Supplementary Information

TE-HQViT-Probs: Full-Probability Quantum Self-Attention Outperforms Expectation-Value Readout in Medical Image Classification

Nguyen Thanh Phong

Faculty of Information Technology, Van Lang University, Ho Chi Minh City, Vietnam

phong.2474802010304@vanlanguni.vn

Scope of this supplement. This document consolidates the extended methodological definitions, control structure, complete numerical tables, statistical summaries, and diagnostic analyses reported in the accompanying manuscript. It deliberately does not add experimental settings that are absent from the manuscript or executed notebook outputs available for this paper. In particular, optimizer, learning-rate schedule, batch size, epoch count, finite-shot sampling settings, and device-level hardware calibration are not specified in the source manuscript and therefore are not reconstructed here.

Contents

Supplementary Note S1. Experimental Scope and Study Logic

The study isolates the *quantum readout map* as an architectural variable in a compact hybrid quantum self-attention value path. The evaluated setting uses four measured qubits ($n = 4$) and two variational layers ($L = 2$), with amplitude embedding followed by a parameterized quantum circuit. Query and key projections remain classical, while the value representation is obtained from a quantum measurement map.

For a token representation $Z \in \mathbb{R}^{N\times d}$, the classical query and key projections are

$$Q = ZW_Q, \qquad K = ZW_K, \tag{1}$$

and the attention matrix is

$$A = \mathrm{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right). \tag{2}$$

For an input value vector $\boldsymbol{v}$, amplitude encoding prepares $|\psi_{\mathrm{in}}(\boldsymbol{v})\rangle$ and the parameterized circuit produces

$$|\psi(\boldsymbol{v},\boldsymbol{\theta})\rangle = U(\boldsymbol{\theta})|\psi_{\mathrm{in}}(\boldsymbol{v})\rangle. \tag{3}$$

The Full-Probs readout exposes all computational-basis probabilities

$$p_b = |\langle b|\psi(\boldsymbol{v},\boldsymbol{\theta})\rangle|^2, \qquad b \in \{0,1\}^4, \tag{4}$$

producing a 16-dimensional representation.

The central comparison is not intended to establish quantum computational advantage. Rather, it tests whether a richer measurement map preserves classification-relevant information that is compressed by selected expectation values.

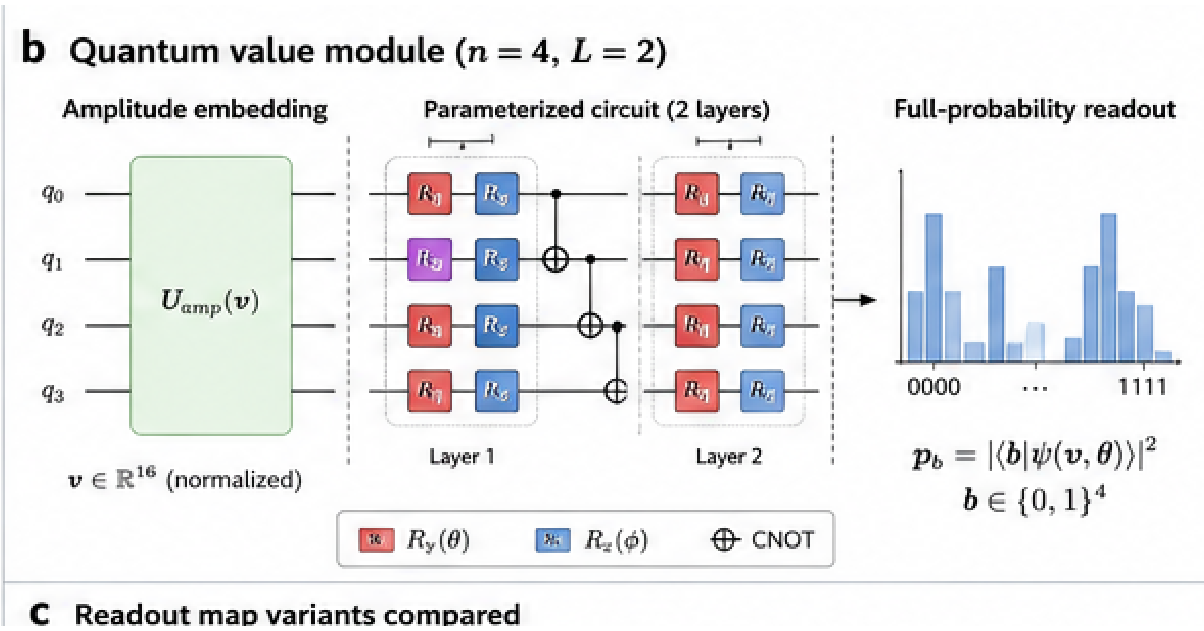

Figure 1: Supplementary Fig. S1 | Quantum value module. The four-qubit, two-layer quantum value path uses amplitude embedding, a shallow parameterized circuit, and full computational-basis probability readout. The figure is a conceptual representation of the module described in the manuscript.

Supplementary Note S2. Readout Maps and Information Compression

The evaluated readout ladder comprises Expval-Z, Z+ZZ, Prob-k4, Prob-k8, and Full-Probs. For single-qubit Pauli-Z readout,

$$\langle Z_i \rangle = \sum_{b \in \{0,1\}^n} (-1)^{b_i} p_b. \tag{5}$$

Thus, each $\langle Z_i \rangle$ is a linear summary of the complete basis-state probability distribution. Pairwise ZZ moments add second-order information,

$$\langle Z_i Z_j \rangle = \sum_{b \in \{0,1\}^n} (-1)^{b_i + b_j} p_b, \tag{6}$$

while truncated probability readouts retain only a fixed subset of entries from the simulator-ordered probability vector. Full-Probs retains all $2^4 = 16$ probabilities.

This hierarchy should be interpreted as an increasing-capacity readout ladder rather than as a claim that full-probability measurement is scalable. In general, its dimensionality grows as 2^n with the number of measured qubits.

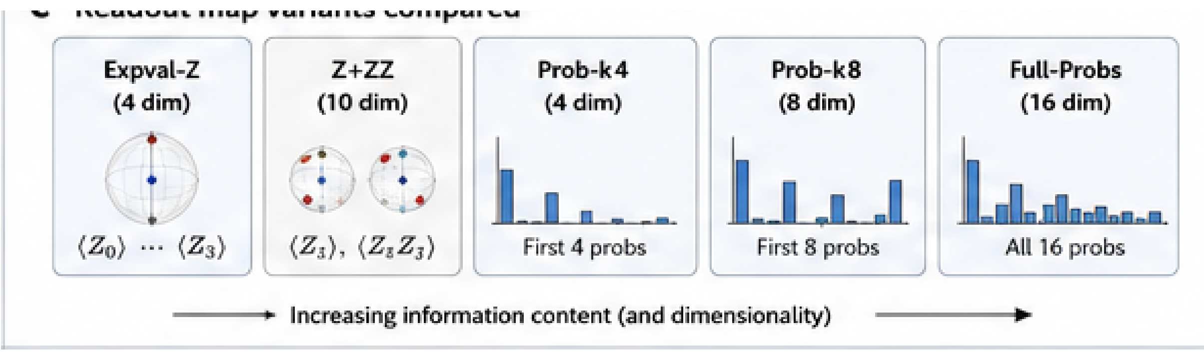

Figure 2: Supplementary Fig. S2 | Readout variants. Conceptual comparison of Expval-Z, Z+ZZ, truncated probability readouts, and Full-Probs. Full-Probs is used as a high-capacity diagnostic readout in the four-qubit simulation setting.

Supplementary Note S3. Controls and Interpretation

The experimental control structure separates three questions:

1. **Readout-map effect:** TE-HQViT-Probs versus TE-HQViT-Expval and the controlled readout ladder test whether changing the measurement representation changes predictive performance.
2. **Classical competitiveness:** classical softmax attention, classical MLP-value attention, ResNet18, and ViT-Tiny prevent the readout result from being misinterpreted as broad superiority over classical vision models.
3. **Trainable-quantum-dynamics effect:** frozen-random TE-HQViT-Probs tests whether the benefit depends on optimizing circuit parameters or whether a fixed probability-based feature map already supplies useful information.

The manuscript reports that trainable and frozen-random probability readout are practically equivalent across the three datasets in the TOST analysis. This narrows the supported interpretation toward readout/representation capacity rather than trained quantum dynamics alone.

Supplementary Note S4. Repeated-Seed and Statistical Protocol

All primary paired comparisons use the same ten random seeds:

$$\{42, 7, 123, 2024, 91, 1337, 314, 2718, 8888, 51\}. \tag{7}$$

The primary endpoint is test AUC, with accuracy and F1 as secondary metrics. Seed-level paired AUC differences are evaluated by the Wilcoxon signed-rank test. Holm correction is used for the reported multiple-comparison families. Practical equivalence is evaluated by TOST with an absolute AUC margin $\Delta = 0.02$.

These repeated-seed tests characterize optimization/run-to-run stability. They do not create additional independent datasets; accordingly, dataset-level generalization claims remain limited by the three MedMNIST benchmarks.

Supplementary Table S1. Complete Main Architecture Benchmark

Table 1: Complete main benchmark results across ten random seeds. Values are reproduced from the manuscript.

Dataset	Model	Params	Accuracy	AUC	F1
PneumoniaMNIST	TE-HQViT-Probs	4,362	0.840 [0.831, 0.849]	0.937 [0.933, 0.942]	0.885 [0.879, 0.890]
PneumoniaMNIST	TE-HQViT-Expval	4,362	0.837 [0.831, 0.844]	0.931 [0.928, 0.934]	0.882 [0.878, 0.886]
PneumoniaMNIST	Classical softmax attention	4,402	0.836 [0.828, 0.844]	0.931 [0.926, 0.936]	0.882 [0.877, 0.887]
PneumoniaMNIST	Frozen-random TE-HQViT-Probs	4,338	0.846 [0.835, 0.856]	0.939 [0.935, 0.942]	0.888 [0.882, 0.895]
PneumoniaMNIST	ResNet18 fine-tuned	11,171,266	0.868 [0.857, 0.880]	0.960 [0.955, 0.965]	0.904 [0.897, 0.912]
BreastMNIST	TE-HQViT-Probs	4,362	0.752 [0.736, 0.767]	0.722 [0.709, 0.737]	0.842 [0.831, 0.852]
BreastMNIST	TE-HQViT-Expval	4,362	0.742 [0.734, 0.750]	0.673 [0.647, 0.695]	0.846 [0.842, 0.849]
BreastMNIST	Classical softmax attention	4,402	0.775 [0.760, 0.789]	0.742 [0.721, 0.758]	0.861 [0.854, 0.869]
BreastMNIST	Frozen-random TE-HQViT-Probs	4,338	0.748 [0.735, 0.761]	0.721 [0.705, 0.738]	0.840 [0.830, 0.848]
BreastMNIST	ResNet18 fine-tuned	11,171,266	0.744 [0.624, 0.838]	0.855 [0.831, 0.879]	0.770 [0.612, 0.888]
OrganAMNIST	TE-HQViT-Probs	4,515	0.699 [0.694, 0.705]	0.950 [0.949, 0.952]	0.688 [0.683, 0.695]
OrganAMNIST	TE-HQViT-Expval	4,515	0.627 [0.620, 0.636]	0.927 [0.924, 0.930]	0.617 [0.608, 0.626]
OrganAMNIST	Classical softmax attention	4,555	0.640 [0.629, 0.649]	0.926 [0.923, 0.929]	0.632 [0.619, 0.642]
OrganAMNIST	Frozen-random TE-HQViT-Probs	4,491	0.696 [0.692, 0.701]	0.950 [0.949, 0.951]	0.686 [0.681, 0.691]
OrganAMNIST	ResNet18 fine-tuned	11,175,883	0.905 [0.901, 0.909]	0.993 [0.992, 0.993]	0.902 [0.899, 0.906]

Supplementary Table S2. Controlled Readout Ladder

Table 2: Mean test AUC for the controlled readout ladder.

Dataset	Expval-Z	Z+ZZ	Prob-k4	Prob-k8	Full-Probs
PneumoniaMNIST	0.9306	0.9350	0.9323	0.9343	**0.9357**
BreastMNIST	0.6310	0.6615	0.6220	0.6655	**0.7157**
OrganAMNIST	0.9218	0.9406	0.9055	0.9375	**0.9484**

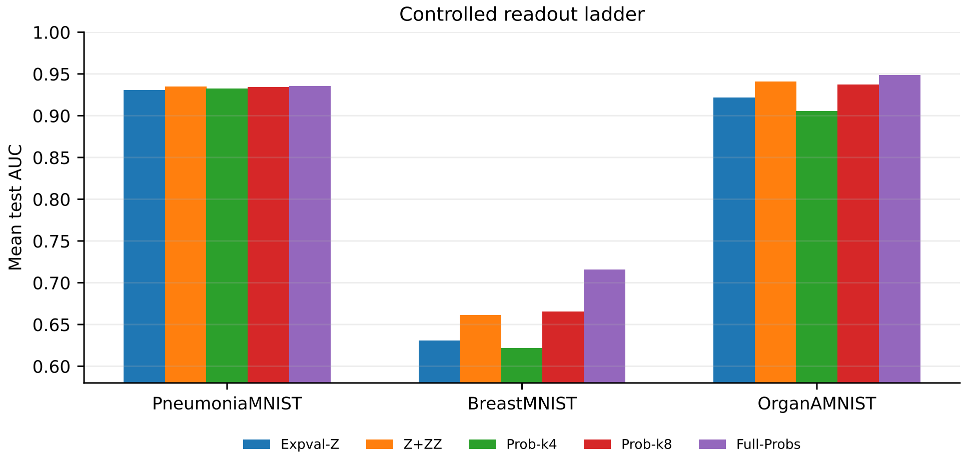

Figure 3: Supplementary Fig. S3 | Controlled readout ladder. Mean test AUC for the five readout schemes reported in Supplementary Table S2.

Table 3: Paired seed-level AUC comparisons for probability versus expectation-value readout.

Experiment	Dataset	Comparison	ΔAUC	Wilcoxon p	Holm-p	Interpretation
Main architecture	PneumoniaMNIST	Probs vs Expval	0.0062	0.0488	0.2441	Small, nominal only
Main architecture	BreastMNIST	Probs vs Expval	0.0497	0.0098	0.0586	Large, near corrected threshold
Main architecture	OrganAMNIST	Probs vs Expval	0.0236	0.0020	0.0137	Large, corrected
Readout ladder	PneumoniaMNIST	Full-Probs vs Expval-Z	0.0052	0.0488	0.1092	Small, nominal only
Readout ladder	BreastMNIST	Full-Probs vs Expval-Z	0.0847	0.0020	0.0180	Large, corrected
Readout ladder	OrganAMNIST	Full-Probs vs Expval-Z	0.0266	0.0020	0.0180	Large, corrected

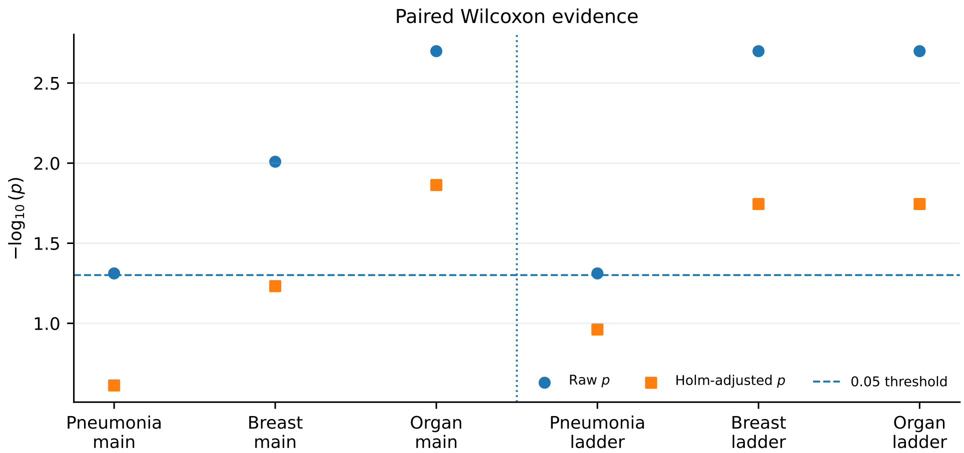

Figure 4: Supplementary Fig. S4 | Paired Wilcoxon evidence. Nominal and Holm-adjusted evidence for probability-readout versus expectation-value comparisons. The dashed line corresponds to the 0.05 threshold.

Table 4: TOST practical-equivalence summary using an absolute 0.02 AUC margin.

Dataset	Comparison	ΔAUC	Conclusion
PneumoniaMNIST	Probs vs Classical attention	+0.0065	Equivalent
PneumoniaMNIST	Trainable vs Frozen Probs	-0.0011	Equivalent
BreastMNIST	Probs vs Classical attention	-0.0195	Inferior
BreastMNIST	Trainable vs Frozen Probs	+0.0012	Equivalent
OrganAMNIST	Probs vs Classical attention	+0.0241	Non-inferior
OrganAMNIST	Trainable vs Frozen Probs	+0.0001	Equivalent

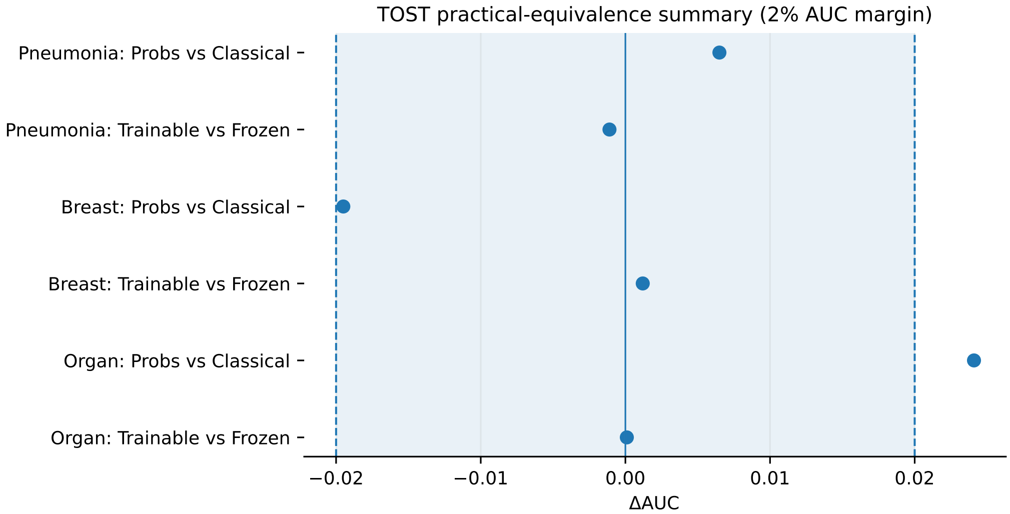

Figure 5: Supplementary Fig. S5 | Practical-equivalence analysis. TOST summary for comparisons with the classical softmax-attention control and the frozen-random probability-readout control.

Supplementary Table S3. Paired Wilcoxon Analysis

Supplementary Table S4. Practical-Equivalence Analysis

Supplementary Note S5. Compression and Frozen-Probe Diagnostics

The manuscript includes two diagnostics designed to separate readout dimensionality from representation content. First, on PneumoniaMNIST, the 16-dimensional Full-Probs representation is compressed to four dimensions by PCA and compared with the native four-dimensional Expval-Z readout. The reported AUC values are 0.8945 for Expval-Z, 0.9286 for PCA-compressed probabilities, and 0.9320 for Full-Probs. The four-dimensional PCA projection therefore remains markedly above the dimension-matched Expval-Z representation in this diagnostic setting.

Second, frozen readout-only probes report AUC values of 0.7213 for Expval-Z, 0.8071 for Z+ZZ, and 0.8462 for Full-Probs. These values are consistent with the interpretation that richer readout maps preserve more class-relevant information for the downstream probe in the evaluated setting.

Supplementary Note S6. Depolarizing-Channel Stress Test

The mixed-state stress test is a simulator diagnostic rather than a hardware validation. The reported AUC decreases from 0.9389 at depolarizing probability $p = 0$ to 0.9322 at $p = 0.05$. This modest change under the selected simulated channel does not establish robustness to physical-device noise, which would additionally involve finite-shot sampling, readout error, gate-dependent noise, calibration drift, and hardware connectivity constraints.

Supplementary Note S7. Claim Boundary and Failure Modes

The evidence supports a narrow readout-representation claim. Full-Probs achieves higher mean AUC than the corresponding expectation-value variant on all three evaluated datasets, and the controlled readout ladder shows the same directional pattern. However, four boundaries are important:

- The effect magnitude is dataset-dependent; PneumoniaMNIST shows only a small improvement.

Table 5: Diagnostic analyses supporting the readout-representation interpretation.

Analysis	**Setting**	**AUC**
Compression	Expval-Z, 4 dimensions	0.8945
Compression	PCA projection of probabilities, 4 dimensions	0.9286
Compression	Full probabilities, 16 dimensions	0.9320
Frozen readout-only probe	Expval-Z	0.7213
Frozen readout-only probe	Z+ZZ	0.8071
Frozen readout-only probe	Full-Probs	0.8462
Depolarizing stress test	$p = 0$	0.9389
Depolarizing stress test	$p = 0.05$	0.9322

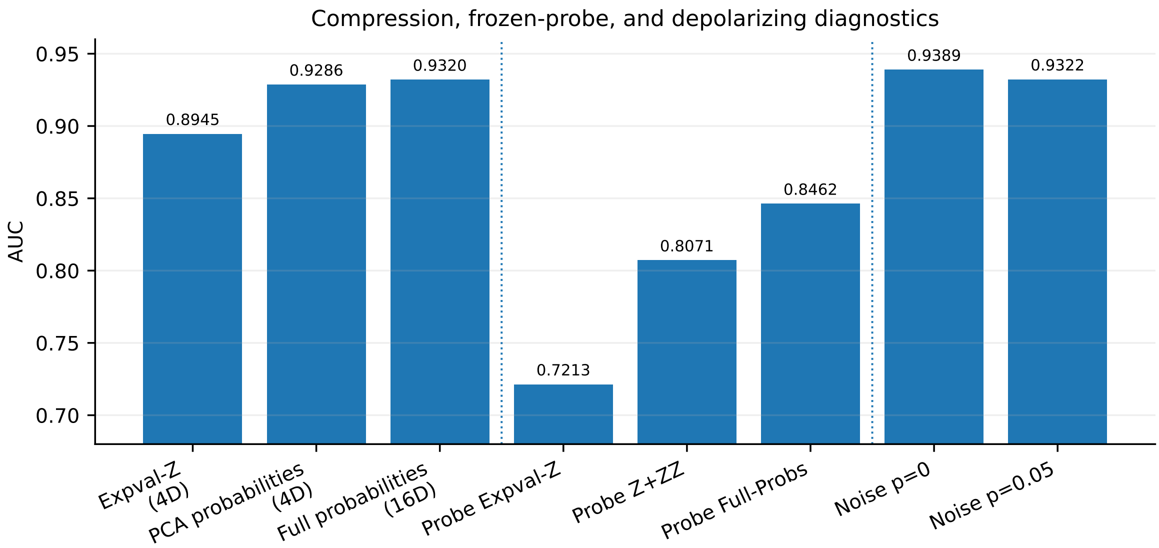

Figure 6: Supplementary Fig. S6 | Diagnostic analyses. Compression, frozen readout-only, and depolarizing-channel diagnostics reproduced from the manuscript results.

- BreastMNIST demonstrates that improved quantum readout does not imply superiority over classical attention; the classical softmax-attention control has higher AUC in the reported benchmark.
- Practical equivalence between trainable and frozen-random probability readouts weakens an interpretation based specifically on learned quantum dynamics.
- Full probability readout scales exponentially with the number of measured qubits and is therefore best viewed here as a small-qubit diagnostic or upper-capacity readout.

Accordingly, the paper does not establish quantum advantage, hardware speedup, or clinical deployment readiness.

Supplementary Note S8. Reproducibility Checklist

Table 6: Reproducibility information explicitly available in the manuscript and information that should be added to a future archival release if available from the experimental notebook.

Item	Information	Status in manuscript
Datasets	PneumoniaMNIST, BreastMNIST, OrganAMNIST	Specified
Quantum circuit size	4 measured qubits, 2 variational layers	Specified
Input encoding	Amplitude embedding	Specified
Primary readouts	Expval-Z, Z+ZZ, Prob-k4, Prob-k8, Full-Probs	Specified
Primary seeds	42, 7, 123, 2024, 91, 1337, 314, 2718, 8888, 51	Specified
Primary metric	Test AUC	Specified
Secondary metrics	Accuracy, F1	Specified
Paired test	Wilcoxon signed-rank	Specified
Multiplicity control	Holm correction	Specified
Equivalence margin	Absolute AUC margin $\Delta = 0.02$	Specified
Noise diagnostic	Mixed-state depolarizing-channel stress test	Specified at study level
Optimizer / learning rate	Not reported in the manuscript text used to prepare this supplement	Not specified
Batch size / epoch count	Not reported in the manuscript text used to prepare this supplement	Not specified
Finite-shot measurement count	Not reported; the study is simulation-based	Not specified
Hardware backend	No physical quantum hardware validation is claimed	Not applicable to main claim

Supplementary Note S9. Recommended Archival Supplement Package

For a future public or journal archival release, the manuscript states that the experimental pipeline, seed-level logs, statistical analysis scripts, and generated figures can accompany the paper. A complete reproducibility package should preserve the same dataset partitions and paired seeds used in the reported analyses. If the executed notebook contains optimizer, batch-size, epoch, preprocessing, or software-version metadata that are not currently stated in the manuscript, those values should be exported directly from the notebook rather than reconstructed retrospectively.

This Supplementary Information is designed to accompany the manuscript and does not introduce new experimental claims beyond the reported study.