

Is Quantum Cross-Band Fusion Worth Its Cost?
A Data-Scaling Ablation of Hierarchical Quantum DCT Encoders
in Biomedical Image Classification

Xavier Font Aragonés^{1,2} and Miguel A. González Ballester^{2,3,4}

¹Tecnocampus, Universitat Pompeu Fabra

²BCN MedTech, Universitat Pompeu Fabra, Barcelona

³BSC, Barcelona Supercomputer Center, Barcelona

⁴ICREA, Barcelona, Spain

Abstract. Hierarchical quantum spectral encoders often include a global “fusion” stage that entangles otherwise disjoint frequency subspaces, at the price of doubling the variational quantum parameter count and the two-qubit gate depth. We ask a single, practical question: *does this fusion stage actually improve classification, and does the answer change as training data grows?* We run a controlled fusion-on/off ablation of the Hierarchical Quantum Discrete Cosine Transform (H-QDCT) encoder across four training-set sizes ($N \in \{500, 1000, 2000, 5000\}$, the largest only on datasets whose native pools allow it) on six MedMNIST benchmarks ($K \in \{2, 5, 7, 8, 9\}$), with five-to-ten seeds per cell and a paired Wilcoxon design. Pooled across 95 paired observations on the four multiclass tasks ($K \geq 5$), the fusion stage yields no consistent benefit (median $\Delta = -0.010$ macro-F1, $p = 0.148$); across all six datasets only 1 of 21 per-cell tests reaches significance, consistent with multiple-testing noise, and increasing N up to 5000 does not turn the small effect into a reliable one. Removing fusion halves the variational quantum parameters ($144 \rightarrow 72$) and the per-patch two-qubit gate count at no measurable accuracy cost—a free simplification for NISQ deployment. We conclude that, for this class of encoder, classical re-integration of the subspace outputs is sufficient and on-circuit cross-band fusion is dispensable.

1 Motivation. Hierarchical Quantum DCT (H-QDCT) encoders [1] extract multi-scale spectral features by composing a unitary Quantum Fourier Transform with frequency-partitioned variational StronglyEntanglingLayers (SEL) [2] acting on disjoint qubit subspaces—a low-frequency “structure” subspace and a high-frequency “texture” subspace—whose per-qubit Pauli- Z outputs are concatenated and read out by a small classical head. A common design choice in such architectures is to add a third SEL stage acting on the *full* register, mixing the two subspaces on-circuit before readout. We call this the *cyclic global fusion* stage. It is intuitively attractive: it lets the structure and texture channels interact inside the quantum device rather than only in the classical head.

But fusion is not free: on the H-QDCT encoder it doubles the variational quantum parameter count ($72 \rightarrow 144$) and adds a block of two-qubit gates to every patch circuit. Since two-qubit gate depth is the dominant source of NISQ decoherence, this stage must earn its place by improving accuracy enough to justify the extra noise—and whether it does, and whether its value grows or shrinks as more training data becomes available, is the question this study answers empirically.

Contribution. (i) A four-point data-scaling ablation of the fusion stage on six biomedical benchmarks, with evaluation sets held fixed so differences are attributable to training-data volume; (ii) the finding that fusion contributes no consistent accuracy benefit at any scale tested, the effect remaining within seed noise even at $N = 5000$; (iii) a concrete architectural recommendation—drop the fusion stage—.

2 Architecture and the ablation. Figure 1 summarises the H-QDCT encoder ($n_q = 8$ qubits, $P = 4$ patches). Each patch is amplitude-embedded, transformed by a unitary QDCT (QFT followed by uniform $\pi/2$ phase shifts), then processed by two parallel StronglyEntanglingLayers acting on disjoint frequency subspaces: subspace A (low frequencies, qubits 0–3) and subspace B (high frequencies, qubits 4–7). The optional fusion stage applies a third SEL across all eight qubits. Readout is the per-qubit expectation $\langle Z_w \rangle$, giving a 32-dimensional feature vector that a small classical head maps to the K class logits. The *with-fusion* and *no-fusion* variants are

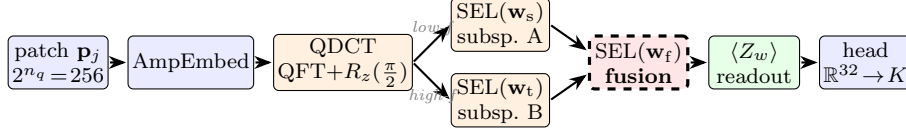


Figure 1: H-QDCT-Coarse encoder ($n_q=8$, $P=4$). Two parallel StronglyEntanglingLayers act on disjoint low- and high-frequency subspaces. The **cyclic global fusion** stage (dashed red) is the subject of the ablation: it adds 72 variational parameters (total 144) and a block of two-qubit gates per patch. The no-fusion variant omits it (72 parameters) without changing the readout or the head.

identical except for the presence of the dashed block in Fig. 1; everything downstream (readout dimension, classical head, training protocol) is held constant, so the ablation isolates exactly the contribution of on-circuit cross-band fusion.

3 Experimental setup. We train both H-QDCT variants at four scales $N_{\text{train}} \in \{500, 1000, 2000, 5000\}$ on six MedMNIST benchmarks [3]: PneumoniaMNIST and BreastMNIST ($K=2$), RetinaMNIST ($K=5$), DermaMNIST ($K=7$), BloodMNIST ($K=8$), PathMNIST ($K=9$). Evaluation sets are held fixed ($N_{\text{val}} = N_{\text{test}} = 150$) at every scale so that trajectory differences reflect only training-data volume. We use five seeds $\{42, \dots, 46\}$ at $N \in \{500, 1000, 2000\}$ and ten seeds $\{42, \dots, 51\}$ at $N=5000$; the $N=5000$ point is restricted to the three datasets (Derma, Blood, Path) whose native training pools exceed 5000 images, and the PathMNIST $N=2000$ cell was additionally extended to ten seeds to harden the comparison at the largest task. Training uses class-weighted cross-entropy, AdamW with separate quantum/classical learning rates ($10^{-3}/10^{-4}$), cosine warmup, and early stopping, simulated with PennyLane `default.qubit` [5]. The fusion stage is compared by a paired Wilcoxon signed-rank test on macro-F1, seed-matched within each (dataset, N) cell.

4 Does fusion help? No, at any scale. Table 1 reports the seed-matched comparison in every (dataset, N) cell. The picture is consistent across the grid: the fusion-vs-no-fusion difference is small, its sign varies by dataset rather than tracking class count or data size, and almost no cell reaches significance. For the pooled test we group the four multiclass tasks ($K \geq 5$), where discriminating several classes is the relevant setting for a cross-band interaction; the two binary tasks are reported per-cell in Table 1 but excluded from the pool. Across all 95 paired observations on these datasets the median difference is -0.010 macro-F1 (mean -0.007); 58/95 paired observations nominally favor fusion and 37/95 favor no-fusion, but the paired Wilcoxon test is far from significant ($p=0.148$). The typical effect size ($|\Delta| \approx 0.02$) is smaller than the within-cell seed standard deviation (≈ 0.05): fusion moves accuracy by less than the run-to-run noise.

Crucially, *adding data does not rescue fusion*. If the fusion stage encoded a real but data-hungry interaction, its benefit should emerge as N grows. It does not: the $N=5000$ cells are as inconclusive as the $N=500$ cells. Of the 21 per-cell tests, exactly one reaches $p < 0.05$ (PathMNIST at $N=5000$, $\Delta = -0.050$ in favor of fusion, $p=0.0488$). With 21 tests at $\alpha=0.05$ one false positive is expected by chance. We note that the two largest effects (PathMNIST/5000, $\Delta = -0.050$; DermaMNIST/1000, $\Delta = -0.050$, $p=0.06$) both lean toward fusion; but the signs scatter overall, and the comparably hard BloodMNIST favors no-fusion at every scale, so we read the single significant cell as multiple-testing noise rather than a real effect.

Resource advantage of the no-fusion variant. Because fusion makes no consistent difference to accuracy, the comparison reduces to cost. The no-fusion variant uses 72 variational quantum parameters instead of 144, and omits an eight-qubit block of two-qubit gates from every patch circuit. On NISQ devices this depth reduction matters directly: halving the gate budget by removing fusion is expected to improve noise survival at no cost in classification capability. For deployment, the no-fusion variant is therefore the better default.

5 Discussion. The plain reading of Table 1 is that on-circuit cross-band fusion is not pulling

Table 1: Fusion ablation, macro-F1, seed-matched within each cell. $\Delta = (\text{no-fusion} - \text{with-fusion})$; a positive Δ favors *removing* fusion. n is the number of paired seeds; p is the paired Wilcoxon signed-rank value. Dashes: dataset pool too small for $N=5000$. No cell except PathMNIST/5000 reaches $p < 0.05$, and the sign of Δ is not consistent across datasets or scales.

Dataset	K	N	n	Δ (NoFus–Fus)	p
PneumoniaMNIST	2	500	5	+0.006	0.81
		1000	5	+0.033	0.31
		2000	5	−0.016	0.81
BreastMNIST	2	500	5	−0.029	0.81
		1000	5	−0.016	0.63
		2000	5	−0.016	0.63
RetinaMNIST	5	500	5	−0.015	0.81
		1000	5	−0.014	0.19
		2000	5	−0.015	0.31
DermaMNIST	7	500	5	−0.001	0.81
		1000	5	−0.050	0.06
		2000	5	−0.005	0.63
		5000	10	+0.024	0.70
BloodMNIST	8	500	5	+0.008	1.00
		1000	5	+0.003	1.00
		2000	5	+0.021	0.44
		5000	10	+0.018	0.28
PathMNIST	9	500	5	+0.034	0.44
		1000	5	−0.034	0.44
		2000	10	−0.023	0.23
		5000	10	−0.050	0.0488

its weight for this encoder: the two frequency subspaces are read out independently and recombined by the classical head, and that classical re-integration evidently suffices, so a quantum stage that mixes the subspaces before measurement adds parameters and gate depth without adding discriminative information. This holds across the full class range ($K \in [2, 9]$) and—most importantly for a scaling study—as training data grows tenfold: the natural hypothesis that fusion is beneficial but data-hungry is not supported.

Following best practice in QML benchmarking [4], we do not over-read the single significant cell; we report the fusion effect as *absent at the resolution of this study*. Either way, a practitioner choosing between the two variants should take the cheaper one.

References

- [1] X. Font Aragones, M. A. González Ballester, “H-QDCT: Hierarchical Quantum DCT for Structural–Textural Feature Fusion in Medical Imaging,” manuscript under review (2026).
- [2] M. Schuld, A. Bocharov, K. M. Svore, N. Wiebe, “Circuit-centric quantum classifiers,” *Phys. Rev. A* 101, 032308 (2020).
- [3] J. Yang et al., “MedMNIST v2 – A large-scale lightweight benchmark for 2D and 3D biomedical image classification,” *Scientific Data* 10, 41 (2023).
- [4] J. Bowles, S. Ahmed, M. Schuld, “Better than classical? The subtle art of benchmarking quantum machine learning models,” arXiv:2403.07059 (2024).
- [5] V. Bergholm et al., “PennyLane: Automatic differentiation of hybrid quantum-classical computations,” arXiv:1811.04968 (2018).