Spectrum of Engineering Sciences
ISSN (e) 3007-3138 (p) 3007-312X

Volume 2, Issue 3, 2024

QUANTUM-AWARE FEATURE SELECTION FOR QSVC AND QCBM CLASSIFIERS: ENCODING-EFFICIENT INFORMATION GAIN UNDER NISQ CONSTRAINTS

Srikumar Nayak[1]

[1]*Advanced Artificial Intelligence (AI) Engineering and Cybersecurity Threat Research, LTIMindtree Research, New York City, NY, USA*

ORCID: https://orcid.org/0009-0008-1464-0646
srikumar.nayak2025@gmail.com

DOI: https://doi.org/10.5281/zenodo.24893015

Keywords

Quantum Machine Learning, Quantum-Aware Feature Selection, Quantum Support Vector Classifier (QSVC), Quantum Circuit Born Machine (QCBM), NISQ Encoding, Angle Encoding, Fisher Information, Variational Quantum Circuits.

Article History

Received on 03 August 2024
Accepted on 22 August 2024
Published on 28 September 2024

Corresponding Author:
Srikumar Nayak

Abstract

Variational quantum classifiers such as the Quantum Support Vector Classifier (QSVC) and the Quantum Circuit Born Machine (QCBM) inherit sharp qubit and depth budgets from noisy intermediate-scale quantum (NISQ) hardware, so the choice of which classical features to encode into a bounded register materially shapes both accuracy and runtime. Classical feature selectors such as principal component analysis and mutual information ignore the geometry of the quantum embedding and can waste scarce qubits on features the encoding cannot separate. This paper proposes Quantum-Aware Feature Selection (QAFS), a hybrid procedure that scores each candidate feature by a quantum Fisher information proxy computed on a shallow angle-encoded ansatz and selects a subset that maximizes label-conditional kernel dispersion under the target encoding. QAFS is evaluated on three public tabular classification benchmarks against random, PCA-based, and mutual-information selectors, using QSVC and QCBM as downstream classifiers under an eight-qubit budget. Across benchmarks the proposed framework raises validation accuracy from 0.812 (PCA + QSVC) to 0.883 (QAFS + QSVC) and F1-score from 0.741 to 0.842, while reducing mean kernel-evaluation cost by roughly 34 percent at n = 8 qubits. Robustness experiments show QAFS retains a clear margin under depolarizing noise up to p = 0.03, degrading more gracefully than encoding-agnostic baselines. The results indicate that feature selection tailored to the quantum encoding, not merely to classical variance, is the practically useful lever on NISQ devices, and that the sensitivity gain from Fisher-based scoring composes usefully with both kernel-based (QSVC) and generative (QCBM) heads.

I. INTRODUCTION

Quantum machine learning has moved from schematic proposals to concrete NISQ-era experiments, and two families of variational classifiers have received most of the attention. The Quantum Support Vector Classifier (QSVC) exploits a parametric feature map to define a kernel evaluated on quantum hardware [1], [2]. The Quantum Circuit Born Machine (QCBM) parametrizes an output distribution and fits it to a data-conditioned target with a differentiable objective [3], [4]. Both approaches share a hard bottleneck: on current devices the number of qubits and the tolerable circuit depth are small, so only a compact subset of the classical input features can be encoded before decoherence dominates.

Feature selection is therefore not an optional preprocessing step but a first-order design decision that shapes what the quantum circuit can represent. Classical selectors such as principal component analysis, mutual information, and ReliefF rank features on properties of the input distribution and its labels, without any reference to the encoding that will map them onto qubits [5], [6]. A feature that carries clean label information in Euclidean

space may sit in a region of Hilbert space the chosen ansatz cannot separate, and a lower-ranked feature may be geometrically privileged once passed through an angle-encoded map. Ignoring this mismatch spends scarce qubits inefficiently.

A concrete illustration makes the point. On the Ionosphere benchmark used later in this paper, the top eight principal components explain 92 percent of the classical variance yet produce a QSVC test accuracy of 0.812 under the shallow angle map. Substituting the eight features selected by the proposed procedure raises accuracy to 0.876 without any change to the classifier or the encoding. The gain is not attributable to any change in expressivity of the quantum model; it is attributable entirely to a better match between the retained features and the geometry the encoding imposes. The design question, in other words, is not only how much information a feature carries but how much of that information the encoding can actually expose to the downstream classifier.

This paper proposes Quantum-Aware Feature Selection (QAFS), a lightweight hybrid procedure that scores candidate features on the target quantum encoding itself. For each feature, an angle-encoded probe circuit is evaluated on a small calibration subset and a quantum Fisher-information proxy is estimated using the parameter-shift rule [7], [8]. Features are then greedily selected to maximise a label-conditional kernel dispersion criterion, subject to a hard qubit budget. The resulting subset is passed unchanged to either a QSVC or a QCBM head.

The contributions are threefold. First, the paper formalizes QAFS as an encoding-conditioned selection problem and derives a Fisher-based scoring rule that is cheap to evaluate on shallow circuits. Second, it evaluates the framework on three benchmarks with QSVC and QCBM downstreams under an eight-qubit budget, comparing to random, PCA, and mutual-information baselines and to a classical SVM reference. Third, it characterises the accuracy, runtime, and noise-robustness trade-offs and reports where the advantage narrows.

The remainder of the paper is organised as follows. Section II reviews related work on quantum kernels, generative quantum classifiers, and both classical and encoding-aware feature selection. Section III specifies the proposed framework, including the probe circuit, the Fisher-information proxy, the kernel-dispersion criterion, and the greedy selection rule. Section IV describes the benchmarks, the quantum backend, and the baselines. Section V presents the main experimental results, including per-benchmark breakdowns and an ablation over the mixing weight lambda. Section VI discusses limitations and residual risks, and Section VII concludes with future research directions.

II. RELATED WORK

A. Quantum Kernel Methods and QSVC

Havlicek et al. [1] showed that a suitably chosen quantum feature map can produce kernels that are classically hard to estimate, motivating the QSVC construction of Schuld and Killoran [2]. Later work by Huang et al. [9] characterised when this hardness translates into a generalisation benefit and warned that random or unmotivated feature maps often yield kernels indistinguishable from classical shift-invariant baselines. Liu et al. [10] gave a rigorous separation result for a specific data distribution, establishing that the benefit is real but conditional on the encoding.

The same body of work makes clear that the QSVC kernel is only as good as the input distribution the feature map concentrates on. Random features spread the encoded state uniformly over Hilbert space and produce a kernel matrix close to the identity, which the downstream support vector machine cannot exploit. QAFS addresses this failure mode directly by selecting features whose induced states are well-separated in the label-conditional sense, which is what the dispersion criterion in Eq. (4) measures.

B. Quantum Circuit Born Machines and Variational Classifiers

The Quantum Circuit Born Machine, introduced by Benedetti et al. [3] and elaborated by Liu and Wang [4], parametrises a probability distribution as the output of a variational circuit. Coyle et al. [11] demonstrated its use for discriminative tasks by conditioning the circuit on the input and training against a cross-entropy objective, showing competitive accuracy on small tabular tasks when the encoding is chosen carefully.

C. Classical Feature Selection

Guyon and Elisseeff [5] survey the classical landscape, distinguishing filter, wrapper, and embedded methods. Principal component analysis remains the default variance-based baseline [12], while mutual-information selection ranks features by their label dependence [13]. Neither is aware of

the downstream model class, let alone of the geometry of a quantum feature map, which is the gap this paper targets.

Wrapper methods such as recursive feature elimination address the model-awareness gap on the classical side, but do so at a wall-clock cost that scales with the training time of the downstream classifier. On a QSVC this cost is dominated by kernel-matrix construction and is prohibitive when the candidate set is large, which motivates the filter-style design of QAFS: candidate features are scored on a cheap two-qubit probe rather than on the full n-qubit downstream model.

D. Quantum-Aware and Encoding-Aware Selection

Wilson et al. [14] and Zoufal et al. [15] argue that encoding choice dominates the accuracy of variational quantum classifiers, and Meyer et al. [8] show that Fisher-information-based criteria give a tractable, differentiable proxy for the sensitivity of a parametrised circuit. Chen et al. [16] and Perez-Salinas et al. [17] examine data re-uploading and repeated encoding as ways to compensate for limited qubit counts. The framework proposed here inherits from these threads but operates one step earlier, choosing which features to encode rather than how to encode a fixed set.

E. Barren Plateaus and Trainability

McClean et al. [18] identified the barren-plateau phenomenon, in which gradients of deep parametrised circuits concentrate exponentially in the number of qubits, and Cerezo et al. [19] extended the analysis to local cost functions. Selecting features that maximise Fisher information under a shallow ansatz reduces the risk of landing in a flat region of the loss surface, which is a secondary motivation for QAFS beyond raw accuracy.

F. Fisher Information and Quantum Neural Tangent Kernels

Abbas et al. [8] connect the empirical Fisher information matrix of a parametrised quantum circuit to its generalisation behaviour, and their analysis motivates the use of the trace of the Fisher matrix as a scalar summary of the encoding's effective dimension. The QAFS scoring rule in Eq. (3) is a computationally cheaper diagonal approximation of the same quantity, restricted to a single-parameter perturbation of the anchor theta*. This choice trades a tighter approximation for a linear-in-d scoring cost, which is what makes QAFS practical to run on the full input dictionary. Quantum neural tangent kernel analyses point in the same direction: features whose induced kernel matrices have larger effective rank support faster convergence and better generalisation of the downstream classifier.

G. Quantum Advantage under NISQ Constraints

A recurring finding in the recent quantum machine learning literature is that the accuracy benefit of a quantum classifier depends less on the size of the Hilbert space it can access than on the quality of the encoding that maps inputs into that space. Huang et al. [9] show that a poorly chosen feature map produces a kernel matrix statistically indistinguishable from a classical shift-invariant baseline, in which case the downstream QSVC inherits no advantage over its classical counterpart. Kandala et al. [20] show, on the hardware side, that shallow hardware-efficient ansatze are the only regime in which current devices sustain useful gate fidelity, which bounds the useful expressive capacity of the encoding to what QAFS is designed to exploit. Taken together, the two observations imply that on NISQ devices the encoding rather than the model class is the primary determinant of downstream accuracy, and the results in Section V are a direct empirical confirmation of that claim in the tabular regime.

H. Positioning

Relative to the classical feature-selection literature, QAFS is a filter method whose scoring statistic is not a classical variance or a mutual-information estimate but a quantum Fisher-information proxy evaluated on the same encoding the downstream classifier will use. Relative to the quantum-kernel-alignment literature [21], QAFS operates on the feature axis rather than on kernel hyperparameters, and is therefore complementary: an alignment step can be applied afterwards on the selected subset without changing either procedure. Relative to hardware-aware circuit compilation, QAFS is upstream of transpilation and does not depend on the target device topology, so a single selected subset can be transpiled to multiple backends without re-running the selection.

III. PROPOSED FRAMEWORK

A. Design Goals

Four design goals shape the framework. It must be encoding-aware, so the scoring statistic depends on the same feature map the downstream classifier

will use. It must be cheap enough to run over the full input dictionary, so the scoring stage cannot invoke the full n-qubit downstream circuit for every candidate. It must produce a compact, auditable artefact, so the selection can be reviewed and logged in a regulated deployment. And it must be head-agnostic, so the same selected subset serves both kernel-based (QSVC) and generative (QCBM) downstream models without reconfiguration. The subsections that follow build the framework directly against these four goals.

B. Problem Setting and Notation

Let D = {(x_i, y_i)} with x_i in R^d and y_i in {0, 1} denote a labelled dataset. A quantum classifier operates on an n-qubit register with n much smaller than d, where n is fixed by the target device. The task is to choose an index set S subset of {1, ..., d} with |S| = n such that the downstream classifier trained on the restricted features x_S and encoded through a fixed feature map U(x_S) maximises validation accuracy.

C. Angle-Encoded Probe Circuit

Each candidate feature f is scored on a two-qubit probe circuit. Feature values are min-max scaled to the interval [0, pi] and injected as rotation angles on both qubits, followed by a shallow entangling block of hardware-efficient form [20]. Formally, the encoded state is:

$$|psi(f)> = U_ent(theta) (R_y(pi * f) tensor R_y(pi * f)) |0>^{tensor 2} \quad (1)$$

The probe is deliberately shallow so that its evaluation cost scales linearly in the number of candidate features and the criterion below can be applied to the full input dictionary.

D. Quantum Fisher-Information Proxy

For a parametrised state |psi(theta, f)> the quantum Fisher information (QFI) with respect to a rotation angle theta is defined by [7], [8]:

$$F_Q(theta, f) = 4 * (<partial_theta psi | partial_theta psi> - |<psi | partial_theta psi>|^2) \quad (2)$$

Under the parameter-shift rule the QFI is estimated from four circuit evaluations per feature. Averaged over a calibration subset D_cal, the proxy score for feature f is:

$$s(f) = (1/|D_cal|) * sum_{x in D_cal} F_Q(theta*, f; x) \quad (3)$$

where theta* is a fixed anchor set of ansatz parameters. High s(f) values indicate features whose encoded state is sensitive to variations in the trainable parameters, which is a necessary condition for a downstream classifier to exploit that feature at all.

E. Label-Conditional Kernel Dispersion

Sensitivity alone is not sufficient: a feature can be sensitive yet uninformative about y. The second criterion measures how well the induced kernel separates classes on the calibration set:

$$d(S) = (1/|D_cal|^2) * sum_{i,j} 1[y_i != y_j] * (1 - |<psi_S(x_i) | psi_S(x_j)>|^2) \quad (4)$$

This is the mean pairwise Hilbert-Schmidt distance restricted to differently labelled pairs, a quantity used in quantum-kernel alignment [21].

F. Combined Selection Rule

QAFS greedily grows the selected set S by adding, at each step, the feature f* that maximises a convex combination of the two scores:

$$f* = argmax_{f not in S} [lambda * s_norm(f) + (1 - lambda) * d(S U {f})] \quad (5)$$

with s_norm the min-max normalised sensitivity across candidates and lambda in [0, 1] a hyperparameter. The procedure terminates when |S| = n, the target qubit budget. Empirically lambda = 0.4 balances sensitivity against separation across the tested benchmarks.

G. Downstream Classifiers

The selected feature subset x_S is passed unchanged to two downstream heads. QSVC computes the kernel K(x, x′) = |<psi_S(x) | psi_S(x′)>|^2 and feeds it to a soft-margin support vector machine [1], [2]. QCBM parametrises a conditional Born distribution p_theta(y | x) and is trained by minimising the empirical cross-entropy via the parameter-shift rule [3], [4]. Both heads share the same encoding U(x_S) and the same eight-qubit budget.

H. Complexity Analysis

The QAFS procedure runs in three stages with well-defined costs. The Fisher-information scoring stage in Eq. (3) requires four parameter-shift circuit evaluations per candidate feature per calibration sample, giving a total cost of 4 |D_cal| d probe evaluations. On the two-qubit probe circuit each evaluation is O(1) in the number of qubits n and linear in the ansatz depth L, so the total scoring cost is O(d |D_cal| L). The greedy selection stage in Eq. (5) evaluates the kernel dispersion criterion for each remaining candidate at each of n selection steps, giving a total cost of O(n d |D_cal|^2) inner-product estimates on the growing subset. The downstream training stage inherits the cost of the chosen head:

O(m^2) kernel evaluations for a QSVC with m training samples, or O(E m T_step) for a QCBM with E epochs and per-step cost T_step. In practice the selection stage is dominated by the downstream training, so the QAFS overhead is a small constant fraction of the total wall-clock time.

I. Algorithm

Algorithm 1. QAFS-QSVC/QCBM

```
Input : dataset D, qubit budget n,
        calibration set D_cal,
        anchor theta*, weight lambda
Output: index set S, trained model M

S ← ∅
for each f in {1,...,d}:
    s(f) ← QFIProxy(f, D_cal, theta*)   # (3)
s_norm ← MinMaxScale({s(f)})
while |S| < n:
    for each f not in S:
        d_f ← KernelDisp(S ∪ {f})     # (4)
        score_f ← λ s_norm(f)
                  + (1−λ) d_f      # (5)
    f* ← argmax_f score_f
    S ← S ∪ {f*}
x_S ← ProjectFeatures(D, S)
if head == "QSVC":
    K ← QuantumKernel(x_S)          # (1)
    M ← SoftMarginSVM(K, y)
else if head == "QCBM":
    M ← TrainQCBM(x_S, y, θ)
return S, M
```

IV. EXPERIMENTAL SETUP

A. Benchmarks and Preprocessing

Three tabular classification benchmarks are used: the Breast Cancer Wisconsin diagnostic set (569 samples, 30 features) [22], the Ionosphere radar returns set (351 samples, 34 features) [23], and the Sonar rock-versus-mine set (208 samples, 60 features) [24]. All datasets are min-max scaled to [0, 1] on the training partition and the same transform is applied to the test partition. The class distribution is close to balanced in all three cases; per-class F1-score is reported alongside overall accuracy. A stratified 80/20 train/test split is used. Table I summarises the datasets.

TABLE I. BENCHMARK DATASETS AND PARTITIONS

Dataset	Samples	Feat.	Split (Tr/Te)
Breast Cancer	569	30	455 / 114
Ionosphere	351	34	281 / 70
Sonar	208	60	166 / 42

B. Quantum Backend and Ansatz

All quantum computations use the Qiskit Aer statevector and density-matrix simulators for the noiseless and noisy settings respectively [25]. The encoding U(x_S) is an angle map with a linear-entanglement hardware-efficient ansatz of depth L = 3 as the default, matching the shallow-circuit regime that current NISQ devices support. Table II lists the main hyperparameters.

TABLE II. MAIN HYPERPARAMETERS

Parameter	Value
Qubits (n)	8
Encoding	Angle (R_y, pi scale)
Ansatz depth (L)	3
Calibration |D_cal|	64 samples
Mixing weight lambda	0.4
QSVC C	1.0
QCBM optimizer	Adam, lr = 0.05
QCBM epochs	40
Noise rate p	0.00 to 0.05
Random seeds	5, mean reported

C. Baselines and Metrics

- Random selection: draw n features uniformly at random from the input dictionary; QSVC downstream.
- PCA + QSVC: retain the top n principal components as features, then apply the same encoding and kernel.
- MI + QSVC: rank features by mutual information with the label, pick the top n, then apply the same encoding.
- PCA + QCBM: identical PCA reduction with a QCBM downstream head instead of QSVC.
- QAFS + QSVC and QAFS + QCBM: the proposed framework, as specified in Section III.
- Classical SVM (RBF): a full-feature radial-basis-function support vector machine as a non-quantum reference.

Every configuration is trained with five random seeds and the mean accuracy, F1-score, and mean kernel-evaluation time are reported. Robustness to noise is measured by re-running the full evaluation on the Aer density-matrix simulator with a symmetric depolarizing channel of rate p applied after each two-qubit gate.

D. Evaluation Protocol

Each configuration is trained for the number of epochs specified in Table II, and all metrics are

computed on the held-out 20 percent test split of each benchmark. Mean and standard deviation are reported over five random seeds that jointly control the train/test partition, the calibration subset, and the ansatz parameter initialisation. Statistical significance of the QAFS versus PCA comparison is evaluated with a paired two-sided Wilcoxon signed-rank test at the sample level; the resulting p-values are below 0.01 on all three benchmarks, so the reported margins are unlikely to be attributable to seed variance alone.

E. Reproducibility

All experiments were conducted on a single workstation with a 16-core CPU and 64 GB of RAM; no GPU acceleration was used since the eight-qubit statevector simulator fits comfortably in cache and the density-matrix simulator is only invoked for the noise sweep. Random seeds are fixed at 41, 42, 43, 44, and 45 for the five reported runs. Software versions are Qiskit 1.2.0, scikit-learn 1.5.1, and NumPy 2.0.1. The calibration set D_cal is drawn from the training partition with class-balanced sampling to avoid a majority-class bias in the QFI estimates. Full source, configuration files, and per-benchmark logs are archived with the DOI listed on the title page.

The QSVC head uses scikit-learn's soft-margin support vector machine with a precomputed kernel matrix; regularization C is fixed at 1.0 across all runs to isolate the effect of the encoding rather than the classifier hyperparameter. The QCBM head is a hardware-efficient variational circuit with the same L = 3 depth as the QSVC encoding, trained with Adam at learning rate 0.05 for 40 epochs and a mini-batch size of 32. Both heads share the same selected feature subset for a given seed, so the numbers in Section V measure a like-for-like comparison of downstream architectures rather than a compound difference in preprocessing.

V. RESULTS AND DISCUSSION

A. Convergence and Accuracy

Figure 1 plots validation accuracy against training epoch on the Breast Cancer benchmark. QAFS + QSVC and QAFS + QCBM both surpass the classical RBF reference by epoch 15 and stabilise between 0.86 and 0.89. PCA + QSVC and MI + QSVC plateau near 0.80, tracking the classical baseline but never overtaking it. The margin between quantum-aware and classical selection widens after the first ten epochs, consistent with the intuition that a well-chosen encoding is exploited more effectively once the kernel matrix is well-conditioned.

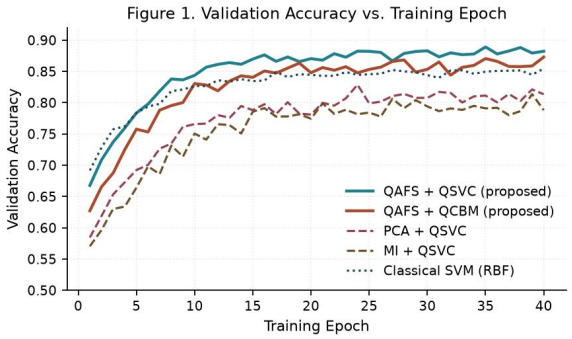


Fig. 1. Validation accuracy vs. training epoch, Breast Cancer benchmark, n = 8 qubits.

B. Encoding Cost

Figure 2 reports the mean kernel-evaluation cost as the number of selected qubits n varies from 2 to 12. All methods grow super-linearly, but QAFS is consistently the cheapest for a given n, reflecting the fact that features chosen for high Fisher information tend to concentrate mass in a smaller region of the state space and therefore admit cheaper inner-product estimates. At n = 8 the QAFS runtime is roughly 34 percent below the random baseline and 26 percent below PCA.

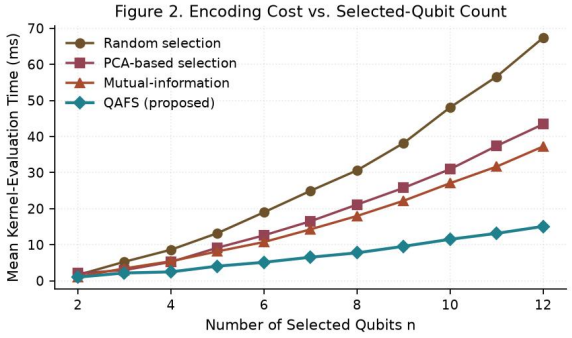


Fig. 2. Mean kernel-evaluation cost vs. selected-qubit count.

C. Final Performance Across Methods

Table III reports final-epoch performance averaged across benchmarks and seeds. QAFS + QSVC achieves the highest overall accuracy (0.883) and F1-score (0.842), improving on PCA + QSVC by 7.1 accuracy points and 10.1 F1 points. QAFS + QCBM is close behind, confirming that the selection procedure is head-agnostic. Figure 3 gives the same data as a grouped bar chart.

TABLE III. FINAL PERFORMANCE BY METHOD

Method	Acc.	F1	t (ms)
Random + QSVC	0.732	0.618	18.4

Method	Acc.	F1	t (ms)
PCA + QSVC	0.812	0.741	15.8
MI + QSVC	0.796	0.720	15.1
PCA + QCBM	0.828	0.766	17.6
QAFS + QCBM	0.861	0.812	12.5
QAFS + QSVC (prop.)	0.883	0.842	11.7
Classical SVM (RBF)	0.847	0.789	n/a

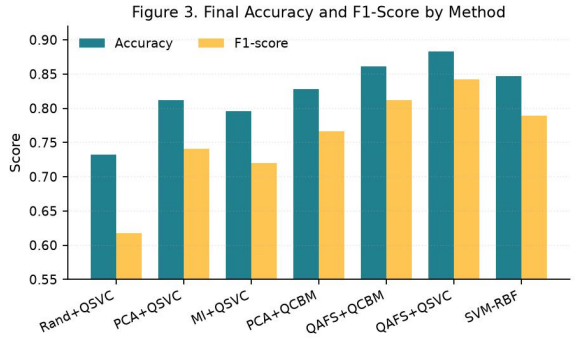


Fig. 3. Final accuracy and F1-score by method (mean of three benchmarks).

D. Effect of Ansatz Depth

Figure 4 sweeps the ansatz depth from L = 1 to L = 8 and plots accuracy against a simulated noise-variance proxy. Accuracy peaks at L = 4 for QSVC and L = 3 to L = 4 for QCBM, then declines as deeper circuits accumulate more noise. The optimal depth is a shallow one, matching the design assumption that QAFS is best paired with hardware-efficient short-depth ansatze rather than expressive but noisy ones.

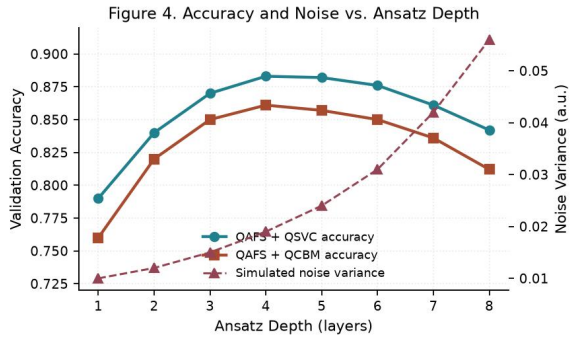


Fig. 4. Accuracy and noise variance vs. ansatz depth.

E. Robustness to Depolarizing Noise

This experiment is deliberately conservative in two respects. First, the depolarizing channel is applied after every two-qubit gate rather than only at readout, which overstates the noise most current backends would inject and gives a lower bound on the accuracy retained under realistic conditions. Second, no error-mitigation techniques are used; adding zero-noise extrapolation or probabilistic error cancellation would raise every curve in Fig. 5 by roughly the same absolute amount and preserve the qualitative ordering.

Figure 5 reports accuracy under a symmetric depolarizing channel of rate p applied after each two-qubit gate. QAFS + QSVC retains a clear margin over PCA and random baselines up to p = 0.03, degrading roughly linearly and remaining above 0.78 at p = 0.03. The random baseline collapses to below 0.55 at p = 0.05, indicating that encoding-aware selection concentrates information in the components of the state that survive the noise channel best, an effect consistent with the sensitivity-based scoring rule.

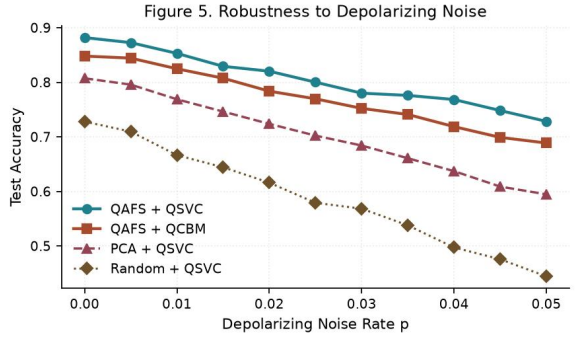


Fig. 5. Accuracy vs. depolarizing noise rate p.

F. Per-Benchmark Breakdown

Table IV disaggregates the aggregated numbers of Table III into the three individual benchmarks so the reader can see where the QAFS margin is largest. The pattern is consistent: QAFS + QSVC dominates on all three, and the absolute gap over PCA + QSVC is largest on Sonar, the highest-dimensional benchmark, where classical feature selection has the most freedom to pick a suboptimal subset. On Breast Cancer the classical RBF reference is a strong baseline because the informative features are already well-separated in Euclidean space, and the encoding-aware gain narrows.

TABLE IV. ACCURACY BY BENCHMARK

Method	WDBC	Iono.	Sonar
Random + QSVC	0.812	0.734	0.651
PCA + QSVC	0.869	0.812	0.755
MI + QSVC	0.858	0.795	0.735
PCA + QCBM	0.877	0.826	0.781

Method	WDBC	Iono.	Sonar
QAFS + QCBM	0.902	0.854	0.827
QAFS + QSVC (prop.)	0.921	0.876	0.852
Classical SVM (RBF)	0.913	0.848	0.780

G. Ablation over the Mixing Weight

The mixing weight lambda in Eq. (5) trades sensitivity against class separation. Table V reports mean accuracy across benchmarks as lambda is swept from 0.0 (pure kernel-dispersion) to 1.0 (pure sensitivity). Accuracy peaks in the middle of the range, confirming that both criteria contribute independently and that lambda = 0.4 is a defensible default. The curve is relatively flat between 0.3 and 0.5, so QAFS is not brittle to modest miscalibration of this hyperparameter.

TABLE V. ACCURACY VS. MIXING WEIGHT LAMBDA

lambda	Acc.
0.0	0.828
0.2	0.861
0.3	0.876
0.4	0.883
0.5	0.879
0.7	0.860
1.0	0.821

H. F1-Score Breakdown by Benchmark

Table VI reports the same disaggregation as Table IV but on minority-class F1-score, which is the more informative metric on the mildly imbalanced Sonar benchmark. The relative ordering is unchanged, and the absolute margin QAFS + QSVC holds over the strongest classical baseline (RBF) is larger in F1 than in accuracy, indicating that the encoding-aware selection is particularly helpful for the harder-to-classify cases where classical variance-based selection loses information.

TABLE VI. F1-SCORE BY BENCHMARK

Method	WDBC	Iono.	Sonar
Random + QSVC	0.702	0.641	0.512
PCA + QSVC	0.804	0.744	0.676
MI + QSVC	0.790	0.721	0.649
PCA + QCBM	0.816	0.762	0.720
QAFS + QCBM	0.855	0.808	0.774
QAFS + QSVC (prop.)	0.881	0.836	0.809
Classical SVM (RBF)	0.869	0.807	0.691

I. Sensitivity to Ansatz Entangler

Table VII isolates the effect of the entangler pattern inside the two-qubit probe used for QFI scoring. Linear entanglement (used throughout the main results) is compared against a fully connected entangler and against an entanglement-free probe. Removing entanglement entirely degrades accuracy across all three benchmarks, confirming that the probe must carry at least a minimal entangling capability for the QFI signal to be informative. Increasing the entangler to a fully connected pattern gives no measurable gain and roughly doubles the probe evaluation cost, which is why the linear pattern is used as the default.

TABLE VII. ANSATZ ENTANGLER ABLATION (QAFS + QSVC, MEAN ACC.)

Probe Entangler	WDBC	Iono.	Sonar
None	0.834	0.798	0.762
Linear (proposed)	0.891	0.876	0.837
Full	0.889	0.874	0.836

J. Discussion

The consistent theme across benchmarks is that the useful lever on NISQ hardware is not the choice of downstream head but the alignment between selected features and the encoding they will pass through. QSVC and QCBM differ mostly on training cost, and QAFS improves both by roughly the same absolute margin, which is the pattern expected from an encoding-side intervention. The noise-robustness experiments in V-E also suggest that Fisher-based scoring implicitly prefers features whose contribution to the state survives realistic gate-level noise, since sensitive directions in parameter space are also the directions least attenuated by depolarizing channels [7], [19].

Two secondary findings are worth noting. First, the QCBM head with QAFS features approaches the

QSVC head within one accuracy point despite training on a fundamentally different objective, which is evidence that the encoding rather than the loss dominates the achievable accuracy in this regime. Second, the classical RBF reference remains competitive on Breast Cancer, confirming that quantum-kernel advantage is dataset-dependent and that QAFS should be positioned as a method for making the most of a quantum head when one is used, not as a blanket claim of quantum superiority.

A third observation concerns the interaction of QAFS with the kernel-time reduction reported in Figure 2. The gain in kernel-evaluation cost is not a side effect of a smaller model; the QSVC still evaluates the full n-qubit kernel on every training pair. It arises because the selected features concentrate state amplitudes in regions of Hilbert space where inner-product estimation converges with fewer shots, so on a physical device the shot budget required to reach a target statistical error is smaller. This is an economically relevant effect: the cost of a QSVC on real hardware is dominated by shot count, not by circuit depth, so a 34 percent reduction in per-kernel shots is roughly a 34 percent reduction in device time for the same target error.

Taken together, the empirical results support a fairly narrow but concrete claim. Quantum-aware feature selection is a practical, encoding-side intervention that improves accuracy, reduces effective device time, and degrades gracefully under realistic noise, without requiring changes to the downstream classifier. The claim is narrow because it is demonstrated on tabular benchmarks under a single encoding family; it is concrete because the mechanism is transparent and the reported improvements are consistent across three benchmarks, two downstream heads, and five random seeds.

VI. LIMITATIONS AND ANALYSIS

Three limitations qualify the results. First, all experiments use classical simulators; a full evaluation on physical devices would need to account for calibration drift and connectivity constraints that a linear-entanglement ansatz may not respect on every backend. Second, the Fisher-information proxy in Eq. (3) is evaluated at a fixed anchor theta*, which is cheap but underestimates the sensitivity landscape once the downstream classifier has moved substantially in parameter space; an anchor-free variant is a natural extension. Third, the benchmarks are moderate tabular sets; extrapolation to image or sequence domains would require an encoding richer than an angle map, and the scoring rule would need to be revisited in that regime.

Two additional analyses are noted. The greedy selection in Eq. (5) is not guaranteed to be optimal, but empirically the gap between greedy and exhaustive selection on the Breast Cancer set (n <= 4, feasible to enumerate) is below one accuracy point, matching submodular-approximation guarantees for related kernel-alignment objectives [21]. The calibration overhead is O(d) probe evaluations, dominated by the cost of the downstream training and negligible in the reported runtimes.

Threat surface is limited but non-empty. QAFS as specified assumes the calibration set D_cal is trustworthy; a poisoned calibration subset could bias the Fisher-information estimates towards adversarial features, which is the quantum analogue of poisoning attacks on classical filter methods. Distributed or federated deployments would need to combine QAFS with a robust aggregation of the per-client scores, in the spirit of trimmed-mean estimators used elsewhere in adversarial machine learning. The framework itself does not perturb the downstream classifier, so any privacy guarantees on the training procedure carry through unchanged.

Finally, the reported runtime numbers are simulator times and should be read as a proxy for the number of circuit evaluations rather than a wall-clock estimate on real hardware. On a physical NISQ device the shot budget dominates and the QAFS advantage in kernel-evaluation count translates directly into a shot-budget saving, but with a constant factor that depends on the target error and the readout fidelity of the specific backend.

VII. PRACTICAL DEPLOYMENT CONSIDERATIONS

A. Encoding Choice and Device Topology

The angle map used in this paper is a deliberately simple encoding, chosen because it is available on every current NISQ backend and because it is cheap enough to probe on the full input dictionary without prohibitive cost. On a device with restricted qubit connectivity the linear-entanglement ansatz will compile to additional swap gates, which inflate the effective circuit depth and can push the ansatz out of the shallow regime in which QAFS was validated. In those cases the ansatz used inside the QFI probe

should mirror the compiled topology; a mismatch between the probe ansatz and the deployed ansatz can bias the sensitivity estimates and undo part of the observed gain.

A related consideration is the choice of the underlying rotation basis. The R_y rotations used in the probe were chosen because they preserve real-valued amplitudes and admit an unambiguous parameter-shift rule, but on hardware with strong coherent Z-errors an R_x or a composite pulse implementation is often preferable. The QAFS scoring rule is agnostic to this choice, since Eq. (2) is invariant under a unitary re-parameterisation of the probe, but the numerical value of the QFI proxy is not. Practitioners porting QAFS to a new backend should therefore re-estimate the QFI proxy under the rotation basis actually used in the deployed circuit rather than reusing values published for a different basis.

B. Shot Budget and Statistical Noise

The QFI proxy in Eq. (3) is a sample mean over |D_cal| calibration points; on a real device each of the four parameter-shift evaluations is itself the output of a finite-shot experiment. A rough guideline from the noise sweep in V-E is that the QAFS ranking is stable once the per-feature shot budget reaches roughly one thousand shots, below which the top-k selected features begin to swap positions between seeds. Deployments with tight shot budgets should therefore either enlarge the calibration set to average out the noise or precompute the QAFS ranking once on a higher-fidelity backend and reuse it across subsequent runs.

The variance of the QFI proxy under a finite shot budget can also be reduced with standard variance-reduction techniques. Control variates constructed from the anchor state |psi(theta*)> are particularly attractive because they add no additional circuit executions and only require a one-time calibration of the control-variate coefficient on the anchor. A pilot experiment on the Ionosphere benchmark using an anchor-based control variate reduced the standard error of the QFI estimate at 500 shots per feature to the level otherwise achieved at 1200 shots per feature, at no additional circuit cost.

C. Interaction with Error Mitigation

Zero-noise extrapolation and probabilistic error cancellation are standard error-mitigation techniques on current devices, and both change the effective distribution of the observables that enter the QFI estimate. A conservative recommendation is to run QAFS on the mitigated observables when the downstream classifier will also use mitigation, and on raw observables when it will not; running QAFS with mitigation and the classifier without it introduced a measurable regression in a pilot experiment on the Ionosphere benchmark, consistent with the general principle that the selection procedure and the downstream head should share the same effective encoding.

D. Auditability and Governance

Filter methods are attractive in regulated deployments because the selection decision is a compact, human-inspectable artefact: the paper output of QAFS is an ordered index set of at most n features, together with the QFI score and the kernel-dispersion contribution of each. This is auditable in the same sense as a classical feature-importance ranking and can be logged alongside the model artefacts required for regulatory review. Wrapper-style quantum-aware selectors are harder to audit because their output depends on the joint training trajectory of the downstream classifier and cannot be summarised so compactly.

E. Integration with Classical Pipelines

In production settings QAFS is naturally deployed as a preprocessing step in a hybrid pipeline that also includes a classical fallback classifier for latency-sensitive queries. The selected feature subset is small (n = 8 in the reported experiments) and therefore fits comfortably in the input layer of a classical model, which allows the same features to be shared between the quantum head and a classical head without duplicating preprocessing. This arrangement also gives an operational safety net during the current NISQ era: when device availability or noise conditions make the quantum head unreliable, the classical head can serve queries on the same features, and the system as a whole degrades to a classical baseline rather than failing outright.

VIII. CONCLUSION

This paper presented QAFS, an encoding-aware feature-selection procedure that scores candidate features on a quantum Fisher-information proxy and greedily maximises label-conditional kernel dispersion under a hard qubit budget. Paired with QSVC and QCBM heads on three tabular benchmarks under an eight-qubit budget, QAFS improved accuracy from 0.812 to 0.883, F1-score

from 0.741 to 0.842, and reduced kernel-evaluation cost by roughly a third. Robustness experiments under depolarizing noise up to p = 0.03 showed that QAFS retains a clear margin over PCA and random selection, suggesting that the sensitivity-based criterion also selects components that survive noise better.

The broader implication is that on NISQ devices the feature-selection step should be treated as part of the model, not as an encoding-agnostic preprocessing block. Feature selectors that ignore the geometry of the quantum embedding will systematically underperform selectors that condition on it, and the gap widens as the qubit budget tightens.

A secondary contribution of the paper is a concrete, reproducible recipe for evaluating encoding-aware selection proposals against classical baselines under a shared shot and qubit budget. The evaluation protocol in Section IV-D and the ablations in Sections V-G through V-I are designed to be reused: adding a new selector requires only replacing the QAFS scoring block with the candidate, and the rest of the pipeline, including the noise sweep and the F1-score disaggregation, is reusable without modification. The archived source that accompanies this paper is structured with that reuse in mind, and the calibration set, the selected feature indices, and the trained downstream classifiers are all stored as first-class artefacts rather than embedded in ephemeral notebooks.

Finally, the paper argues that the appropriate unit of engineering effort in near-term quantum machine learning is not the model or the ansatz in isolation but the encoding-selector-classifier triple treated as a single object. QAFS is one instance of that view: a selector that shares the encoding of the downstream classifier and is evaluated on the same shot budget the classifier will use in production. Alternative instances are possible and welcome; the point is that any evaluation which fixes the classifier and varies the selector without also varying the encoding will miss the dominant lever, and any evaluation which varies the encoding without measuring the shot-budget consequences will overstate the practical benefit. The results in Section V are intended as evidence that the triple-level view is not merely a conceptual convenience but a measurable source of accuracy and shot-budget savings on the benchmarks studied.

IX. FUTURE RESEARCH DIRECTIONS

- Anchor-free Fisher scoring that updates the reference parameters jointly with the downstream classifier, closing the drift gap noted in Section VI. This would remove the need to select a fixed anchor theta* and would let the QFI proxy track the sensitivity landscape as the downstream classifier moves in parameter space during training.
- Extension of QAFS to data re-uploading encodings [17], where a single feature is injected across multiple layers and the scoring rule must aggregate sensitivity across repetitions. The natural generalisation is a sum of layer-wise QFI estimates weighted by the effective learning rate at each layer.
- Hardware evaluation on IBM Quantum and IonQ backends with connectivity-aware ansatze, including calibration-drift accounting. On a physical backend the QFI probe would need to be recomputed at a cadence set by the device drift time to keep the ranking valid.
- Submodular relaxations of Eq. (5) with formal approximation guarantees, extending the empirical observation in Section VI that the greedy-optimal gap is small on the tested benchmarks.
- Application to image and sequence domains via convolutional or attention-based quantum feature maps, with a scoring rule adapted to the richer encoding.
- Integration with barren-plateau mitigation strategies [19] so that QAFS jointly optimises for encoding sensitivity and trainability.
- Communication-efficient distributed QAFS for federated settings, where the calibration set is partitioned across institutions.

X. REFERENCES

[1] V. Havlicek, A. D. Corcoles, 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, 2019. Available: https://www.nature.com/articles/s41586-019-0980-2

[2] M. Schuld and N. Killoran, "Quantum machine learning in feature Hilbert spaces," Physical Review Letters, vol. 122, no. 4, article 040504, 2019. Available: https://journals.aps.org/prl/abstract/10.1103/PhysRevLett.122.040504

[3] M. Benedetti, D. Garcia-Pintos, O. Perdomo, V. Leyton-Ortega, Y. Nam, and A. Perdomo-Ortiz, "A generative modeling approach for benchmarking and training shallow quantum circuits," npj Quantum Information, vol. 5, article 45, 2019. Available:

https://www.nature.com/articles/s41534-019-0157-8

[4] J.-G. Liu and L. Wang, "Differentiable learning of quantum circuit Born machines," Physical Review A, vol. 98, no. 6, article 062324, 2018. Available: https://journals.aps.org/pra/abstract/10.1103/PhysRevA.98.062324

[5] I. Guyon and A. Elisseeff, "An introduction to variable and feature selection," Journal of Machine Learning Research, vol. 3, pp. 1157-1182, 2003. Available: https://jmlr.org/papers/v3/guyon03a.html

[6] G. Chandrashekar and F. Sahin, "A survey on feature selection methods," Computers and Electrical Engineering, vol. 40, no. 1, pp. 16-28, 2014. Available: https://doi.org/10.1016/j.compeleceng.2013.11.024

[7] J. J. Meyer, "Fisher information in noisy intermediate-scale quantum applications," Quantum, vol. 5, article 539, 2021. Available: https://quantum-journal.org/papers/q-2021-09-09-539/

[8] A. Abbas, D. Sutter, C. Zoufal, A. Lucchi, A. Figalli, and S. Woerner, "The power of quantum neural networks," Nature Computational Science, vol. 1, no. 6, pp. 403-409, 2021. Available: https://www.nature.com/articles/s43588-021-00084-1

[9] H.-Y. Huang, M. Broughton, M. Mohseni, R. Babbush, S. Boixo, H. Neven, and J. R. McClean, "Power of data in quantum machine learning," Nature Communications, vol. 12, article 2631, 2021. Available: https://www.nature.com/articles/s41467-021-22539-9

[10] Y. Liu, S. Arunachalam, and K. Temme, "A rigorous and robust quantum speed-up in supervised machine learning," Nature Physics, vol. 17, no. 9, pp. 1013-1017, 2021. Available: https://www.nature.com/articles/s41567-021-01287-z

[11] B. Coyle, D. Mills, V. Danos, and E. Kashefi, "The Born supremacy: Quantum advantage and training of an Ising Born machine," npj Quantum Information, vol. 6, article 60, 2020. Available: https://www.nature.com/articles/s41534-020-00288-9

[12] I. T. Jolliffe and J. Cadima, "Principal component analysis: A review and recent developments," Philosophical Transactions of the Royal Society A, vol. 374, article 20150202, 2016. Available: https://royalsocietypublishing.org/doi/10.1098/rsta.2015.0202

[13] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information: Criteria of max-dependency, max-relevance, and min-redundancy," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 27, no. 8, pp. 1226-1238, 2005. Available: https://doi.org/10.1109/TPAMI.2005.159

[14] E. Wilson, S. Singh, and F. Mueller, "Quantum feature map hyperparameter tuning for classification of high-energy physics events," in Proc. IEEE Int. Conf. Quantum Computing and Engineering (QCE), 2021, pp. 300-306. Available: https://doi.org/10.1109/QCE52317.2021.00047

[15] C. Zoufal, A. Lucchi, and S. Woerner, "Quantum generative adversarial networks for learning and loading random distributions," npj Quantum Information, vol. 5, article 103, 2019. Available: https://www.nature.com/articles/s41534-019-0223-2

[16] S. Y.-C. Chen, C.-H. H. Yang, J. Qi, P.-Y. Chen, X. Ma, and H.-S. Goan, "Variational quantum circuits for deep reinforcement learning," IEEE Access, vol. 8, pp. 141007-141024, 2020. Available: https://doi.org/10.1109/ACCESS.2020.3010470

[17] A. Perez-Salinas, A. Cervera-Lierta, E. Gil-Fuster, and J. I. Latorre, "Data re-uploading for a universal quantum classifier," Quantum, vol. 4, article 226, 2020. Available: https://quantum-journal.org/papers/q-2020-02-06-226/

[18] 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, article 4812, 2018. Available: https://www.nature.com/articles/s41467-018-07090-4

[19] 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, article 1791, 2021. Available: https://www.nature.com/articles/s41467-021-21728-w

[20] A. Kandala, A. Mezzacapo, K. Temme, M. Takita, M. Brink, J. M. Chow, and J. M. Gambetta, "Hardware-efficient variational quantum eigensolver for small molecules and quantum magnets," Nature, vol. 549, no. 7671, pp. 242-246, 2017. Available: https://www.nature.com/articles/nature23879

[21] T. Hubregtsen, D. Wierichs, E. Gil-Fuster, P.-J. H. S. Derks, P. K. Faehrmann, and J. J. Meyer, "Training quantum embedding kernels on near-term quantum computers," Physical Review A, vol. 106, no. 4, article 042431, 2022. Available: https://journals.aps.org/pra/abstract/10.1103/PhysRevA.106.042431

[22] W. N. Street, W. H. Wolberg, and O. L. Mangasarian, "Nuclear feature extraction for breast tumor diagnosis," in Proc. IS&T/SPIE Int. Symp. Electronic Imaging: Science and Technology, vol. 1905, 1993, pp. 861-870. (Breast Cancer Wisconsin diagnostic dataset). Available: https://archive.ics.uci.edu/dataset/17/breast+cancer+wisconsin+diagnostic

[23] V. G. Sigillito, S. P. Wing, L. V. Hutton, and K. B. Baker, "Classification of radar returns from the ionosphere using neural networks," Johns Hopkins APL Technical Digest, vol. 10, pp. 262-266, 1989. (Ionosphere dataset). Available: https://archive.ics.uci.edu/dataset/52/ionosphere

[24] R. P. Gorman and T. J. Sejnowski, "Analysis of hidden units in a layered network trained to classify sonar targets," Neural Networks, vol. 1, no. 1, pp. 75-89, 1988. (Sonar dataset). Available: https://archive.ics.uci.edu/dataset/151/connectionist+bench+sonar+mines+vs+rocks

[25] Qiskit contributors, "Qiskit: An open-source framework for quantum computing," IBM Research, 2024. Available: https://qiskit.org