Geometry Matters: Kernel Concentration in Quantum Feature Maps

Anonymous authors
Paper under double-blind review

Abstract

Quantum machine learning (QML) is motivated by the premise that quantum circuits can embed classical data into feature Hilbert spaces in which learning may become more effective. We evaluate this premise empirically by comparing two feature maps, the entangling ZZFeatureMap and simple angle encoding, on multiclass MNIST classification. Using four kernel geometry metrics (kernel target alignment, intra/inter similarity ratio, effective rank, and off-diagonal variance), we show that ZZFeatureMap suffers severe kernel concentration as task complexity increases: its similarity matrix becomes nearly uniform, its effective rank exceeds 300, and its alignment with the labels collapses. In contrast, angle encoding preserves a structured, low-rank kernel and achieves 83% accuracy on 8-class, 8-qubit MNIST, outperforming ZZFeatureMap by 42 percentage points. We find that kernel target alignment and off-diagonal variance are reliable predictors of classification performance, while effective rank reveals when the kernel becomes diffusely concentrated. These metrics offer practical diagnostics for selecting feature maps in near-term quantum kernels. Overall, the results indicate that practical QML performance depends less on generic circuit expressivity than on whether the chosen feature map induces a useful geometry for the data.

1 Introduction

Quantum machine learning (QML) is motivated by the possibility that quantum circuits can embed classical data into high-dimensional Hilbert spaces in which learning may become more effective. This idea is closely related to the feature-space view of learning: rather than operating on the original input representation alone, a model may benefit from a transformed representation in which the data are easier to separate. In the quantum setting, a data encoding circuit implements this transformation by mapping a classical input vector to a quantum state. The usefulness of a QML model therefore depends not only on the size of the Hilbert space but also on whether the chosen quantum feature map induces a geometry that is meaningful for the task.

A natural test of this idea is to compare feature maps that differ fundamentally in structure. Simple angle encoding maps each classical feature to an independent single-qubit rotation, producing a separable quantum state. In contrast, the entangling ZZFeatureMap introduces controlled interactions between qubits, coupling features and generating a more complex, higher-order representation. The angle-encoded map can be simulated efficiently on a classical computer, since the kernel of a separable state is the product of independent single-qubit terms. The entangled map is believed, though not proven, to be difficult to simulate classically, which can prevent this kind of factorization. This pairing is not arbitrary: Havlíček et al. (2019) introduced a feature-map construction incorporating two-qubit ZZ interactions and contrasted it with product-state feature maps, arguing that the latter admit a classical factorization and therefore offer no expected quantum advantage. We use the same pairing here to isolate the affects of entangling feature maps while holding the qubit count and single-qubit gates fixed.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

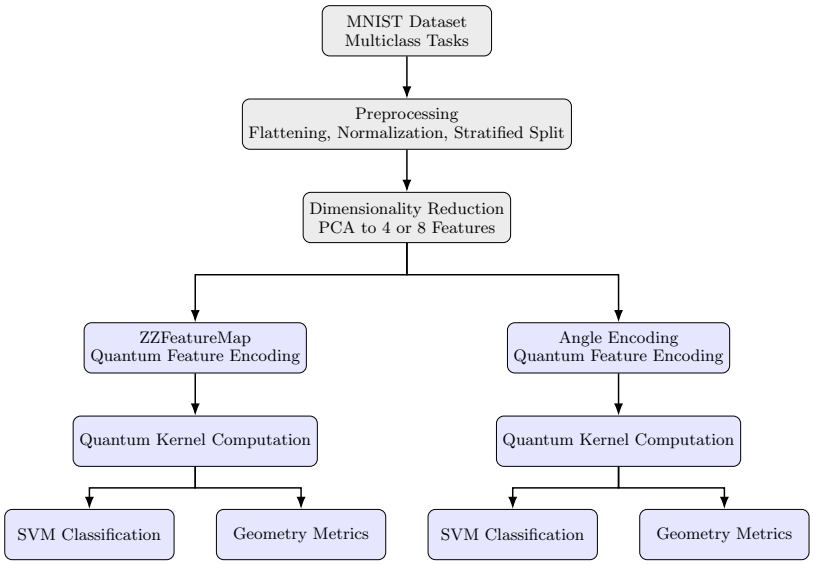


Figure 1: Overall experimental framework. MNIST data are preprocessed and reduced via PCA, then encoded using either ZZFeatureMap or angle encoding before quantum kernel computation and classification.

Figure 1 shows the experimental design. We first preprocess the MNIST data and reduce its dimensionality with principal component analysis (PCA) to match the number of qubits. We then feed the reduced data into two QSVM (Quantum Support Vector Machine) models, one with ZZFeatureMap encoding and one with angle encoding. After computing each kernel, we measure classification accuracy and the kernel geometry metrics.

Kernel concentration is a known failure mode for expressive quantum feature maps: as circuit depth or qubit count grows, pairwise similarities between data points collapse towards a single value, and the kernel loses the ability to distinguish between classes. It has not been established whether this failure mode occurs for the ZZFeatureMap under realistic classification settings, or how the ZZFeatureMap compares with a feature map that has no entangling structure at all. Existing evaluations of quantum feature maps typically report accuracy alone, without examining the geometry of the kernel that produced it, which leaves little insight into why a given map succeeds or fails. This gap motivates a direct comparison: does the entangling structure of the ZZFeatureMap yield a kernel that is more useful for classification than the kernel produced by simple angle encoding?

1.1 Contributions

- We evaluate angle encoding and ZZFeatureMap on multiclass MNIST classification under matched qubit count, single-qubit gate structure, and task, isolating entangling feature map as the variable of interest.
- We use four complementary metrics (kernel target alignment, intra/inter-class similarity ratio, effective rank, and off-diagonal variance) to characterize the structure of a quantum kernel independently of downstream accuracy.
- As task complexity increases, the ZZFeatureMap's kernel becomes nearly uniform, its effective rank exceeds 300, and its alignment with class labels collapses. Angle encoding exhibits none of these symptoms and retains a structured, low-rank kernel throughout.
- The difference in kernel geometry is reflected in classification accuracy: angle encoding reaches 83% on 8-class, 8-qubit MNIST, 42 percentage points above the ZZFeatureMap.
- Kernel target alignment and off-diagonal variance correlate with classification performance, while effective rank flags concentration as it emerges.

The remainder of this paper is organized as follows. Section 2 reviews related work on quantum feature maps and kernel-based models. Section 3 describes the tasks, the preprocessing pipeline, the kernel-based models, the two feature maps, and the four kernel geometry metrics used in the evaluation. Section 4 presents our empirical results, comparing classification accuracy and kernel geometry across both feature maps as task complexity increases. Section 5 discusses the implications of these findings for feature map selection in near-term quantum kernel methods and concludes the study.

2 Background and Related Work

Quantum machine learning can be understood through the lens of feature-space learning. In classical kernel methods, data are mapped into a higher-dimensional feature space where classification may become easier. Schuld and Killoran extend this idea to quantum models by interpreting quantum data encoding circuits as feature maps into Hilbert space (Schuld & Killoran, 2019). Under this view, a classical input x is encoded into a quantum state $|\phi(x)\rangle$, and learning depends on the geometry induced by the chosen encoding circuit. Theoretical work has shown that quantum kernels can offer provable advantages over classical methods in certain settings (Liu et al., 2021; Huang et al., 2021). These advantages are governed more by the geometric relationship between kernels than by circuit expressivity or entanglement alone. However, realizing these advantages in practice remains challenging, particularly in the noisy intermediate-scale quantum (NISQ) era (Preskill, 2018), where qubit counts, circuit depth, and simulation cost impose severe constraints.

A key obstacle is kernel concentration. In high-dimensional spaces, the off-diagonal entries of the kernel matrix can become almost constant for a random data distribution, reducing discriminative power. Thanasilp et al. (2024) proved that, for a wide class of quantum feature maps, the variance of the kernel estimator decays exponentially with the number of qubits. This concentration effect is particularly relevant for entangling feature maps, which generate complex representations but risk producing near-uniform matrices that may erase class structure. The geometry of the feature map determines the kernel's inductive bias (Kübler et al., 2021), making it essential to quantify whether a given map produces a useful geometry for the data.

To quantify whether a kernel matrix preserves useful class structure, we rely on kernel target alignment (Cortes et al., 2012). This metric measures the agreement between the kernel similarity matrix and the ideal label-based similarity matrix, providing a direct way to assess whether a feature map's geometry aligns with the classification task. While alignment has been established as a reliable predictor of performance in classical settings, its application to diagnosing concentration in quantum feature maps, particularly for entangling encodings, remains underexplored.

This paper directly addresses this gap by comparing two feature maps, the entangling ZZFeatureMap and simple angle encoding, as task complexity increases. Rather than treating the feature map as an incidental detail, we evaluate how encoding choice alters kernel geometry and whether geometric metrics, including alignment, predict classification performance. We use four diagnostic metrics to quantify concentration and demonstrate that these metrics reliably predict when a feature map will succeed or fail.

3 Methodology

3.1 Tasks and Preprocessing

We evaluate quantum kernel geometry on MNIST-derived image classification tasks with increasing complexity. The experiments include 4-class and 8-class classification tasks. Multiclass tasks test whether the kernel geometry remains stable as the number of classes increases.

All images are flattened from their original 28×28 pixel representation and normalized before dimensionality reduction. We apply principal component analysis (PCA) to reduce each input to either 4 or 8 features, corresponding to 4-qubit and 8-qubit quantum circuits, and

use the reduced feature vectors to construct train-test splits for each experimental condition. This preprocessing step reflects the qubit limitations of near-term quantum simulation and allows the same reduced feature representation to be used consistently across experiments.

3.2 Quantum Kernel SVM

The implicit quantum model is implemented as a quantum kernel support vector classifier (QSVC). For each input vector x, a quantum feature map $U_\phi(x)$ prepares an encoded quantum state,

$$|\phi(x)\rangle = U_\phi(x)|0\rangle^{\otimes n}.$$

The similarity between two inputs is then computed as the squared overlap between their encoded quantum states,

$$K(x, x') = |\langle\phi(x)|\phi(x')\rangle|^2.$$

These pairwise similarities form a quantum kernel matrix, which is passed to a support vector classifier as a precomputed kernel. In this setup, the quantum circuit defines the feature-space geometry, while the kernel method solves the final classification problem. This makes QSVC a direct empirical test of the implicit quantum feature-space approach described by Schuld and Killoran Schuld & Killoran (2019).

For multiclass classification, QSVC uses a "one-vs-one" decomposition. For a task with C classes, this requires training

$$N_{\text{OvO}} = \frac{C(C-1)}{2}$$

binary classifiers, one for each pair of classes. Therefore, 4-class experiments require 6 classifiers, and 8-class experiments require a total of 28 classifiers.

3.3 Feature Encodings

We study two feature maps that encode classical data into quantum states through fundamentally different circuit structures.

3.3.1 Angle Encoding

Angle encoding maps each classical feature directly to a rotation on an independent qubit. For an input vector $x = (x_1, x_2, \ldots, x_d)$, the resulting quantum state is

$$|\phi(x)\rangle = \bigotimes_{i=1}^{d} R_y(x_i)|0\rangle,$$

where $R_y(x_i)$ denotes a single-qubit rotation about the y-axis by the angle x_i. Each qubit is rotated independently, so no entangling operations are applied during the feature map stage, and the resulting state is a product state across all d qubits. This gives angle encoding a simple, low-depth circuit whose kernel can be evaluated efficiently on a classical computer, since the inner product between two states factorizes into a product of independent single-qubit terms.

3.3.2 ZZFeatureMap

The ZZFeatureMap Havlíček et al. (2019) extends this construction by introducing entangling interactions between qubits. In addition to feature-dependent single-qubit rotations, the circuit applies two-qubit ZZ interaction terms of the form

$$\exp\left(if(x_i, x_j)Z_iZ_j\right),$$

where Z_i and Z_j are operators acting on qubits i and j, and $f(x_i, x_j)$ is a feature-dependent interaction function. Conceptually, the single-qubit rotations encode individual feature values, while the ZZ interactions encode pairwise relationships between features. This additional structure makes the ZZFeatureMap more expressive than angle encoding, but it may also produce feature-space geometries that become less discriminative as dimensionality or task complexity increases.

3.4 Kernel Geometry Metrics

To understand why different feature maps lead to different classification performance, we directly analyze the geometry of the quantum kernel matrix. For each experimental condition, the kernel matrix $\mathbf{K}$ with entries $K_{ij} = K(x_i, x_j)$ is computed on the training set, and we evaluate four complementary metrics that diagnose kernel concentration and class separability.

Kernel Target Alignment (KTA) We measure how well the kernel aligns with the ideal label kernel, defined as $Y_{ij} = 1$ if $y_i = y_j$ and 0 otherwise. Using centered alignment Cortes et al. (2012):

$$\text{KTA} = \frac{\langle \mathbf{K}_c, \mathbf{Y}_c \rangle_F}{\|\mathbf{K}_c\|_F \|\mathbf{Y}_c\|_F}$$

where $\mathbf{K}_c = \mathbf{K} - \bar{K}$, $\mathbf{Y}_c = \mathbf{Y} - \bar{Y}$. Higher KTA indicates better agreement with class structure.

Intra/Inter Similarity Ratio Let $\mathcal{S}_{\text{intra}}$ be the set of kernel values for same-class pairs and $\mathcal{S}_{\text{inter}}$ for different-class pairs. The ratio

$$R = \frac{\text{mean}(\mathcal{S}_{\text{intra}})}{\text{mean}(\mathcal{S}_{\text{inter}})}$$

quantifies class separation; $R > 1$ is desirable, and $R \approx 1$ signals loss of discriminative power.

Effective Rank The participation ratio of the kernel eigenvalues λ_i:

$$\text{rank}_{\text{eff}} = \frac{(\sum_i \lambda_i)^2}{\sum_i \lambda_i^2}.$$

This measures how many dimensions the kernel effectively uses. A very low rank (close to 1) indicates that all points are nearly identical, which suggests a collapsed kernel. A low-to-moderate rank suggests a compact, potentially discriminative representation. A very high rank implies the kernel spreads data thinly across multiple dimensions, which can dilute class separation.

Off-Diagonal Variance The variance of all off-diagonal kernel entries,

$$C = \text{Var}\big(\{K_{ij} : i < j\}\big),$$

directly measures kernel concentration. High variance means structured similarities; low variance implies that the off-diagonal entries are nearly constant, the hallmark of an uninformative, concentrated kernel.

These metrics are computed on the training set for each configuration, using a representative random seed (42) and the largest sample size for multiclass classification ($n = 400$), where concentration effects are most visible. Accuracy results over five seeds confirm the trends are robust.

3.5 Evaluation Protocol

All experiments are repeated using the same set of seed values across experimental conditions. These seeds control stochastic components such as train-test splitting and model initialization where applicable, allowing results to be compared consistently across feature maps, feature

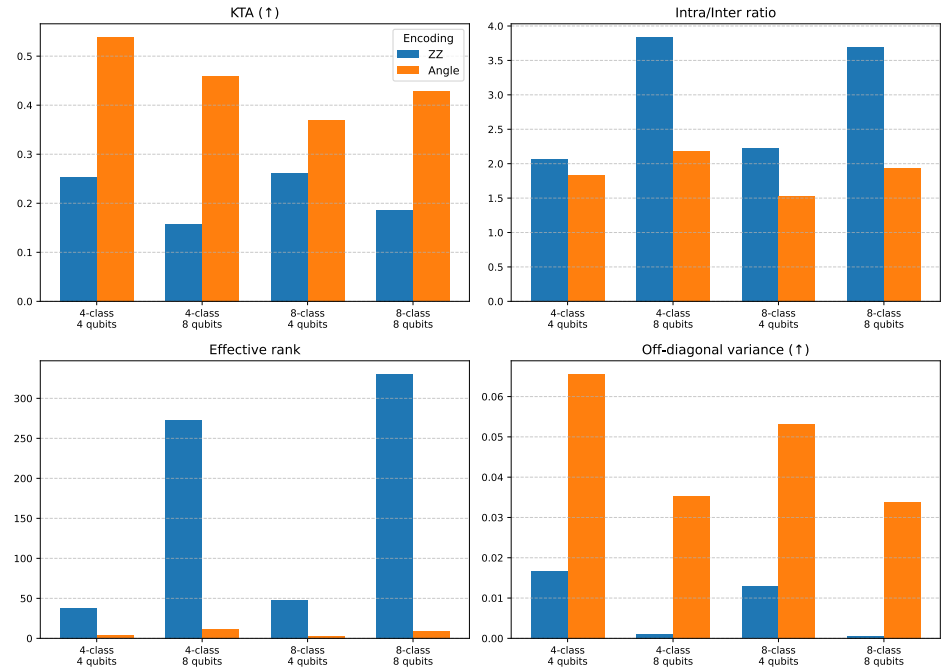


Figure 2: Kernel Geometry Metrics Chart

dimensions, and sample sizes. Although the data splits and model initializations vary across seeds, the seed set itself is fixed throughout the study.

Performance is measured primarily using classification accuracy, with runtime and computational feasibility serving as secondary indicators of practical scalability. Quantum experiments are performed using noiseless statevector simulation, which isolates algorithmic behavior from hardware noise and allows direct comparison between the tested feature maps.

The experimental design is intended to compare kernel behavior under controlled simulation conditions rather than to claim hardware-level quantum advantage. By using the same task definitions, preprocessing pipeline, qubit counts, and evaluation protocol across all experiments, the study isolates the effects of feature encoding on kernel geometry and classification performance.

In addition to accuracy, we report the kernel geometry metrics defined above to explain the observed performance differences between feature encodings.

4 Experiments and Results

4.1 Kernel Geometry Reveals Concentration

The central question of this work is whether different feature encodings produce fundamentally different kernel geometries, and whether those geometries explain classification performance. We directly analyze kernel matrices using the four metrics defined in Section 3. Table A1 and Figure 2 report kernel target alignment, intra/inter similarity ratio, effective rank, and off-diagonal variance for each configuration (400 training samples, seed 42). Higher KTA, higher intra/inter ratio, and higher off-diagonal variance indicate more structured, discriminative kernels. For effective rank, very low values indicate collapse, while very high values indicate diffuse spreading.

Angle encoding maintains high KTA (≥ 0.37), high off-diagonal variance (≥ 0.033), and moderately low effective rank (≤ 12) across all tasks. In contrast, ZZFeatureMap on 8-qubit

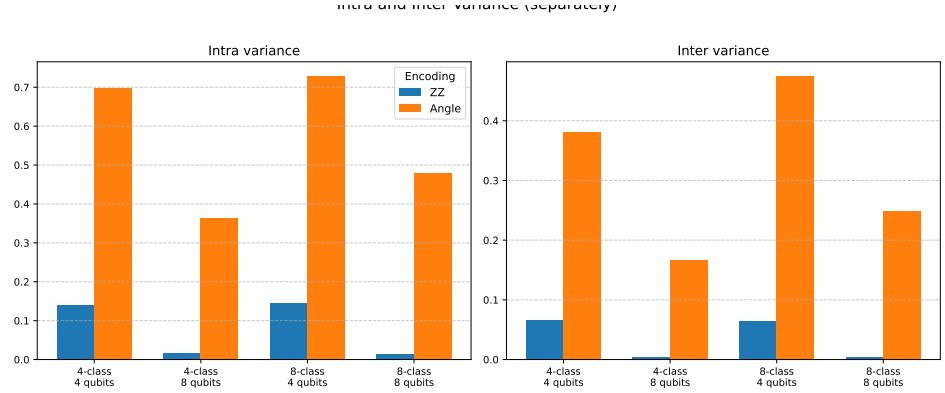
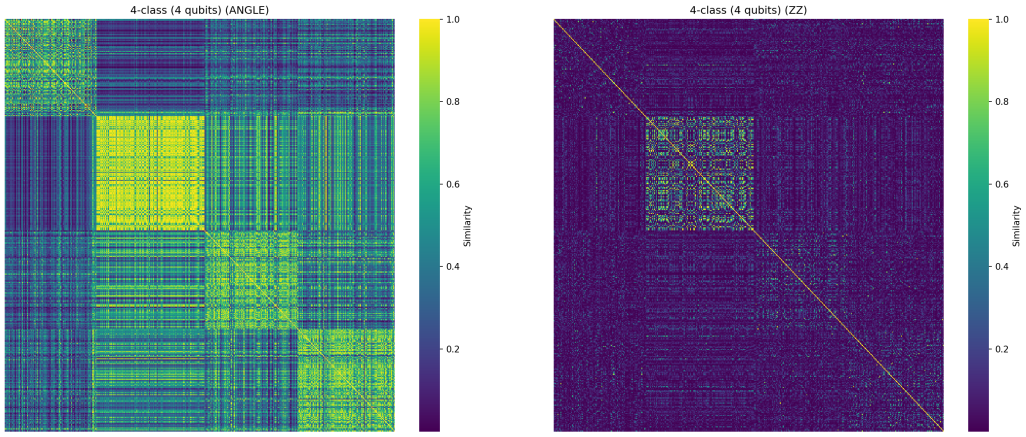


Figure 3: Intra and Inter Absolute Values



(a) Angle encoding: clear block-diagonal structure.

(b) ZZFeatureMap: nearly uniform (concentration).

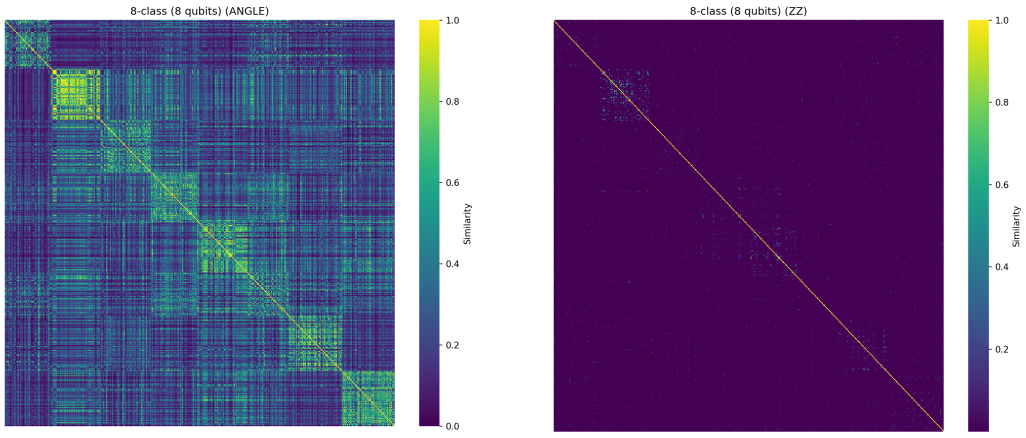
Figure 4: Kernel heatmaps for 4-class, 4 qubits.

tasks exhibits low KTA (≤ 0.19), extremely low off-diagonal variance (0.0011 for 4-class, 0.0005 for 8-class), and very high effective rank (> 270). These values are direct evidence of kernel concentration. The similarity matrix becomes nearly uniform, losing discriminative structure. The intra/inter ratio alone can be misleading (ZZFeatureMap shows a ratio > 3.6 on 8-qubit tasks), but the absolute similarities are near zero (intra ≈ 0.015), confirming that the high ratio is an artifact of near-zero values, as shown in Figure 3.

Figure 4 and Figure 5 show kernel heatmaps for the 4-class and 8-class classification tasks with corresponding qubit counts. Each block corresponds to samples from a single class. Within a block, high similarity indicates strong intra-class consistency, while low similarity between blocks indicates good class separation. As shown in the figures, angle encoding produces a clear block-diagonal pattern, indicating clear class boundaries, while ZZFeatureMap produces a nearly uniform heatmap, suggesting that entanglement has reduced class separability.

4.2 Accuracy Follows Geometry

If kernel geometry determines learnability, classification accuracy should correlate with the metrics. Table 1 reports mean test accuracy over 5 seeds for the quantum kernel SVM. Angle encoding consistently achieves higher accuracy, and the gap widens with task complexity. On the hardest 8-class, 8-qubit task, angle encoding reaches 83.0%, while ZZFeatureMap achieves



(a) Angle encoding: clear block-diagonal structure.

(b) ZZFeatureMap: nearly uniform (concentration).

Figure 5: Kernel heatmaps for 8-class, 8 qubits.

only 41.0%. This mirrors the geometric metrics directly: higher KTA, higher off-diagonal variance, and moderate effective rank all predict better classification.

Table 1: Mean test accuracy (%) for QSVC (5 seeds, n = 400)

Task	Qubits	ZZFeatureMap	Angle Encoding
4-class	4	74.0	91.0
4-class	8	58.0	96.0
8-class	4	47.0	67.0
8-class	8	41.0	**83.0**

As shown in Table 2, the F-1 score tells a similar story: angle encoding consistently outperforms ZZFeatureMap. Angle encoding also benefits from increased qubit and PCA feature count, whereas ZZFeatureMap suffers performance degradation as qubit and feature count increase, the same pattern visible in Table 1.

Table 2: Mean F-1 Score for QSVC across all sample sizes (5 seeds, n = 100, 250, 400)

Task	Qubits	ZZFeatureMap	Angle Encoding
4-class	4	69.7	90.0
4-class	8	53.1	94.8
8-class	4	35.7	58.8
8-class	8	29.0	**77.8**

The improvement from 4 to 8 features under angle encoding contrasts with the degradation under ZZFeatureMap. Rather than becoming less useful as the feature dimension increases, the angle-encoded QSVC benefits from additional PCA features. This indicates that the issue is not inherent to the QSVC model but tied to the geometry induced by the chosen feature map.

5 Conclusion

This paper evaluated how feature map geometry affects quantum kernel performance on MNIST-derived classification tasks. The results show that the quantum kernel approach is strongly affected by the chosen feature map. ZZFeatureMap-based QSVC degraded as

task complexity increased due to kernel concentration, while angle encoding substantially improved QSVC performance on the 8-class tasks. This suggests that, as Huang et al. predicted, adding more qubits to the kernel results in worse performance Huang et al. (2021).

These findings align with prior theory suggesting that practical QML performance depends not only on circuit expressivity, but also on whether the encoding induces a useful feature-space geometry for the data Huang et al. (2021). Future work should explore mitigation strategies for kernel concentration and test these diagnostic metrics in other data domains and with alternative feature encodings.

6 Impact Statement

This work contributes to the empirical understanding of near-term quantum machine learning by demonstrating that feature map geometry, not just circuit expressivity, determines practical performance. The primary impact is scientific: the results show that entangling feature maps like ZZFeatureMap can suffer from kernel concentration on multiclass tasks, while simple angle encoding preserves discriminative structure. These findings caution against assuming that greater entanglement or expressivity automatically improves learning. By reporting encoding-dependent improvements alongside direct evidence of kernel concentration, this work supports more realistic expectations for NISQ-era quantum learning research.

7 Limitations

Several limitations of this study should be acknowledged. First, all experiments were performed using noiseless statevector simulation, which isolates algorithmic behavior from hardware noise. While this choice facilitates direct comparison between feature maps, it does not reflect the performance of these models on actual quantum hardware with noise and finite shots. Second, the study focuses on MNIST-derived image classification tasks using PCA-reduced features; the extent to which these findings generalize to other data types, dimensionalities, or domains remains an open question. Third, the kernel geometry metrics were computed on relatively small training sets ($n = 400$); future work should investigate whether these metrics scale reliably to larger datasets. Despite these limitations, the diagnostic framework introduced here provides a practical tool for evaluating feature maps before expensive multiclass training, and the observed correlation between geometry and accuracy offers a foundation for future work on mitigation strategies.

8 Future Work

Planned future work includes further exploration of additional settings, such as binary classification with larger training and testing sample sizes. We also plan to incorporate classical baselines and other QML methods, such as variational quantum classifiers (VQC), for direct comparison against the quantum kernel results presented here.

Beyond these planned extensions, the findings of this paper motivate several further directions. Validating whether the observed differences in kernel geometry persist under hardware noise and finite-shot sampling would clarify the relevance of these diagnostics for near-term quantum devices. Testing the four kernel geometry metrics on data domains beyond MNIST-derived images would help establish whether they generalize as general-purpose diagnostic tools. Finally, the kernel concentration observed for the ZZFeatureMap motivates work on concrete mitigation strategies, such as circuit modifications or regularization techniques, and on evaluating whether the geometry metrics introduced here can help guide such interventions.

References

Cortes, C., Mohri, M., and Rostamizadeh, A. Algorithms for learning kernels based on centered alignment. *J. Mach. Learn. Res.*, 13(1):795–828, March 2012. ISSN 1532-4435.

Havlíček, V., Córcoles, A. D., Temme, K., Harrow, A. W., Kandala, A., Chow, J. M., and Gambetta, J. M. Supervised learning with quantum-enhanced feature spaces. *Nature*, 567(7747):209–212, March 2019. ISSN 1476-4687. doi: 10.1038/s41586-019-0980-2. URL http://dx.doi.org/10.1038/s41586-019-0980-2.

Huang, H.-Y., Broughton, M., Mohseni, M., Babbush, R., Boixo, S., Neven, H., and McClean, J. R. Power of data in quantum machine learning. *Nature Communications*, 12(1), 2021. ISSN 2041-1723. doi: 10.1038/s41467-021-22539-9. URL http://dx.doi.org/10.1038/s41467-021-22539-9.

Kübler, J. M., Buchholz, S., and Schölkopf, B. The inductive bias of quantum kernels, 2021. URL https://arxiv.org/abs/2106.03747.

Liu, Y., Arunachalam, S., and Temme, K. A rigorous and robust quantum speed-up in supervised machine learning. *Nature Physics*, 17(9):1013–1017, July 2021. ISSN 1745-2481. doi: 10.1038/s41567-021-01287-z. URL http://dx.doi.org/10.1038/s41567-021-01287-z.

Preskill, J. Quantum computing in the nisq era and beyond. *Quantum*, 2:79, 2018.

Schuld, M. and Killoran, N. Quantum machine learning in feature hilbert spaces. *Physical Review Letters*, 122(4), February 2019. ISSN 1079-7114. doi: 10.1103/physrevlett.122.040504. URL http://dx.doi.org/10.1103/PhysRevLett.122.040504.

Thanasilp, S., Wang, S., Cerezo, M., and Holmes, Z. Exponential concentration in quantum kernel methods, 2024. URL https://arxiv.org/abs/2208.11060.

AI Statement

We have not used generative AI tools to Generate synthetic data sets, help develop theoretical models or conceptual frameworks, assist in the writing of proofs, propose or refine hypotheses, design or provide feedback on research methodology or experiments, implement methods, clean and reformat dataset, support qualitative and thematic data analysis, interpret results, and formulate mathematical claims, provide critical ingredients for proving mathematical claims, assist with translation are not applicable to this work.

Additionally, we used generative AI tools to edit the paper to improve readability and propose a title and keywords. All AI-suggested edits were reviewed by the authors, who accepted, modified, or rejected each suggestion individually; no wording was incorporated without author review. We take responsibility for the final content of this work, including text, claims, or artifacts produced with the aid of generative AI.

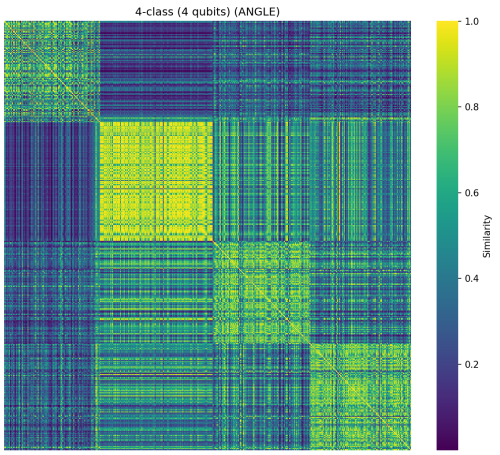
A Kernel Geometry Metrics table

Table A1: Kernel geometry metrics for 4-class and 8-class tasks (n=400). Higher is better for KTA, intra/inter ratio, and off diagonal variance. For effective rank, low to moderate is desirable.

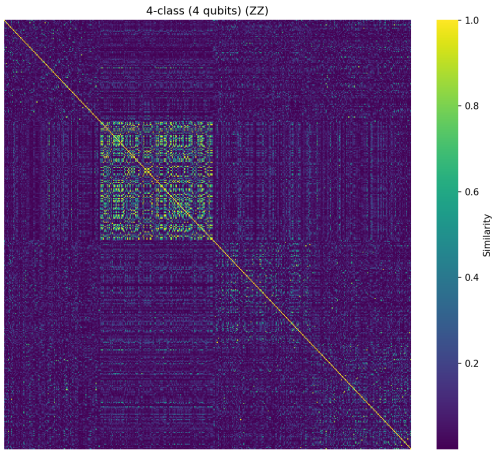
Task	Qubits	Encoding	KTA ↑	Intra/Inter	Eff. rank	Off-diag var ↑
4-class	4	ZZ	0.253	0.139/0.067 (2.06)	37.9	0.0167
4-class	4	Angle	**0.539**	0.697/0.381 (1.83)	**3.6**	**0.0656**
4-class	8	ZZ	0.157	0.016/0.004 (3.84)	272.5	0.00112
4-class	8	Angle	**0.459**	0.364/0.167 (2.18)	**11.9**	**0.0354**
8-class	4	ZZ	0.262	0.144/0.065 (2.23)	47.7	0.0129
8-class	4	Angle	**0.370**	0.729/0.475 (1.53)	**3.2**	**0.0532**
8-class	8	ZZ	0.187	0.015/0.004 (3.69)	330.4	**0.00050**
8-class	8	Angle	**0.428**	0.479/0.248 (1.93)	**8.9**	**0.0339**

B Kernel Heatmap for All Configurations

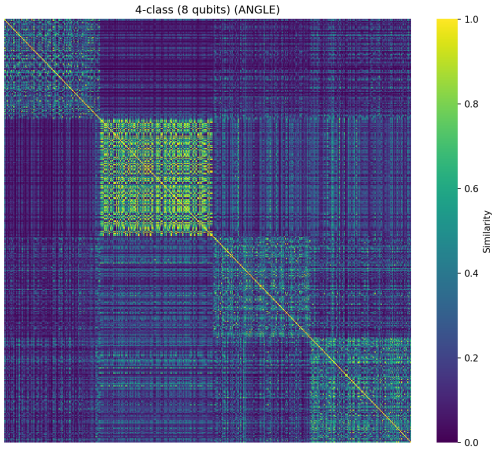
B.1 4-class, 4 qubits, angle encoding



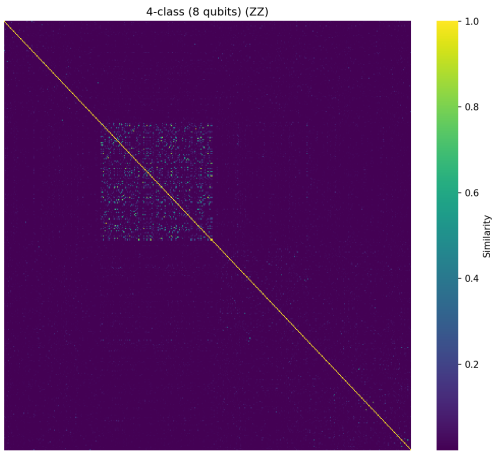
B.2 4-class, 4 qubits, ZZFeatureMap



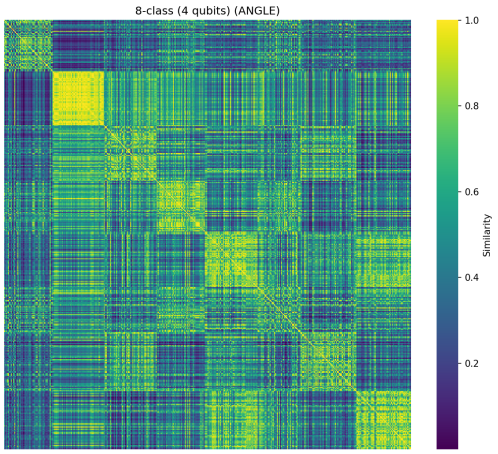
B.3 4-class, 8 qubits, angle encoding



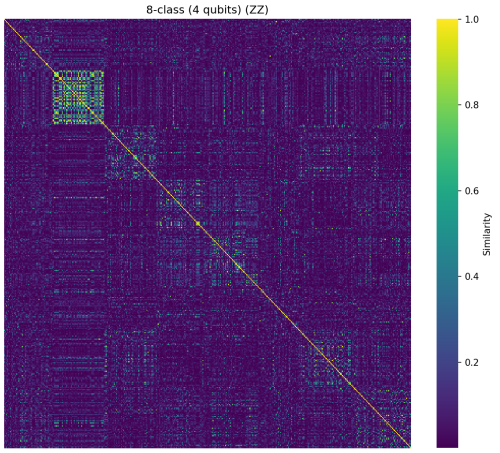
B.4 4-class, 8 qubits, ZZFeatureMap



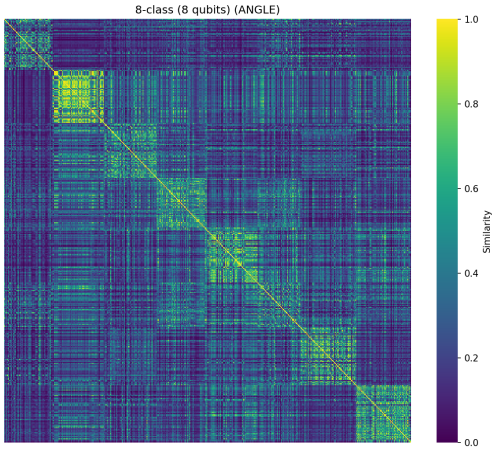
B.5 8-class, 4 qubits, angle encoding



B.6 8-class, 4 qubits, ZZFeatureMap



B.7 8-class, 8 qubits, angle encoding



B.8 8-class, 8 qubits, ZZFeatureMap

