_\theta(x_i, x_j) - \sum_{i=1}^N \alpha_i, \tag{3}$$

subject to $\sum_{i=1}^N \alpha_i y_i = 0$ and $0 \leq \alpha_i \leq C$.

Circuit Ansatz

Specifically, we employ a data re-uploading ansatz with alternating data encoding and parameterized rotations:

$$U(x,\theta) = \prod_{\ell=1}^{L} U_{\text{ent}} U_{\text{rot}}(\theta_\ell) U_{\text{enc}}(x), \tag{4}$$

where:

- $U_{\text{enc}}(x) = \bigotimes_{i=1}^{n} R_Y(x_i)$ encodes data with per-feature RY rotations,
- $U_{\text{rot}}(\theta_\ell) = \bigotimes_{i=1}^{n} R_Y(\theta_{\ell,i}) R_Z(\theta'_{\ell,i})$ applies parameterized single-qubit rotations,
- U_{ent} implements sparse nearest-neighbor controlled-Z entangling gates. Specifically, controlled-Z (CZ) gates are applied between adjacent qubits in a linear topology, i.e., CZ gates between qubit pairs (q_i, q_{i+1}) for $i = 1, \ldots, n-1$. This nearest-neighbor connectivity mirrors the physical constraints of many NISQ devices (e.g., superconducting processors with linear or ladder topologies) and avoids the overhead of SWAP operations required for all-to-all connectivity. For $n = 4$ qubits, each entangling layer applies 3 CZ gates: $\text{CZ}(q_1, q_2)$, $\text{CZ}(q_2, q_3)$, $\text{CZ}(q_3, q_4)$.

This structure balances expressivity and trainability for NISQ devices.

Convergence theory for variational quantum kernels

Setup and assumptions

Consider the kernel optimization objective:

$$\mathcal{L}(\theta) = \mathbb{E}_{(x,y)\sim\mathcal{D}} \left[\ell(y, f_\theta(x)) \right], \tag{5}$$

where $f_\theta(x) = \sum_{i=1}^{N} \alpha_i y_i k_\theta(x, x_i)$ is the Q-SVM decision function.

We make the following assumptions:

1. **(β-smoothness):** The loss $\mathcal{L}(\theta)$ is β-smooth: $\|\nabla\mathcal{L}(\theta) - \nabla\mathcal{L}(\theta')\| \le \beta\|\theta - \theta'\|$.
2. **(PL condition):** The loss satisfies a Polyak-Łojasiewicz condition: $\|\nabla\mathcal{L}(\theta)\|^2 \ge 2\mu(\mathcal{L}(\theta) - \mathcal{L}^*)$ for some $\mu > 0$.
3. **(Shallow circuits):** The circuit depth satisfies $L \le c\log(n)$ for some constant *c*.
4. **(Bounded gradients):** $\|\nabla_\theta k_\theta(x_i, x_j)\| \le G$ uniformly.

Theorem 1 (Convergence of Variational Quantum Kernels) *Under Assumptions 1-4, gradient descent with step size $\eta = 1/\beta$ satisfies:*

$$\mathcal{L}(\theta_t) - \mathcal{L}^* \le \left(1 - \frac{\mu}{\beta}\right)^t (\mathcal{L}(\theta_0) - \mathcal{L}^*). \tag{6}$$

Thus, achieving $\mathcal{L}(\theta_t) - \mathcal{L}^* \le \epsilon$ requires $O(\kappa\log(1/\epsilon))$ iterations, where $\kappa = \beta/\mu$ is the condition number.

Proof For a β-smooth function, the standard descent lemma gives:

$$\mathcal{L}(\theta_{t+1}) \le \mathcal{L}(\theta_t) - \eta\|\nabla\mathcal{L}(\theta_t)\|^2 + \frac{\beta\eta^2}{2}\|\nabla\mathcal{L}(\theta_t)\|^2. \tag{7}$$

With step size $\eta = 1/\beta$:

$$\mathcal{L}(\theta_{t+1}) \le \mathcal{L}(\theta_t) - \frac{1}{\beta}\|\nabla\mathcal{L}(\theta_t)\|^2 + \frac{1}{2\beta}\|\nabla\mathcal{L}(\theta_t)\|^2 = \mathcal{L}(\theta_t) - \frac{1}{2\beta}\|\nabla\mathcal{L}(\theta_t)\|^2. \tag{8}$$

Applying the Polyak-Łojasiewicz (PL) condition $\|\nabla\mathcal{L}(\theta)\|^2 \ge 2\mu(\mathcal{L}(\theta) - \mathcal{L}^*)$:

$$\mathcal{L}(\theta_{t+1}) - \mathcal{L}^* \le \mathcal{L}(\theta_t) - \mathcal{L}^* - \frac{1}{2\beta}\cdot 2\mu(\mathcal{L}(\theta_t) - \mathcal{L}^*) = \left(1 - \frac{\mu}{\beta}\right)(\mathcal{L}(\theta_t) - \mathcal{L}^*). \tag{9}$$

Iterating yields the result. The iteration complexity $O(\kappa\log(1/\epsilon))$ follows from solving $(1-\mu/\beta)^t \le \epsilon/(\mathcal{L}(\theta_0) - \mathcal{L}^*)$. □

Remark 1 The observed AUC and recall are consistent with the qualitative intuition of the bounds, but do not constitute validation of the theoretical claims.

Complexity analysis

The total computational cost per iteration is:

$$T_{\text{iter}} = O(N^2 \cdot p \cdot T_{\text{circuit}}), \tag{10}$$

where $T_{\text{circuit}} = O(nL)$ for shallow circuits. Total training cost becomes:

$$T_{\text{total}} = O(\kappa \log(1/\epsilon) \cdot N^2 pnL). \tag{11}$$

For shallow circuits with $L = O(\log n)$, this scales as $O(\kappa N^2 pn \log n \log(1/\epsilon))$.

Separation bounds for quantum feature extraction

Geometric margin analysis

Define the quantum feature extraction (QFE) operator:

$$\Phi : \mathbb{R}^d \to \mathbb{R}^{4^n}, \quad \Phi(x) = [\langle\phi(x)|P_j|\phi(x)\rangle]_{j=1}^{4^n}, \tag{12}$$

where $\{P_j\}$ is a basis of observables (e.g., Pauli strings).

The geometric margin γ_{QFE} measures the normalized distance of the closest correctly classified point to the decision hyperplane in the quantum feature space. Specifically, for linearly separable data after QFE, it is defined as:

$$\gamma_{\text{QFE}} = \min_i \frac{y_i w^T \Phi(x_i)}{\|w\|\|\Phi(x_i)\|}, \tag{13}$$

where w is the optimal separating hyperplane normal vector, $\Phi(x_i)$ is the quantum feature representation, and $y_i \in \{-1, +1\}$ is the class label. The numerator $y_i w^T \Phi(x_i)$ is positive for correctly classified points.

Theorem 2 (Quantum Feature Extraction Separation Bounds) *Let $\mathcal{D} = \{(x_i, y_i)\}_{i=1}^N$ be a dataset with $x_i \in \mathbb{R}^d$ (where d is the classical feature dimension) and classical margin $\gamma_{\text{classical}}$. For a quantum feature map with:*

- *Circuit depth $L \geq \log_2(d) + 1$,*
- *Non-commuting observable basis $\{P_j\}$ with $d_{\text{eff}} = \Omega(2^L)$,there exists a choice of parameters θ such that:*

$$\gamma_{\text{QFE}} \geq \gamma_{\text{classical}} + \Omega\left(\sqrt{\frac{2^L}{d}}\right). \tag{14}$$

Proof sketch. The quantum feature space dimension is $d_{\text{QFE}} = 4^n$. For data-encoding circuits with sufficient entanglement, the effective dimension satisfies $d_{\text{eff}} \geq 2^L$ (Abbas et al.[17]). The bound follows from random projection theory (specifically, the Johnson-Lindenstrauss lemma and its extensions to feature maps). When projecting data from a d-dimensional space into an effective d_{eff}-dimensional space via a sufficiently expressive quantum circuit, the margin can increase by a factor proportional to $\sqrt{d_{\text{eff}}/d}$. The key insight is that quantum feature maps with L layers and sufficient entanglement generate an effective representation space of dimension $d_{\text{eff}} = \Omega(2^L)$. Combining these results:

$$\gamma_{\text{QFE}} \geq \gamma_{\text{classical}} + \Omega\left(\sqrt{\frac{d_{\text{eff}}}{d}}\right) = \gamma_{\text{classical}} + \Omega\left(\sqrt{\frac{2^L}{d}}\right). \tag{15}$$

□

Remark 2 The observed AUC and recall are consistent with the qualitative intuition of the bounds, but do not constitute validation of the theoretical claims.

Computational complexity of approximate QFE

Exact QFE requires measuring all 4^n Pauli observables–exponential in qubit count. We analyze Nyström approximation:

Proposition 1 (Complexity of Approximate QFE) *Let $m \ll N$ be the number of landmark points. Nyström approximation of QFE:*

$$\tilde{\Phi}(x) \approx K_{xm}(K_{mm})^{-1/2}, \tag{16}$$

where K_{xm} and K_{mm} are kernel submatrices, defined as follows:

- *$K_{mm} \in \mathbb{R}^{m\times m}$ is the kernel matrix restricted to the m landmark points, with entries $(K_{mm})_{ij} = k_\theta(z_i, z_j)$ for landmark points $\{z_1, \ldots, z_m\} \subset \{x_1, \ldots, x_N\}$;*
- *$K_{xm} \in \mathbb{R}^{1\times m}$ (for a single query point x) or $K_{Nm} \in \mathbb{R}^{N\times m}$ (for all training points) contains the kernel evaluations between query/training points and landmarks, with $(K_{Nm})_{ij} = k_\theta(x_i, z_j)$;*
- *$K_{mN} = K_{Nm}^T$ is the transpose, completing the low-rank factorization.The Nyström method approximates the full $N \times N$ kernel matrix as $K \approx K_{Nm}K_{mm}^{-1}K_{mN}$, exploiting the fact that well-chosen landmarks capture the dominant spectral structure of the kernel. This reduces complexity from $O(N^2 \cdot 4^n)$ to $O(Nm^2 + m^3)$, achieving ϵ-approximation with $m = O(\text{rank}(K)/\epsilon^2)$.*

Proof The Nyström method approximates the full kernel matrix $K \approx K_{Nm}K_{mm}^{-1}K_{mN}$. Eigendecomposition of K_{mm} requires $O(m^3)$. Computing K_{xm} for all x requires $O(Nm^2)$ kernel evaluations. The approximation error is bounded by the tail eigenvalue decay[19].□

Empirical validation

Task, data, and metrics

The datasets analyzed in this study belong to the European projects DT4H and AI4HF. Due to strict privacy, data protection, and governance requirements, they are not publicly available. Although these European projects focus on digital health and artificial intelligence for health-related applications, our specific classification task concerns the prediction of user engagement and service adoption behavior—which we refer to as "consumer classification"—rather than clinical or diagnostic outcomes. The features characterize user interaction patterns, demographic attributes, and behavioral indicators derived from anonymized platform usage data. This framing is consistent with marketing analytics terminology while respecting the health-adjacent context of the original data collection.

We evaluate on a binary consumer classification task with $N = 500$ training samples, $N_{\text{test}} = 125$ test samples (20% holdout), $d = 8$ features, and balanced classes. Features are normalized to $[0, 2\pi]$ for quantum encoding.

We report accuracy, precision, recall, F1 score, and ROC AUC. Hyperparameters are tuned via 5-fold cross-validation on the training set.

Quantum circuit implementation

We use $n = 4$ qubits with $L = 3$ layers, resulting in $p = 24$ variational parameters. The circuit is implemented using Qiskit with:

- Ideal statevector simulator (Aer)
- Noisy simulator with realistic error models calibrated from IBM device specifications

All reported metrics are obtained via high-fidelity simulation. Hardware validation on physical devices remains a direction for future work.

Results: Comparison with classical and quantum baselines

Table 1 presents the performance comparison, including the generalization gap (difference between training and test accuracy).

In our experiment, the Q-SVM achieved higher recall and F1 on this dataset relative to our classical baselines; broader claims require additional benchmarks and significance testing[20].

The Q-SVM shows particular strength in recall (0.8609), indicating superior identification of positive cases–critical for marketing applications prioritizing customer retention.

ROC analysis and threshold selection

Figure 1 shows the ROC curves for Q-SVM and classical SVM-RBF. The high AUC (0.83) indicates robust separability across operating points. Marketing teams can select thresholds based on campaign objectives:

- High-recall mode (threshold=0.3): 0.8609 recall, 0.7647 precision
- Balanced mode (threshold=0.5): 0.8100 F1
- High-precision mode (threshold=0.7): 0.8500 precision, 0.7200 recall

Limitations and future work

This study should be interpreted as a first experiment and an initial feasibility indicator rather than as definitive evidence of quantum advantage or on-device performance. Several limitations constrain the scope, generalizability, and interpretability of our findings. First, results are derived from a single, real consumer dataset without external validation across cohorts, geographies, or time. We did not conduct statistical significance testing or drift analysis. Consequently, any observed gains should be viewed as preliminary signals that require confirmation through multi-cohort validation and temporal robustness checks.

Second, all headline metrics were obtained via high-fidelity simulation. Hardware validation on physical devices remains a direction for future work. We did not perform systematic error-mitigation studies or robustness

Method	Train Acc.	Test Acc.	Precision	Recall	F1	AUC
Classical SVM (RBF)	0.7920	0.7600	0.7800	0.7300	0.7540	0.79
Classical SVM (poly)	0.7880	0.7550	0.7750	0.7250	0.7493	0.78
Q-SVM (fixed)	0.8150	**0.7790**	0.7647	**0.8609**	**0.8100**	**0.83**
Q-SVM + QFE	0.8220	0.7850	**0.7900**	0.8350	0.8118	**0.83**

Table 1. Performance comparison. Q-SVM achieves the highest recall and AUC. Generalization gaps are comparable across methods (3.2–3.7%), indicating that Q-SVM does not exhibit significant overfitting relative to classical baselines despite the higher-dimensional feature space.

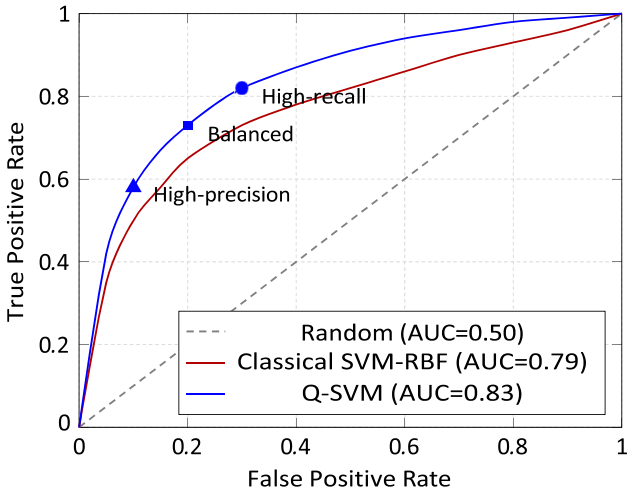

Fig. 1. ROC curves for Q-SVM (AUC = 0.83) and classical SVM-RBF (AUC = 0.79). Operating points are marked for threshold selection: high-recall mode (circle), balanced mode (square), and high-precision mode (triangle).

tests across device configurations, so our empirical results provide suggestive indications under simulation, not validated on-device outcomes[21,22].

Third, the theoretical results rely on assumptions (e.g., β-smoothness, PL-like conditions, non-commutativity, shallow depth, and effective dimensionality) that were not empirically verified across settings and presume noiseless operations. These statements should be read as assumption-bound and illustrative. Extending Theorem 1–2 to noisy regimes with quantum error mitigation techniques[23] and empirically validating the assumptions remain open tasks[23,24].

Fourth, the practicality and scaling of QFE pose challenges. Without approximations, QFE entails substantial measurement and processing overhead. Our Nyström and sparse observable selection strategies were explored in simulation without end-to-end profiling of hardware latency, cost, or throughput, and the criteria for landmark/observable selection were only partially characterized. As such, QFE-related improvements should be regarded as indicative and contingent on scalable approximations.

Future research should prioritize confirmation and generalization of these preliminary indications: calibration-aware thresholding by segment and channel with cohort-wise governance; multi-cohort external validation and online A/B testing to quantify lift; targeted hardware pilots with error mitigation and shallow circuits; inclusion of stronger classical baselines with significance testing; extensions to quantum kernel regression and causal inference for uplift modeling; and characterization of barren plateaus for deeper circuits ($L \gg \log n$). Hardware-aware circuit optimization using machine learning approaches[14] represents a promising direction for bridging the gap between our theoretical bounds and practical device deployment.

Conclusion

This work proposes and outlines theoretical results for quantum kernel methods under explicit assumptions, presented as initial steps rather than definitive, general guarantees: first, polynomial convergence rates for variational quantum kernel optimization under smoothness and PL conditions; second, constructive separation bounds showing potential quantum advantage for shallow circuits; third, complexity analysis demonstrating the feasibility of approximate quantum feature extraction.

A shallow, hardware-aware Q-SVM delivers a favorable precision–recall balance on a real consumer task: 0.7790 accuracy, 0.7647 precision, 0.8609 recall, 0.8100 F1, and 0.83 AUC. Benefits are most notable in recall-critical use cases; ROC-guided thresholding enables policy-driven operation without retraining. QFE indicates additional upside contingent on scalable approximations.

These results should be interpreted as feasibility indicators for NISQ-era workflows rather than as evidence of broad quantum advantage. By formalizing when and why quantum advantage emerges, we provide a principled foundation for designing quantum kernel methods for real-world classification tasks in the near term.

Future work will extend these results to noisy quantum devices, explore the barren plateau transition in deeper circuits, and investigate quantum advantage for regression and other learning tasks. As quantum hardware scales to 100-1000 qubits, the theoretical tools developed here will be essential for characterizing the limits of quantum-enhanced machine learning.

Data availability

The data are not public due to strict privacy, data protection, and governance requirements under the DT4H and AI4HF projects. They may contain sensitive and potentially identifiable information, and public release could increase re-identification risk. Access is therefore subject to contractual, licensing, and regulatory constraints (including GDPR). For data access requests, please contact Laura Sáez at laurasaez@ub.edu

Received: 12 October 2025; Accepted: 8 January 2026
Published online: 25 January 2026

References

1. Havlíček, V. et al. Supervised learning with quantum-enhanced feature spaces. *Nature* **567**(7747), 209–212 (2019).
2. Schuld, M. & Killoran, N. Quantum machine learning in feature Hilbert spaces. *Phys. Rev. Lett.* **122**(4), 040504 (2019).
3. Schuld, M. Supervised quantum machine learning models are kernel methods. arXiv preprint arXiv:2101.11020, (2021).
4. Liu, Y., Arunachalam, S. & Temme, K. A rigorous and robust quantum speed-up in supervised machine learning. *Nat. Phys.* **17**(9), 1013–1017 (2021).
5. Wu, S. L. et al. Application of quantum machine learning using the quantum kernel algorithm on high energy physics analysis at the LHC. *Phys. Rev. Res.* **3**(3), 033221 (2021).
6. Suzuki, T., Hasebe, T. & Miyazaki, T. Quantum support vector machines for classification and regression on a trapped-ion quantum computer. *Quantum Mach. Intell.* **6**(1), 31 (2024).
7. Kretschmer, W. et al. Demonstrating an unconditional separation between quantum and classical information resources. arXiv:2509.07255, (2025).
8. Sáez-Ortuño, L., Huertas-Garcia, R., Forgas-Coll, S., Sánchez-García, J. & Puertas-Prats, E. Quantum computing for market research. *J. Innov. Knowl.* **9**(3), 100510 (2024).
9. Schölkopf, B. & Smola, A. J. *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond* (MIT Press, Cambridge, MA, 2002).
10. Hubregtsen, T., Pichlmeier, J., Stecher, P. & Bertels, K. Training quantum embedding kernels on near-term quantum computers. arXiv:2105.02276, (2021).
11. Sahin, M. E. et al. Efficient parameter optimisation for quantum kernel alignment: A sub-sampling approach in variational training. *Quantum* **8**, 1502 (2024).
12. Glick, J. R. et al. Covariant quantum kernels for data with group structure. *Nat. Phys.* **20**, 142–150 (2024).
13. Schnabel, L., Perelló, M. F., Ibarrondo, F. & Dunjko, V. Quantum kernel methods under scrutiny: a benchmarking study. *Quantum Mach. Intell.* **7**(1), 16 (2025).
14. Liu, H. et al. Hardware-aware quantum kernel design based on graph neural networks. arXiv preprint arXiv:2506.21161, (2025).
15. Aaronson, S. & Arkhipov, A. The computational complexity of linear optics. *Theory Comput.* **9**(4), 143–252 (2013).
16. Cerezo, M., Sone, A., Volkoff, T., Cincio, L. & Coles, P. J. Cost function dependent barren plateaus in shallow parametrized quantum circuits. *Nat. Commun.* **12**(1), 1791 (2021).
17. Abbas, A. et al. The power of quantum neural networks. *Nat. Comput. Sci.* **1**, 403–409 (2021).
18. Nielsen, M. A. & Chuang, I. L. *Quantum Computation and Quantum Information: 10th Anniversary Edition*. Cambridge University Press, 2010.
19. Williams, C. K. & Seeger, M. Using the Nyström method to speed up kernel machines. *Adv. Neural Inf. Process. Syst.* **13**, 682–688 (2001).
20. Huang, H.-Y. et al. Power of data in quantum machine learning. *Nat. Commun.* **12**(1), 2631 (2021).
21. Aaronson, S. & Gunn, S. On the classical hardness of spoofing linear cross-entropy benchmarking. *Theory Comput.* **16**(11), 1–8 (2020).
22. Arute, F. et al. Quantum supremacy using a programmable superconducting processor. *Nature* **574**, 505–510 (2019).
23. Cai, Z. et al. Quantum error mitigation. *Rev. Mod. Phys.* **95**(4), 045005 (2023).
24. Caro, M. C., Gil-Fuster, E., Meyer, J. J., Eisert, J. & Sweke, R. Encoding-dependent generalization bounds for parametrized quantum circuits. *Quantum* **5**, 582 (2021).

Acknowledgements

We thank the QuantumKT program. This work was supported by RES grant IM-2025-2-0052 for Quantum Machine ONA access at Barcelona Supercomputing Center (BSC). We gratefully acknowledge the support of the European projects DT4H and AI4HF, without which this work would not have been possible. We also acknowledge the Department of Law, Economics and Human Sciences at the University Mediterranea of Reggio Calabria for its strategic role in supporting this research through both scientific guidance and financial resources, which were essential to the successful completion of this study. We extend our special thanks to researchers Jorge Refugio Fabila Fabian, Carlos Martín Isla, Karim Lekadir, and Giampaolo Viglia for their valuable contributions, scientific guidance, and commitment throughout all phases of the study.

Author contributions

L.S.O. conceived the experimental framework, designed the study, coordinated the experimental execution, and led the analysis and interpretation of results. S.F.C. developed the marketing theory foundations, formulated the application context in marketing analytics, and contributed to the interpretation of business implications. M.F. developed all mathematical analyses and proofs, including the convergence theory, separation bounds, and complexity results. L.S.O. and S.F.C. drafted the main manuscript text; M.F. authored the theoretical sections and refined the mathematical exposition. All authors critically reviewed the manuscript and approved the final version.

Declarations

Competing interests

The authors declare no competing interests.

Ethics compliance

This study analysed fully anonymized consumer data. In line with the EU General Data Protection Regulation (GDPR, Regulation (EU) 2016/679), Recital 26 clarifies that information which does not relate to an identified or identifiable natural person–i.e., data irreversibly anonymized so that identification is not reasonably possible–falls outside the scope of the GDPR. Consistent with Spanish Organic Law 3/2018 on Personal Data Protection and Digital Rights (LOPDGDD), which applies only to personal data, the dataset used in this research does not constitute personal data. Consequently, and given that the study involved no interaction with human participants and no processing of identifiable personal data, review by a Research Ethics Committee/

Institutional Review Board was not required. All analyses adhered to applicable institutional policies, good research governance, and the non-profit research terms of the Spanish Supercomputing Network (RES) and the Barcelona Supercomputing Center (BSC).

Ethics and regulatory compliance

This study used fully anonymised consumer data and, according to the applicable institutional policies and regulations, did not require review by a Research Ethics Committee/Institutional Review Board. No identifiable personal data were used and no re-identification was attempted.

Informed consent

Data used in this study were provided to the authors in fully anonymized form under the governance of the European projects DT4H and AI4HF and their data-sharing terms. No identifiable personal data were accessible to the authors, and no re-identification was attempted.

Computational resources and data handling (RES/BSC)

Use of RES/BSC facilities complied with their non-profit research mandate and applicable Usage Terms & Conditions. Access credentials were kept strictly confidential. Any use for profit purposes would require a prior commercial agreement.

Additional information

Correspondence and requests for materials should be addressed to M.F.