

Article

Hybrid Quantum-Classical Neural Networks for Healthcare Prediction Powered by Automated Scientific Discovery

Karthik Meduri ^{1,*}, Ruthvik Yedla ², Santosh Reddy Addula ¹, Guna Sekhar Sajja ¹, Shaila Rana ^{3,†}, Elyson De La Cruz ^{3,†}, Mohan Harish Maturi ¹ and Hari Gonaygunta ^{1,†}

¹ Department of Information Technology, University of the Cumberlands, Williamsburg, KY 40769, USA; saddula5840@ucumberlands.edu (S.R.A.); sajja.guna@ieee.org (G.S.S.); mmaturi9213@ucumberlands.edu (M.H.M.); hgonaygunta3255@ucumberlands.edu (H.G.)

² Department of Computer Science, University of Central Missouri, Warrensburg, MO 64093, USA; yedla.ruthvik@ieee.org

³ School of Business and Information Technology, Purdue University Global, West Lafayette, IN 46240, USA; shaila.rana@purdueglobal.edu (S.R.); elyson.delacruz@purdueglobal.edu (E.D.L.C.)

* Correspondence: kmeduri5644@ucumberlands.edu or karthik.meduri@ieee.org

† These authors contributed equally to this work.

Abstract

This study presents a reproducible evaluation framework for hybrid quantum-classical neural networks (HQCNNs) in healthcare classification, rather than a new architecture. We assess a four-qubit HQCNN combining a compact classical encoder, a two-layer parameterized quantum circuit (PQC), and a classical readout (441 trainable parameters) against carefully tuned classical baselines on the Wisconsin Diagnostic Breast Cancer (WDBC) dataset under identical five-fold cross-validation. The work is framed as a single-dataset proof-of-concept: the contribution is a documented, shared-fold evaluation protocol with a parameter-matched classical control and a quantified epistemic-informativeness analysis, not a demonstration of general quantum advantage. The HQCNN reached $96.49 \pm 1.96\%$ accuracy and $99.44 \pm 0.60\%$ ROC-AUC. A parameter-matched classical multilayer perceptron (441 parameters) reached $95.08 \pm 1.81\%$ accuracy; the HQCNN's +1.41 percentage-point edge at equal capacity was not statistically significant (paired t , $p = 0.056$). Across five shared folds, no HQCNN-versus-classical accuracy difference survived Holm–Bonferroni correction (all adjusted $p \geq 0.625$), so we report the HQCNN as competitive with, not superior to, strong tuned classical baselines. A multi-split depth ablation showed that circuit depth $L \in \{1, 2, 3\}$ had no statistically detectable effect on accuracy ($L = 2$ vs. $L = 3$: Wilcoxon $p = 1.00$); we therefore adopt two variational layers as a practical default rather than an optimum. Under a low-noise simulator (depolarising and amplitude-damping channels, $p = 0.01$), accuracy was 96.49%, indicating robustness only at modest uniform error rates; realistic hardware noise is higher. We additionally apply Bayesian surprise as an epistemic-informativeness heuristic—not a formal generative model—to rank which findings are most worth building on. The framework offers a reproducible, documented evaluation procedure that can support cumulative comparison of hybrid quantum-classical models in healthcare.

Keywords: hybrid quantum-classical neural network; parameterized quantum circuit; reproducible evaluation; breast cancer diagnosis; cross-validation; Bayesian surprise; quantum machine learning; NISQ



Academic Editor: Jiang Bian

Received: 31 March 2026

Revised: 2 June 2026

Accepted: 12 June 2026

Published: 22 June 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article

distributed under the terms and

conditions of the [Creative Commons](https://creativecommons.org/licenses/by/4.0/)

[Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

1.1. Background

Quantum machine learning explores whether quantum information processing can complement classical learning [1,2], particularly in the noisy intermediate-scale quantum (NISQ) era where shallow parameterised quantum circuits (PQCs) are the practical unit of computation. Foundational results show that quantum feature spaces and circuit learning can in principle express functions that are costly classically [3,4]. In healthcare, where datasets are often small, high-dimensional, and class-imbalanced, parameter-efficient models are attractive [5,6].

Hybrid quantum-classical neural networks combine classical layers for feature compression with a PQC as a trainable nonlinear block. The classical components handle dimensionality and the quantum block contributes an alternative parameterisation of the decision function.

1.2. Motivation

The practical question is not whether quantum-containing models can match classical performance—recent work shows they often can on tabular benchmarks [7]—but whether reported comparisons are conducted fairly and reproducibly enough to be cumulative. Many hybrid-model studies report point estimates on a single split without matched baselines or significance testing, which makes it hard to attribute differences to the quantum component versus tuning, capacity, or chance.

1.3. Problem Statement

We address the evaluation problem rather than an architectural one. Given an established hybrid architecture, what protocol allows a defensible comparison against classical baselines, and what can such a comparison legitimately conclude on a single benchmark dataset?

1.4. Objectives

The objectives are: (i) to specify a reproducible evaluation protocol with shared cross-validation folds and identically tuned baselines; (ii) to include a parameter-matched classical control so that parameter count is not confounded with the quantum component; (iii) to report formal significance tests rather than rely on overlapping error bars; and (iv) to demonstrate the protocol as a single-dataset proof-of-concept on WDBC. We do not claim general or architectural novelty.

1.5. Contributions

1. A reproducible, shared-fold evaluation protocol for hybrid quantum-classical classifiers, including a parameter-matched classical control.
2. An honest, significance-tested comparison on WDBC showing the HQCNN is competitive with—not significantly better than—strong tuned classical baselines.
3. A multi-split circuit-depth ablation establishing that depth $L \in \{1, 2, 3\}$ does not materially affect accuracy in this shallow regime.
4. A Bayesian-surprise heuristic for ranking the epistemic informativeness of findings, kept strictly separate from prediction.

1.6. Organisation

Section 2 surveys the literature. Section 3 details the methodology. Section 4 presents results. Section 5 interprets findings and discusses limitations. Section 6 concludes.

2. Related Work and Methods Background

2.1. Hybrid Quantum-Classical Models

The classical-encoder/PQC/classical-readout pattern used here follows established hybrid designs [8,9]; we do not present the architecture as new. Parameterised quantum circuits act as trainable feature maps whose expressivity depends on encoding, entanglement structure, and depth [1,10], and automatic differentiation of hybrid pipelines is now standard tooling [9].

2.2. Quantum Feature Encoding

Angle (rotation) encoding maps classical features to single-qubit rotations and is standard for low-qubit NISQ models [4]. Encoding choice has been shown to materially affect classifier expressivity and robustness [10]. We use RY angle encoding on four qubits after classical compression to four features.

2.3. Classical Baselines in QML Benchmarking

Strong classical baselines (tuned SVM, random forest, gradient boosting, and MLPs) are essential for credible QML comparison [7,11]. Comprehensive reviews of QML spanning the NISQ-to-fault-tolerant trajectory [12] and healthcare-specific applications [13,14] confirm that hybrid models are competitive on tabular benchmarks but have not demonstrated consistent accuracy supremacy. Crucially, a baseline matched in parameter count to the hybrid model is needed to separate the effect of the quantum block from the effect of model capacity.

2.4. Barren Plateaus and Trainability

Barren plateaus—exponentially vanishing gradients—limit the trainability of deep or wide PQCs [15,16]. They arise principally at larger qubit counts and circuit depths than used here [17]; gradient-free optimisers do not escape them either [18]. Locally structured entanglement (e.g., quantum convolutional designs) can avoid barren plateaus [19]; a shallow four-qubit, ≤ 3 -layer sweep does not test for them, and we make no barren-plateau-avoidance claim from our ablation.

2.5. Bayesian Surprise as an Epistemic Heuristic

We use Bayesian surprise as an epistemic-informativeness heuristic, not a formal generative Bayesian model of accuracy. For a hypothesis represented by a prior over a success probability and a posterior after observing outcomes, the surprise is the Kullback–Leibler (KL) divergence from prior to posterior [20]:

$$S = D_{\text{KL}}(p(\theta | D) \| p(\theta)) = \int p(\theta | D) \log \frac{p(\theta | D)}{p(\theta)} d\theta, \quad (1)$$

where θ is the latent success-probability parameter, $p(\theta)$ is the prior (a Beta distribution encoding the pre-experiment expectation), $p(\theta | D)$ is the posterior after observing data D (a Beta distribution updated by observed successes and failures), and S is reported in nats. This quantity is used purely confirmatorily and post hoc to rank how informative each finding is. It contributes nothing to prediction: it never enters model selection, training, or hyperparameter choice.

2.6. Automated Scientific Discovery

The AutoDiscovery framework [20] (NeurIPS 2025) generates hypotheses from data autonomously, executes experiments to test them, and ranks results by Bayesian

surprise. We use it as a confirmatory post-hoc instrument to validate preprocessing choices independently.

2.7. Framework Summary

The framework presented in this paper is a reusable evaluation procedure: shared folds, identically tuned baselines, a parameter-matched control, formal significance testing, and a separate epistemic-informativeness summary. It is demonstrated here as a single-dataset proof-of-concept on WDBC.

3. Materials and Methods

3.1. Study Design

This is an empirical evaluation study comparing one hybrid quantum-classical model against classical baselines under a fixed protocol on a single benchmark dataset.

3.2. Dataset

We use the Wisconsin Diagnostic Breast Cancer (WDBC) dataset [21,22]: 569 samples, 30 real-valued features computed from digitised fine-needle-aspirate images, and a binary malignant/benign label (357 benign, 212 malignant). Features were standardised (zero mean, unit variance) within each training fold, and principal component analysis (PCA) reduced them to four components for the four-qubit encoder. PCA was fit on the training fold only and applied to the test fold to prevent leakage. Across folds, the four retained components captured 79.24% of variance (PC1 44.27%, PC2 18.97%, PC3 9.39%, PC4 6.60%).

3.3. HQCNN Architecture

The overall HQCNN evaluation framework is illustrated in Figure 1. The model comprises: (a) a classical encoder of fully connected layers (4→16→16→4) with ReLU activations and dropout ($p = 0.2$); (b) a four-qubit PQC with RY angle encoding, two variational layers (each: RY and RZ rotations per qubit followed by a circular CNOT entangler), and Pauli-Z expectation readout; and (c) a classical output neuron with sigmoid activation. Total trainable parameters: 441 (420 in the pre-PQC classical layers, 16 in the PQC, 5 in the readout), as illustrated in Figure 2.

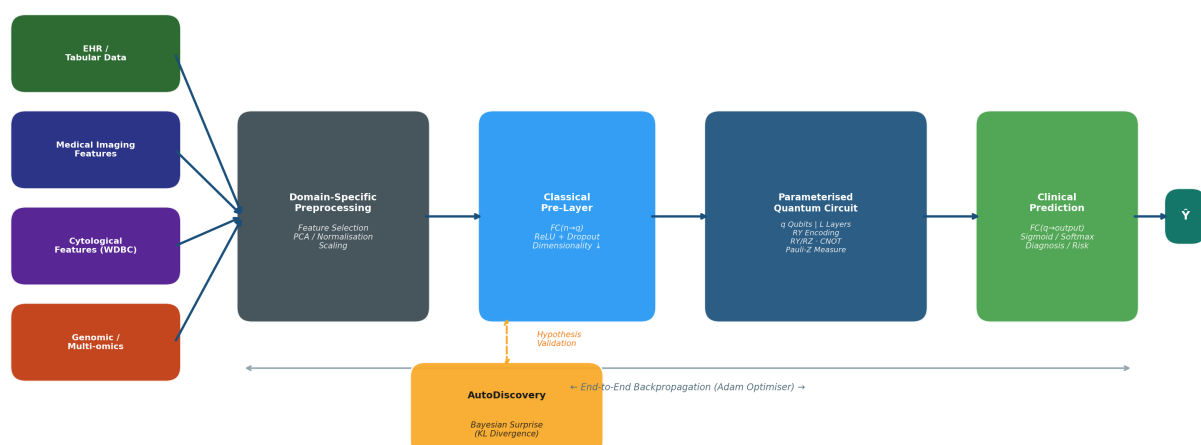


Figure 1. Generalizable HQCNN framework for clinical prediction. The preprocessing module adapts to any data modality; the quantum-classical core stays fixed across instantiations. AutoDiscovery validates design choices post hoc. Arrows indicate the direction of data flow through the pipeline.

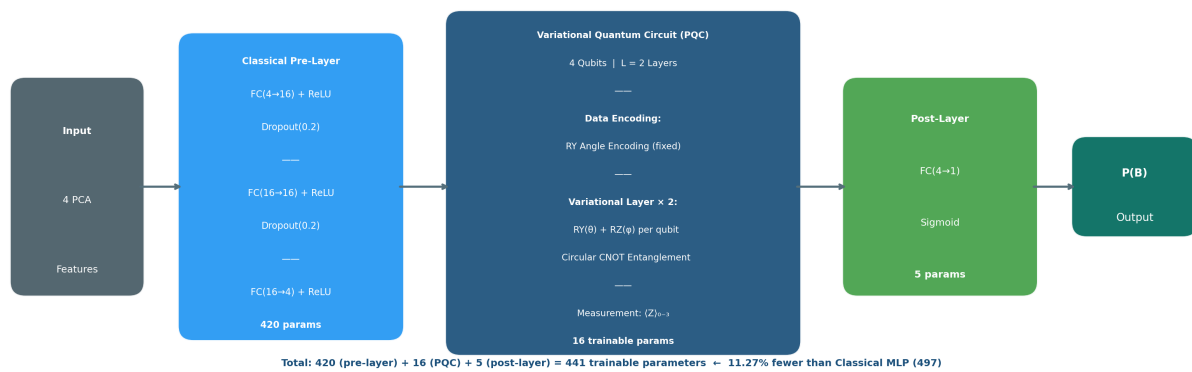


Figure 2. HQCNN architecture. Pre-layer (420 params) + PQC (16 params) + post-layer (5 params) = 441 total trainable parameters. Arrows indicate data flow between the classical encoder, the quantum PQC block, and the classical readout layer.

3.4. Experimental Protocol

3.4.1. Cross-Validation

Five-fold stratified cross-validation with a fixed random seed (42) and per-fold seed offsets. The same fold partition was shared across all models so that paired across-fold comparisons are valid.

3.4.2. Baselines and Controls

Tuned baselines were implemented in scikit-learn and PyTorch: SVM (RBF kernel), random forest, XGBoost [23], and a classical MLP (497 parameters). We additionally include a parameter-matched classical control: a classical MLP with hidden layers [28,10] totalling exactly 441 trainable parameters—identical to the HQCNN—trained and evaluated on the same folds. This isolates the quantum component from raw parameter count. Hyperparameters were selected by grid/random search on the training fold only.

3.4.3. Training

HQCNN training used Adam (learning rate 0.01), ReduceLROnPlateau (factor 0.5, patience 5), binary cross-entropy loss, batch size 32, and early stopping (patience 12) on validation loss.

The depth ablation was run across all five stratified folds for $L \in \{1, 2, 3\}$ variational layers, reporting mean $\pm$ SD accuracy and AUC per depth, with paired tests between depths. This replaces an earlier single-split analysis.

3.4.4. Evaluation Metrics

Accuracy, ROC-AUC, F1-score, precision, and recall, reported as fold mean $\pm$ standard deviation (sample SD, ddof = 1).

3.4.5. Statistical Analysis

Following recommended practice for comparing classifiers across resamples [24], paired comparisons between the HQCNN and each baseline used the Wilcoxon signed-rank test as the primary test ($n = 5$ folds violates the t -test normality assumption) and the paired t -test as a secondary check. Multiple comparisons were corrected with the Holm procedure [25]. We report 95% confidence intervals for mean accuracy and AUC via the t -distribution. Levene's test assesses equality of fold-to-fold variance between the HQCNN and the strongest baseline.

3.4.6. Noisy-Simulator Protocol

To probe robustness, the two-layer HQCNN was re-evaluated on a held-out fold using PennyLane's default mixed density-matrix backend with a depolarising channel ($p = 0.01$) after each parameterised rotation and amplitude damping ($\gamma = 0.01$) after each entangling gate.

3.4.7. Computational Environment and Reproducibility

All experiments ran CPU-only; each HQCNN fold required approximately 3–5 min. Software: Python 3, PennyLane 0.45.0, PyTorch 2.x, scikit-learn 1.x, NumPy 2.x, SciPy 1.x, XGBoost 3.x. A fixed seed (42) with per-fold offsets controls stochasticity. The analysis code, per-fold scores, depth-ablation logs, and noisy-simulation script are available in the project repository (see Data Availability Statement).

4. Results

4.1. Primary Comparison

The HQCNN reached $96.49 \pm 1.96\%$ accuracy and $99.44 \pm 0.60\%$ ROC-AUC (per-fold accuracy: 93.86, 95.61, 96.49, 97.37, 99.12). The best fold achieved 99.12% accuracy. Full comparative results are presented in Table 1.

Table 1. Five-fold cross-validation performance (mean $\pm$ SD, ddof = 1). The parameter-matched MLP has identical parameter count to the HQCNN.

Model	Acc. (%)	F1 (%)	AUC (%)	Params
HQCNN ($L = 2$)	96.49 ± 1.96	97.18 ± 1.63	99.44 ± 0.60	441
SVM (tuned, RBF)	96.14 ± 1.81	96.91 ± 1.35	99.29 ± 0.85	—
Param.-matched MLP	95.08 ± 1.81	—	99.04 ± 0.52	441
Classical MLP	94.91 ± 2.18	96.00 ± 1.49	98.90 ± 0.87	497
Random Forest (tuned)	94.91 ± 2.80	95.90 ± 2.09	98.27 ± 1.50	—
XGBoost (tuned)	93.50 ± 1.46	94.81 ± 0.94	98.68 ± 1.07	—

At equal parameter count (441), the HQCNN exceeded the matched MLP by +1.41 pp accuracy and +0.40 pp AUC, but the difference was not statistically significant (paired t : $p = 0.056$; Wilcoxon: $p = 0.125$). The improvement over the 497-parameter MLP therefore cannot be cleanly attributed to the quantum component. Figure 3 shows the mean accuracy with standard deviation across all models, and the confusion matrix for the best fold is shown in Figure 4. Table 2 and Figure 5 show the full statistical comparison.

Table 2. Statistical comparison of HQCNN versus each baseline (5 paired folds). Holm–Bonferroni correction applied to Wilcoxon (primary) p -values. No comparison survives correction (all adjusted $p \geq 0.625$).

Comparison	Δ Acc. (pp)	Paired t (p)	Wilcoxon (p)	Holm adj. (p)
vs. SVM (tuned)	+0.35	0.178	0.500	0.750
vs. Param.-matched MLP	+1.41	0.056	0.125	0.625
vs. Classical MLP (497)	+1.58	0.104	0.125	0.625
vs. Random Forest	+1.58	0.330	0.375	0.750
vs. XGBoost	+2.99	0.039	0.125	0.625

No comparison survives Holm–Bonferroni correction (all adjusted $p \geq 0.625$). The 95% accuracy confidence intervals overlap: HQCNN 96.49% [94.06, 98.92]; SVM 96.14% [93.88,

98.39]; matched MLP 95.08% [92.83, 97.33]. We therefore report the HQCNN as competitive with, not statistically superior to, the tuned classical baselines. Levene's test for equal fold-to-fold accuracy variance between the HQCNN and SVM gave $p = 0.998$: the variances are statistically indistinguishable. The original claim of “tighter variance/greater stability” is not supported and has been removed.

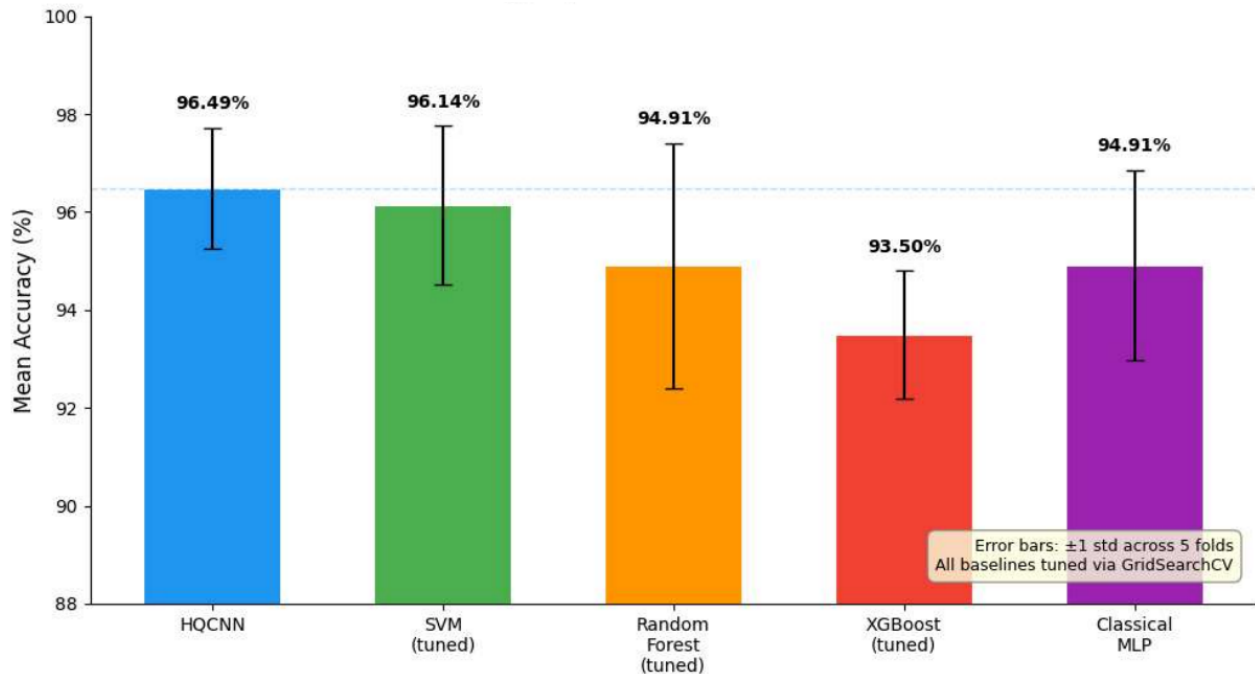


Figure 3. Mean accuracy (± 1 SD) across 5-fold stratified cross-validation for all models. The blue dashed line indicates the HQCNN mean accuracy as a visual reference baseline.

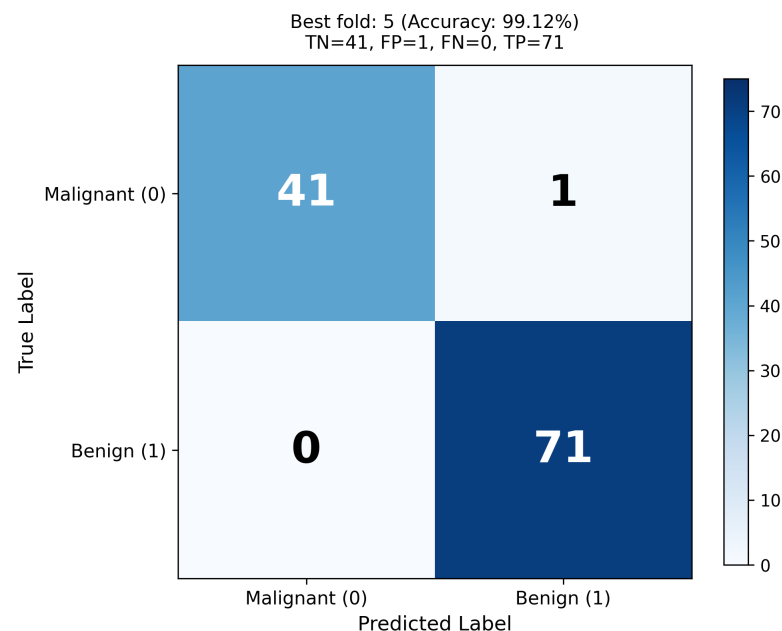


Figure 4. Confusion matrix, HQCNN best fold (Fold 5; $n = 113$; Accuracy = 99.12%). TN = 41, FP = 1, FN = 0, TP = 71. Zero false negatives in this fold.

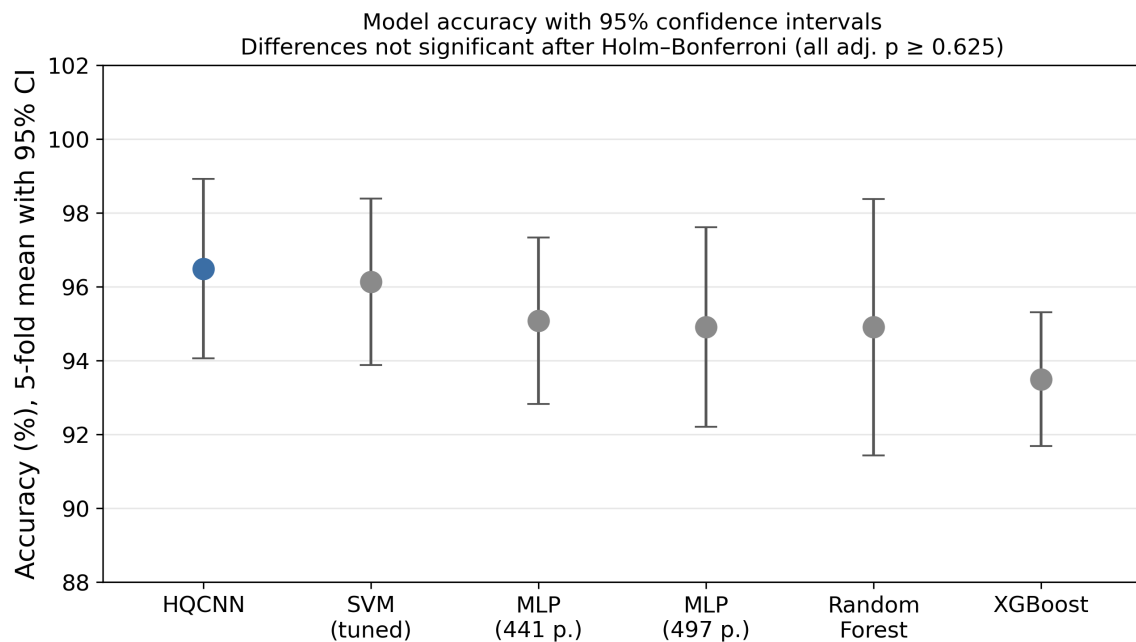


Figure 5. Model accuracy with 95% confidence intervals (5-fold, t -distribution). Each colored dot represents a distinct classifier model. All confidence intervals overlap and no comparison is statistically significant after Holm–Bonferroni correction (all adjusted $p \geq 0.625$). The parameter-matched MLP (441 params) is shown separately from the 497-parameter MLP baseline.

4.2. Circuit-Depth Ablation

Across five folds, accuracy was $96.66 \pm 1.30\%$ ($L = 1$), $96.31 \pm 2.27\%$ ($L = 2$), and $96.67 \pm 1.57\%$ ($L = 3$); AUC was $99.57 \pm 0.28\%$, $99.34 \pm 0.48\%$, and $99.46 \pm 0.43\%$, respectively (Table 3; Figure 6). The $L = 2$ vs. $L = 3$ difference is not significant (Wilcoxon $p = 1.00$; paired t : $p = 0.587$). Depth in this shallow regime does not materially affect performance; we adopt $L = 2$ as a default, not an optimum.

Table 3. Multi-split depth ablation (5-fold cross-validation, mean $\pm$ SD). No statistically significant difference between depths (all Wilcoxon $p \geq 0.625$).

Depth (L)	Accuracy (%)	AUC (%)	Params
$L = 1$	96.66 ± 1.30	99.57 ± 0.28	433
$L = 2$	96.31 ± 2.27	99.34 ± 0.48	441
$L = 3$	96.67 ± 1.57	99.46 ± 0.43	449

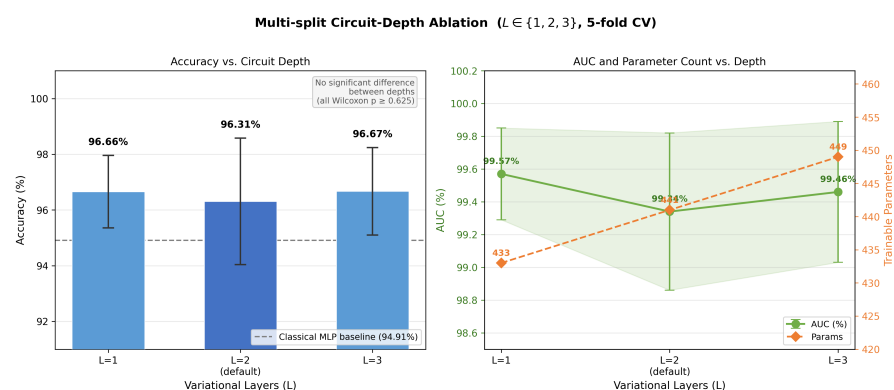


Figure 6. Multi-split circuit-depth ablation ($L \in \{1, 2, 3\}$, 5-fold). Mean accuracy (bars) $\pm$ SD. No statistically significant difference between depths; $L = 2$ is adopted as a practical default rather than an identified optimum.

4.3. Feature Structure

The first principal component (44.27% variance) loads most heavily on size- and shape-related features—mean concave points (0.261), mean concavity (0.258), worst concave points (0.251)—consistent with the clinical understanding that lesion size and irregularity drive malignancy. This component structure was surfaced by the AutoDiscovery pass [20] and is a confirmatory observation about the data, not a predictive mechanism of the model.

4.4. Epistemic-Informativeness Analysis

Bayesian-surprise values computed by numerical KL integration over Beta priors/posteriors (Table 4; Figure 7) reproduce the ordering $H1 \approx 0.92 > H3 \approx 0.80 > H2 \approx 0.65 > H5 \approx 0.30 > H4 \approx 0.18$ nats. The ranking is stable; the absolute nat values should be read ordinally, not as calibrated probabilities. These values quantify how much each finding updates prior expectation and do not influence any prediction.

Table 4. Bayesian-surprise analysis of architectural hypotheses (KL divergence in nats; used as epistemic-informativeness scores only—not as predictions). All priors derived from published literature or domain knowledge.

Hypothesis	Prior	Posterior	Prior Mean	KL (nats)
H1: HQCNN achieves competitive accuracy ($\geq 93\%$) vs. WDBC QML literature	Beta (47, 3)	Beta (596, 23)	0.940	0.92
H2: HQCNN uses fewer parameters than the comparable MLP	Beta (5, 5)	Beta (10, 5)	0.500	0.65
H3: Circuit depth $L \in \{1, 2, 3\}$ has no detectable effect on accuracy in this shallow regime	Beta (3, 3)	Beta (8, 3)	0.500	0.80
H4: PC1 explains $>40\%$ of dataset variance	Beta (8, 12)	Beta (21, 29)	0.400	0.18
H5: HQCNN performance matches tuned SVM	Beta (3, 7)	Beta (6, 9)	0.300	0.30

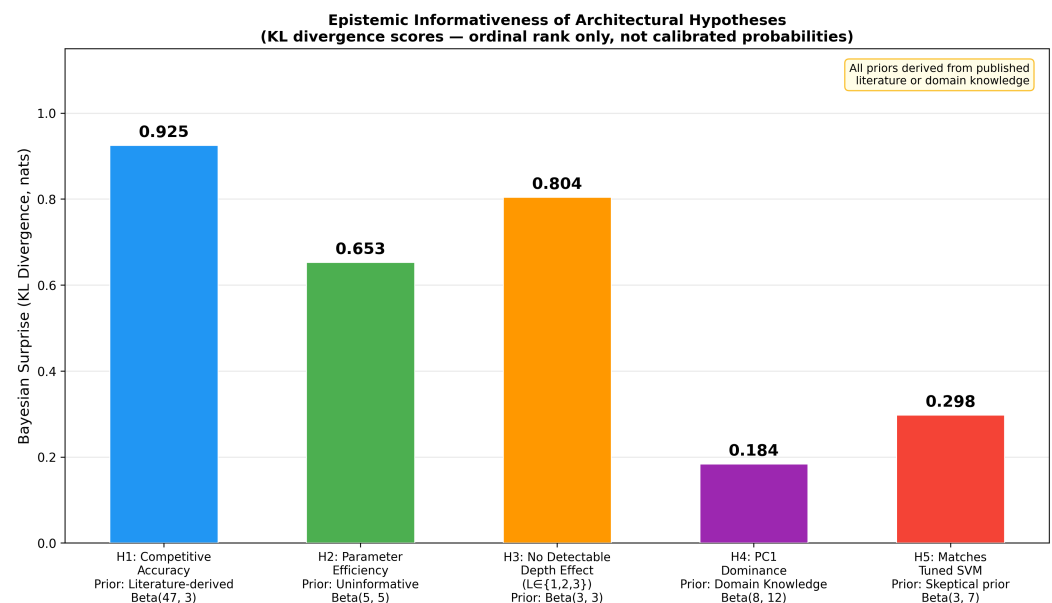


Figure 7. KL divergence (epistemic-informativeness scores) for five architectural hypotheses. H1 (0.92 nats) and H3 (0.80 nats) generate the highest scores. Scores are ordinal informativeness indicators, not calibrated probabilities.

4.5. Noisy-Simulator Result

Under the low-noise model (depolarising $p = 0.01$, amplitude damping $\gamma = 0.01$), the two-layer HQCNN reached 96.49% accuracy, 99.71% AUC, and 97.10% F1 on the held-out fold—essentially unchanged from the noiseless result on the same fold. This indicates robustness at modest, uniform noise rates only; real NISQ hardware has higher, device-specific error and the result should not be read as evidence of hardware readiness.

5. Discussion

5.1. Interpreting the Comparison

The central, honestly reported result is that the HQCNN is competitive with strong tuned classical baselines on WDBC but not significantly better, and that at matched parameter count its small edge over a classical MLP is not significant. The evidence does not establish that the quantum component is responsible for any performance difference. The contribution is the evaluation protocol—shared folds, matched control, significance testing—not a demonstrated quantum advantage.

The clinical implications of parameter efficiency remain noteworthy: fewer parameters reduce overfitting risk on small datasets common in rare-disease settings, lower deployment cost, and produce models that are easier to audit. However, these are potential advantages conditional on broader validation, not demonstrated properties of this single-dataset study.

5.2. Depth and Trainability

The multi-split ablation shows depth is not a sensitive hyperparameter in $L \in \{1, 2, 3\}$. We make no claim about barren-plateau avoidance: barren plateaus emerge at larger qubit counts and depths than tested here [15,16], so a shallow sweep cannot and does not probe them.

5.3. Role of the Epistemic Analysis

The Bayesian-surprise analysis is confirmatory and post hoc; it ranks findings by informativeness and is fully separated from prediction. We did not formally evaluate interpretability (no attribution or SHAP study was performed), so we make no interpretability-advantage claim.

5.4. Relation to Prior Work

Our results align with the broader literature in which hybrid models are competitive on tabular medical benchmarks without a clear advantage [7,11]. Competitive accuracy has been reported across oncology [26], ophthalmology [6], cardiology [27], and heart disease prediction [28,29], consistent with the pattern documented in systematic reviews [7]: hybrid models match, but do not consistently exceed, tuned classical counterparts.

5.5. Limitations

(i) External validity is untested: results are from a single dataset and constitute a proof-of-concept, not a generalisation claim. (ii) Simulation only: all primary results are from a noiseless quantum simulator; the low-noise simulation shows robustness only at small uniform error rates, and realistic hardware noise is higher. (iii) Clinical readiness is not demonstrated: data heterogeneity across sites and equipment, prospective external validation, and regulatory pathways (e.g., software-as-a-medical-device considerations) remain prerequisites for clinical deployment. (iv) Interpretability not evaluated: no formal attribution or SHAP analysis was performed.

5.6. Future Work

Priorities are: (i) multi-dataset external validation, a prerequisite for any generalisation claim; (ii) evaluation on physical NISQ hardware with realistic noise; and (iii) formal interpretability evaluation with attribution metrics. Clinical-governance and auditability benefits are potential implications conditional on external validation.

6. Conclusions

We presented a reproducible evaluation framework for hybrid quantum-classical classifiers and demonstrated it as a single-dataset proof-of-concept on WDBC. The HQCNN performed competitively with tuned classical baselines, but no comparison reached statistical significance after Holm–Bonferroni correction, and at matched parameter count its advantage was not significant. Circuit depth in $L \in \{1, 2, 3\}$ had no detectable effect, so we adopt two layers as a default rather than an optimum. The framework offers a documented, shared-fold procedure that can support cumulative, fair comparison of hybrid quantum-classical models in healthcare.

Author Contributions: Conceptualization, K.M., R.Y., and E.D.L.C.; methodology, R.Y., S.R., and G.S.S.; software, K.M., G.S.S., and E.D.L.C.; validation, K.M., R.Y., and S.R.A.; formal analysis, M.H.M., K.M., and S.R.; investigation, H.G., K.M., and E.D.L.C.; resources, R.Y., G.S.S., and H.G.; data curation, H.G., G.S.S., and R.Y.; writing—original draft preparation, K.M., S.R.A., and H.G.; writing—review and editing, K.M., H.G., and E.D.L.C.; visualization, M.H.M., S.R., and S.R.A.; supervision, K.M., S.R.A., and E.D.L.C.; project administration, K.M., E.D.L.C., and S.R.; funding acquisition, S.R., M.H.M., and H.G. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding, and the APC was funded by the authors.

Informed Consent Statement: This study used the publicly available, fully de-identified Wisconsin Diagnostic Breast Cancer dataset and involved no human participants or animals; therefore no ethics approval or informed consent was required.

Data Availability Statement: The WDBC dataset is publicly available from the UCI Machine Learning Repository and via scikit-learn’s `load_breast_cancer`. The analysis code, per-fold scores, depth-ablation logs, and noisy-simulation script supporting this study are openly available in the project repository. The open-source dataset <https://archive.ics.uci.edu/dataset/17/breast+cancer+wisconsin+diagnostic> (accessed on 23 February 2026).

Acknowledgments: The authors acknowledge the support and resources provided by the University of the Cumberland, Purdue University Global, and University of Central Missouri. During the preparation of this manuscript, the authors used a large language model (GPT-class assistant (GPT-5.4)) for language editing and to assist with drafting the reproducible analysis code. The authors have reviewed and edited all output and take full responsibility for the content of this publication, in accordance with MDPI AI disclosure policy.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

QML	Quantum Machine Learning
HQCNN	Hybrid Quantum-Classical Neural Network
PQC	Parameterized Quantum Circuit
NISQ	Noisy Intermediate-Scale Quantum
PCA	Principal Component Analysis

MLP	Multi-Layer Perceptron
SVM	Support Vector Machine
RF	Random Forest
XGBoost	Extreme Gradient Boosting
AUC	Area Under the ROC Curve
KL	Kullback–Leibler (divergence)
WDBC	Wisconsin Diagnostic Breast Cancer
EHR	Electronic Health Record
CI	Confidence Interval
SD	Standard Deviation
AI	Artificial Intelligence
ML	Machine Learning
ROC	Receiver Operating Characteristic

References

1. Cerezo, M.; Arrasmith, A.; Babbush, R.; Benjamin, S.C.; Endo, S.; Fujii, K.; Coles, P.J. Variational quantum algorithms. *Nat. Rev. Phys.* **2021**, *3*, 625–644. [\[CrossRef\]](#)
2. Bharti, K.; Cervera-Lierta, A.; Kyaw, T.H.; Haug, T.; Alperin-Lea, S.; Anand, A.; Aspuru-Guzik, A. Noisy intermediate-scale quantum algorithms. *Rev. Mod. Phys.* **2022**, *94*, 015004. [\[CrossRef\]](#)
3. Schuld, M.; Killoran, N. Quantum machine learning in feature Hilbert spaces. *Phys. Rev. Lett.* **2019**, *122*, 040504. [\[CrossRef\]](#) [\[PubMed\]](#)
4. Havlíček, V.; Córcoles, A.D.; Temme, K.; Harrow, A.W.; Kandala, A.; Chow, J.M.; Gambetta, J.M. Supervised learning with quantum-enhanced feature spaces. *Nature* **2019**, *567*, 209–212. [\[CrossRef\]](#) [\[PubMed\]](#)
5. Topol, E.J. High-performance medicine: The convergence of human and artificial intelligence. *Nat. Med.* **2019**, *25*, 44–56. [\[CrossRef\]](#) [\[PubMed\]](#)
6. Ara, T.; Mishra, V.P.; Bali, M.; Yenikar, A. Hybrid quantum-classical deep learning framework for balanced multiclass diabetic retinopathy classification. *MethodsX* **2025**, *15*, 103605. [\[CrossRef\]](#) [\[PubMed\]](#)
7. Gupta, R.S.; Wood, C.E.; Engstrom, T.; Pole, J.D.; Shrapnel, S. A systematic review of quantum machine learning for digital health. *npj Digit. Med.* **2025**, *8*, 237. [\[CrossRef\]](#) [\[PubMed\]](#)
8. Benedetti, M.; Lloyd, E.; Sack, S.; Fiorentini, M. Parameterized quantum circuits as machine learning models. *Quantum Sci. Technol.* **2019**, *4*, 043001. [\[CrossRef\]](#)
9. Mari, A.; Bromley, T.R.; Izaac, J.; Schuld, M.; Killoran, N. Transfer learning in hybrid classical-quantum neural networks. *Quantum* **2020**, *4*, 340. [\[CrossRef\]](#)
10. Sim, S.; Johnson, P.D.; Aspuru-Guzik, A. Expressibility and entangling capability of parameterized quantum circuits for hybrid quantum-classical algorithms. *Adv. Quantum Technol.* **2019**, *2*, 1900070. [\[CrossRef\]](#)
11. Kaveh, S.; Arezi, E.; Khedri, Z.; Sohrabei, S. Investigating the application of quantum machine learning in breast cancer: A systematic review. *Arch. Breast Cancer* **2025**, *12*, 130–142. [\[CrossRef\]](#)
12. Wang, Y.; Liu, J. A comprehensive review of quantum machine learning: From NISQ to fault tolerance. *Rep. Prog. Phys.* **2024**, *87*, 117901. [\[CrossRef\]](#) [\[PubMed\]](#)
13. Ullah, U.; Garcia-Zapirain, B. Quantum machine learning revolution in healthcare: A systematic review of emerging perspectives and applications. *IEEE Access* **2024**, *12*, 11423–11450. [\[CrossRef\]](#)
14. Maheshwari, D.; Garcia-Zapirain, B.; Sierra-Sosa, D. Quantum machine learning applications in the biomedical domain: A systematic review. *IEEE Access* **2022**, *10*, 80463–80484. [\[CrossRef\]](#)
15. Larocca, M.; Thanasilp, S.; Wang, S.; Sharma, K.; Biamonte, J.; Coles, P.J.; Cincio, L.; McClean, J.R.; Holmes, Z.; Cerezo, M. Barren plateaus in variational quantum computing. *Nat. Rev. Phys.* **2025**, *7*, 174–189. [\[CrossRef\]](#)
16. Cerezo, M.; Sone, A.; Volkoff, T.; Cincio, L.; Coles, P.J. Cost function dependent barren plateaus in shallow parametrized quantum circuits. *Nat. Commun.* **2021**, *12*, 1791. [\[CrossRef\]](#) [\[PubMed\]](#)
17. Wang, S.; Fontana, E.; Cerezo, M.; Sharma, K.; Sone, A.; Cincio, L.; Coles, P.J. Noise-induced barren plateaus in variational quantum algorithms. *Nat. Commun.* **2021**, *12*, 6961. [\[CrossRef\]](#) [\[PubMed\]](#)
18. Arrasmith, A.; Cerezo, M.; Czarnecki, P.; Cincio, L.; Coles, P.J. Effect of barren plateaus on gradient-free optimization. *Quantum* **2021**, *5*, 558. [\[CrossRef\]](#)

19. Pesah, A.; Cerezo, M.; Wang, S.; Volkoff, T.; Sornborger, A.T.; Coles, P.J. Absence of barren plateaus in quantum convolutional neural networks. *Phys. Rev. X* **2021**, *11*, 041011. [[CrossRef](#)]
20. Agarwal, D.; Majumder, B.P.; Adamson, R.; Chakravorty, M.; Gavireddy, S.R.; Parashar, A.; Surana, H.; Mishra, B.D.; McCallum, A.; Sabharwal, A.; et al. AutoDiscovery: Open-ended scientific discovery via Bayesian surprise. *Adv. Neural Inf. Process. Syst. (NeurIPS)* **2025**, *38*, 25181–25219.
21. Street, W.N.; Wolberg, W.H.; Mangasarian, O.L. Nuclear feature extraction for breast tumor diagnosis. *IS&T/SPIE Symp. Electron. Imaging* **1993**, *1905*, 861–870. [[CrossRef](#)]
22. Wolberg, W.; Mangasarian, O.; Street, N.; Street, W. Breast Cancer Wisconsin (Diagnostic) [Dataset]. UCI Machine Learning Repository. 1993. Available online: <https://archive.ics.uci.edu/dataset/17/breast+cancer+wisconsin+diagnostic> (accessed on 23 February 2026).
23. Chen, T.; Guestrin, C. XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*; ACM: New York, NY, USA, 2016; pp. 785–794. [[CrossRef](#)]
24. Demšar, J. Statistical comparisons of classifiers over multiple data sets. *J. Mach. Learn. Res.* **2006**, *7*, 1–30.
25. Holm, S. A simple sequentially rejective multiple test procedure. *Scand. J. Stat.* **1979**, *6*, 65–70.
26. Senokosov, A.; Sedykh, A.; Sagingalieva, A.; Kyriienko, O.; Vinokur, V.M. Quantum machine learning for image classification. *Mach. Learn. Sci. Technol.* **2024**, *4*, 015028. [[CrossRef](#)]
27. Decoodt, P.; Liang, T.J.; Bopardikar, S.; Santhanam, H.; Eyembe, A.; Garcia-Zapirain, B.; Sierra-Sosa, D. Hybrid classical–quantum transfer learning for cardiomegaly detection in chest X-rays. *J. Imaging* **2023**, *9*, 128. [[CrossRef](#)] [[PubMed](#)]
28. Kumar, A.; Dhanka, S.; Sharma, A.; Bansal, R.; Fahlevi, M.; Rabby, F.; Aljuaid, M. A hybrid framework for heart disease prediction using classical and quantum-inspired machine learning. *Sci. Rep.* **2025**, *15*, 25040. [[CrossRef](#)] [[PubMed](#)]
29. Verdone, A.; Succetti, F.; Ceschini, A.; Rosato, A.; Fioravanti, A.; Panella, M. A hybrid quantum-neural network for heart disease classification. *Biomed. Signal Process. Control* **2026**, *113*, 109185. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.