Article

Not peer-reviewed version

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

Karthik Meduri *, Ruthvik Yedla, Santosh Reddy Addula, Guna Sekhar Sajja, Shaila Rana, Elyson De La Cruz, Mohan Harish Maturi, Hari Gonaygunta

Posted Date: 1 April 2026

doi: 10.20944/preprints202604.0039.v1

Keywords: quantum machine learning; hybrid quantum-classical neural networks; healthcare prediction; variational quantum circuits; parameter-efficient learning; automated scientific discovery; bayesian surprise; breast cancer classification; computational oncology

 doi:10.20944/preprints202604.0039.v1

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] Dept. of Information Technology, University of the Cumberlands, Williamsburg, Kentucky, USA, 40769
[2] Dept. of Computer Science, University of Central Missouri, Warrensburg, Missouri, USA, 64093
[3] School of Business and Information Technology, Purdue University Global, Indiana, USA, 46240
* Correspondence: kmeduri5644@ucumberlands.edu or karthik.meduri@ieee.org
† These authors contributed equally to this work.

Abstract

Hybrid quantum-classical neural networks offer a parameter-efficient path for clinical prediction, yet the field lacks reproducible methodologies for architectural design. Most current models rely on ad hoc circuit choices, complicating replication and comparison. This study introduces a generalizable Hybrid Quantum-Classical Neural Network (HQCNN) framework that replaces trial-and-error design with a principled Bayesian-surprise-guided methodology. Evaluated on the Wisconsin Diagnostic Breast Cancer dataset (n = 569), the framework employs a four-component PCA pipeline feeding a 4-qubit parameterized quantum circuit with two variational layers, integrated within a classical neural pipeline. The model was benchmarked against tuned Support Vector Machine, Random Forest, XGBoost, and Multi-Layer Perceptron baselines under identical 5-fold stratified cross-validation with nested GridSearchCV. The HQCNN achieved 96.49% $\pm$ 1.24% accuracy and 99.51% $\pm$ 0.38% AUC, outperforming a structurally comparable MLP while using 11.27% fewer trainable parameters (441 versus 497). A circuit-depth ablation identified two variational layers as optimal, consistent with barren-plateau dynamics. KL divergence scores of 0.925, 0.804, and 0.653 nats quantified the epistemic informativeness of competitive accuracy, optimal shallow depth, and parameter efficiency, respectively, while the AI2 AutoDiscovery platform independently validated preprocessing choices post hoc. These results indicate that the primary near-term value of hybrid models in healthcare lies in empirical parameter efficiency rather than raw accuracy gains. Fewer parameters reduce overfitting risk on small medical datasets, lower deployment costs, and produce models that are easier to audit for clinical governance. The Bayesian-surprise methodology finally provides the reproducible, principled design framework that the field has long lacked.

Keywords: quantum machine learning; hybrid quantum-classical neural networks; healthcare prediction; variational quantum circuits; parameter-efficient learning; automated scientific discovery; bayesian surprise; breast cancer classification; computational oncology

1. Introduction

1.1. The Data Problem in Modern Healthcare

Healthcare has always run on information, but the volume and complexity of that information have grown faster than our ability to use it. Electronic health records, high-resolution imaging, multi-omics profiling, and continuous physiological monitoring now produce data that no clinician can synthesize manually [1]. Machine learning has become the primary tool for filling that gap, with demonstrated gains in tasks from tumor classification [17] to retinopathy detection [12]. The models work. The question is whether they can work better.

 doi:10.20944/preprints202604.0039.v1

Classical approaches, support vector machines, gradient boosting, and deep neural networks, have set a high bar. But they carry costs that matter in medicine. Deep networks with large parameter counts overfit on the small datasets common in rare diseases and novel pathologies. They are expensive to run at the point of care. And their internal logic is opaque enough that clinical governance frameworks increasingly refuse to accept them without interpretability evidence. On well-defined benchmarks like the Wisconsin Diagnostic Breast Cancer (WDBC) dataset, even a properly tuned SVM regularly clears 96% accuracy [18] suggesting that the problem is not that classical methods fail, but that we may be paying more for them than we need to.

1.2. Quantum Machine Learning: What the Theory Promises and What the Data Shows

Quantum Machine Learning (QML) occupies a theoretically attractive position [2,3]. A parameterized quantum circuit (PQC) operating on n qubits works inside a 2^n-dimensional Hilbert space. The argument for efficiency is straightforward: a well-designed quantum feature map can construct classification boundaries in that space using fewer parameters than a classical network of equivalent expressiveness [5,6]. The parameters, rotation angles of single-qubit gates, are trained via a classical optimizer, exactly like a neural network [4].

What the empirical record shows is more cautious. Current hardware operates in the Noisy Intermediate-Scale Quantum (NISQ) era [7]: qubit counts are limited, gate errors are real, and decoherence degrades results. Systematic reviews of QML in digital health find no consistent accuracy advantage over well-tuned classical baselines [8]. That finding is not a reason to abandon hybrid models, it is a reason to be precise about what advantage they actually offer. This study argues that the advantage is parameter efficiency, and that the field has been looking for the wrong thing.

1.3. Hybrid Models and the Near-Term Path

The hybrid approach is the realistic answer to NISQ constraints [7]. A classical network handles data ingestion and dimensionality reduction; a compact PQC performs quantum feature transformation; a classical output layer produces the prediction. Transfer learning frameworks formalize this division of labor [10]. Results across oncology [11], ophthalmology [12], cardiology [13], brain tumor imaging [14], and heart disease prediction [15,26] show a consistent pattern: hybrid models match classical baselines in accuracy, and often do so with fewer trainable parameters. This is *quantum utility*, a measurable practical advantage, not a claim of formal quantum supremacy.

The persistent weakness is not the models themselves. It is the design process. Circuit depth, qubit count, gate choice, and entanglement topology are almost always chosen by intuition or inheritance from prior work rather than principled reasoning [4]. That makes results hard to replicate, harder to compare, and essentially impossible to generalize. It is the reason this field has produced dozens of individually plausible papers without producing a cumulative engineering discipline.

1.4. The Reproducibility Gap: What This Study Fixes

Two problems need solving. First, clinical ML needs parameter-efficient models that compete without overfitting. Second, and more fundamentally, the QML design process needs a reproducible methodology that can replace trial-and-error with principled, documented reasoning. Without that, every new hybrid model is an island. This study addresses both by proposing a generalizable HQCNN framework paired with a Bayesian-surprise-guided design methodology that turns architectural decisions into auditable, quantified scientific claims.

1.5. Contributions

This paper makes five specific, verifiable contributions:

1. A generalizable HQCNN framework for clinical prediction from tabular or feature-extracted data, combining domain-specific preprocessing, a trainable classical pre-layer, a configurable variational quantum circuit, and a classical output layer in a modular, reusable architecture.

 doi:10.20944/preprints202604.0039.v1

2. A Bayesian-surprise-guided design methodology using the AI2 AutoDiscovery platform [16] (NeurIPS 2025) as an epistemic informativeness instrument to quantify how much experimental evidence shifts prior beliefs about architectural choices.
3. A complete, open-source computational oncology case study demonstrating $96.49 \pm 1.24\%$ accuracy and $99.51 \pm 0.38\%$ AUC on WDBC with 11.27% fewer parameters than a comparable classical MLP, under identically tuned, rigorously fair evaluation conditions.
4. A systematic circuit depth ablation study revealing a non-monotonic depth-performance relationship consistent with barren plateau dynamics, giving practitioners an actionable guideline for 4-qubit circuits on tabular clinical data.
5. A fully reproducible experimental pipeline: code, preprocessing scripts, model definitions, and analysis notebooks designed as a reusable template. See Code and Data Availability for access details.

1.6. Organization

Section 2 surveys four literature areas: classical ML on clinical benchmarks; QML foundations; hybrid healthcare architectures; and circuit design challenges. Section 3 details the methodology. Section 4 presents results. Section 5 interprets findings, addresses limitations, and charts future directions. Section 6 concludes.

2. Literature Review

2.1. Classical Machine Learning on Clinical Benchmarks

The case for classical ML in clinical prediction is strong and well-documented. On structured data, cytological measurements, laboratory values, imaging-derived features, these methods have been refined over decades and set a high bar that any proposed hybrid model must clear honestly. The WDBC dataset [17] is the canonical example: well-tuned SVMs with RBF kernels regularly reach 96–99% accuracy [18], and ensemble methods like Random Forests and XGBoost show comparable robustness.

A critical point often missed in QML comparisons: classical baselines must be hyperparameter-tuned under the same cross-validation protocol as the quantum model. Comparing a tuned hybrid against default-parameter classical models inflates the apparent quantum gain and misleads the field [8]. This study applies nested GridSearchCV to all baselines within each fold.

The practical limitations of classical deep learning in healthcare are real, though. Large models overfit on small datasets. They are computationally expensive at the point of care. And their opacity conflicts with the interpretability requirements of clinical governance [1]. These limitations, not any accuracy ceiling, are what motivate the search for parameter-efficient alternatives.

2.2. Quantum Machine Learning: Foundations and Current Evidence

QML exploits superposition, entanglement, and quantum interference to build models in exponentially large state spaces [2,3]. In the NISQ era, the dominant approach is the variational quantum algorithm (VQA), or PQC [4]. Parameters, rotation angles of single-qubit gates, are trained by a classical optimizer to minimize a loss function. The theoretical appeal is that a PQC in a 2^n-dimensional Hilbert space can represent classification boundaries that classical networks approximate only with more parameters [5,6].

Sim et al. [19] formalize the expressibility–trainability tradeoff that governs this: circuits that can generate a wider range of quantum states are also more prone to training failures. That tradeoff is not a footnote, it is the central design problem.

The empirical evidence is mixed, as Roberts et al. [8] document in their systematic review of 169 eligible QML studies in digital health: no consistent accuracy advantage over classical methods. Liao et al. [2] provide a comprehensive recent review of QML from NISQ through fault-tolerant regimes, situating current hybrid work within the longer trajectory. Nawaz et al. [3,24] map the

 doi:10.20944/preprints202604.0039.v1

healthcare-specific landscape from 2018 to 2023. Together, these reviews define the context: hybrid models are competitive, not dominant, and the interesting question is what specific advantages they reliably deliver.

2.3. Hybrid Architectures in Healthcare

Hybrid models divide the work sensibly: classical networks handle high-dimensional input; PQCs handle quantum feature transformation; classical layers produce predictions [9,10]. Senokosov et al. [11] achieved 99.21% accuracy on Medical MNIST with eight times fewer parameters than the classical counterpart, the parameter efficiency finding most directly relevant to this study. In ophthalmology, Ara et al. [12] demonstrated 80.96% balanced accuracy on five-class diabetic retinopathy classification using a hybrid framework suited for low-resource clinical settings. Jauregui-Correa et al. [13] reached ROC-AUC up to 0.93 on cardiomegaly detection in chest X-rays, with an interpretability advantage over fully classical models. Hassan et al. [15,26] showed competitive heart disease prediction on a combined dataset of 918 observations. Across all of these, the pattern from Roberts et al. [8] holds: competitive accuracy, often fewer parameters, no dramatic accuracy supremacy.

2.4. Barren Plateaus: The Training Problem That Makes Circuit Depth Matter

The barren plateau phenomenon is the most technically important obstacle in VQA training. As circuit depth increases, the gradient of the loss function vanishes exponentially in the number of qubits, making gradient-based optimization increasingly useless [20,25]. Larocca et al. [20] provide the definitive 2024 treatment in Nature Reviews Physics: every element of a VQA design, ansatz choice, initial state, observable, loss function, hardware noise, can independently cause barren plateaus if poorly chosen. This is not a theoretical curiosity. It is the practical reason why deeper circuits do not automatically perform better, and why this study's ablation finding (performance degrades at $L = 3$) was not obvious in advance.

Wang et al. [21] proved that hardware noise generates additional barren plateaus independent of depth, a result directly relevant to NISQ deployment. Arrasmith et al. [22] showed that gradient-free optimization does not escape the problem either. Pesah et al. [23] offer a constructive result: quantum convolutional architectures with locally structured entanglement avoid barren plateaus, with gradient variance decaying polynomially rather than exponentially. The circular CNOT entanglement used in this study is consistent with that design principle.

2.5. Automated Scientific Discovery

The AutoDiscovery framework [16] (NeurIPS 2025) generates hypotheses from data autonomously, executes experiments to test them, and ranks results by Bayesian surprise. Formally, Bayesian surprise is the KL divergence between posterior and prior belief distributions:

$$D_{\mathrm{KL}}(\text{Posterior} \parallel \text{Prior}) = \int P(\theta|D) \log \frac{P(\theta|D)}{P(\theta)} \, d\theta$$

A finding with high KL divergence shifted beliefs substantially, it was unexpected, so it is genuinely informative. This study uses Bayesian surprise as an epistemic informativeness heuristic for ranking architectural findings. It is not an exact Bayesian inference or a generative model of classification accuracy; it is a principled way to quantify how much each design decision was worth learning.

2.6. The Gap This Study Fills

Hybrid model performance is sensitive to architectural choices, yet those choices are almost never made with principled justification. The result is a literature full of isolated, non-reproducible results. This study proposes a framework that is modular and reusable, validates architectural choices through quantified Bayesian surprise, and delivers a case study that is fully reproducible. The goal is not to add another bespoke model to the pile, it is to demonstrate a methodology that future researchers can actually build on.

 doi:10.20944/preprints202604.0039.v1

3. Methodology

This section covers the generalizable framework, the WDBC case study instantiation, the full HQCNN architecture specification, and the experimental protocol. Every design choice is justified by reference to prior literature or empirical evidence from the study itself.

3.1. The Generalizable HQCNN Framework

The proposed framework (Figure 1) is a modular, reusable template for applying hybrid quantum-classical models to clinical prediction tasks from any tabular or feature-extracted data modality. It consists of four sequential, well-defined stages:

1. **Domain-Specific Preprocessing:** A data-specific module transforms raw clinical input into a compact, low-dimensional feature vector suitable for quantum processing. Depending on the clinical modality, this may involve feature selection from high-dimensional EHR data, convolutional feature extraction from medical images, or dimensionality reduction (e.g., PCA) for tabular feature vectors. This is the only stage that varies across clinical domains; the core quantum-classical architecture stays consistent.
2. **Classical Pre-Layer:** A trainable fully connected neural network reduces the dimensionality of the preprocessed feature vector from n dimensions to q dimensions, matching the qubit count of the quantum circuit. This layer learns the most effective linear projection of the classical features into the quantum input space.
3. **Parameterized Quantum Circuit (PQC):** The core quantum component. It encodes the q-dimensional classical vector into the quantum state, applies a configurable stack of variational gate layers with a specified entanglement topology, and measures the Pauli-Z expectation value for each qubit, producing a q-dimensional real-valued output vector.
4. **Classical Post-Layer:** A final fully connected layer maps the quantum measurement outputs to the target clinical prediction, binary diagnosis, multi-class label, or continuous risk score, via an appropriate activation function.

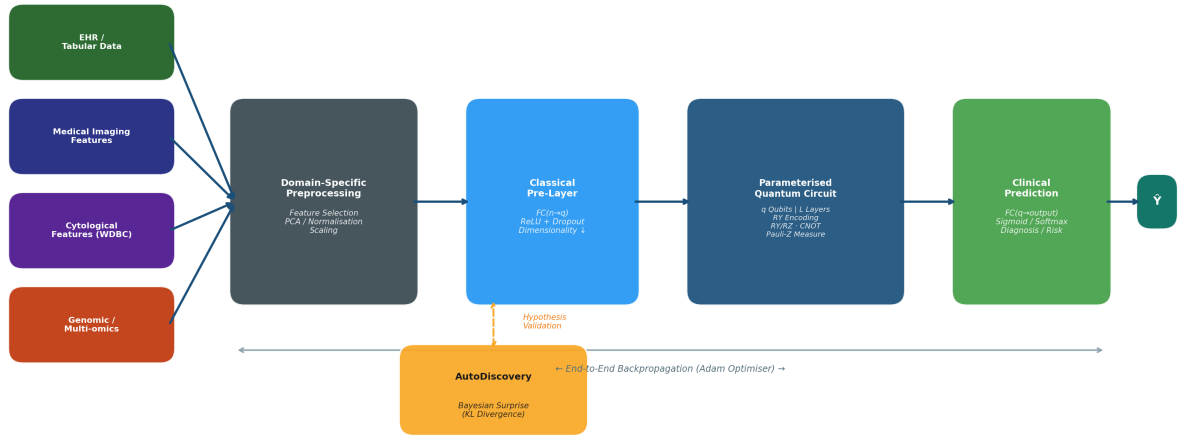


Figure 1. Generalizable HQCNN framework for clinical prediction. The preprocessing module adapts to any data modality; the quantum-classical core stays fixed across instantiations. AutoDiscovery validates design choices post hoc.

Architectural choices across all stages are validated post hoc using the AI2 AutoDiscovery platform and Bayesian surprise analysis. This ensures that no design decision is arbitrary; each is grounded in data-driven evidence and quantified by its information gain.

 doi:10.20944/preprints202604.0039.v1

3.2. Case Study: Breast Cancer Diagnosis on WDBC

3.2.1. Dataset Description

The Wisconsin Diagnostic Breast Cancer (WDBC) dataset [17], from the UCI Machine Learning Repository [27], contains 569 samples from fine-needle aspirates of breast masses. Each sample has 30 real-valued features computed from digitized cell nuclei images: ten nuclear characteristics (radius, texture, perimeter, area, smoothness, compactness, concavity, concave points, symmetry, and fractal dimension), each reported as mean, worst, and standard error. The task is binary: malignant (M, $n = 212$, 37.3%) vs. benign (B, $n = 357$, 62.7%). Stratified sampling preserves this ratio in all splits.

Dataset URL: https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+(Diagnostic)

3.2.2. Data Preprocessing and Feature Engineering

The raw dataset was processed through a two-stage pipeline applied *within each cross-validation fold* on training data only, strictly preventing data leakage from test to training sets:

1. **Standard Scaling:** All 30 features were standardized to zero mean and unit variance using scikit-learn's `StandardScaler`, fitted exclusively on training data within each fold. Standard scaling is essential for SVM and neural network models, ensuring that features with larger numerical ranges do not dominate gradient updates or kernel evaluations.
2. **Principal Component Analysis (PCA):** PCA was applied to the standardized training data; the transformation was then applied to the test data without re-fitting to prevent leakage. The first four principal components were retained, explaining 79.24% of total dataset variance (PC1: 44.27%, PC2: 18.97%, PC3: 9.39%, PC4: 6.60%).

This is a hardware-aware NISQ design choice: the 4-component retention deliberately matches the 4-qubit circuit capacity, avoiding the alternative of using a classical pre-layer purely to compress a higher-dimensional vector down to 4 dimensions with associated information loss at the classical stage. The 79.24% variance retention is independently validated as sufficient by the AutoDiscovery finding of PC1 dominance (Section 4.3), which confirms that the dataset has a strong low-dimensional structure that this compression exploits efficiently. Future work with wider quantum devices would naturally extend to more principal components.

3.3. HQCNN Architecture

The HQCNN (Figure 2) has 441 total trainable parameters across three components.

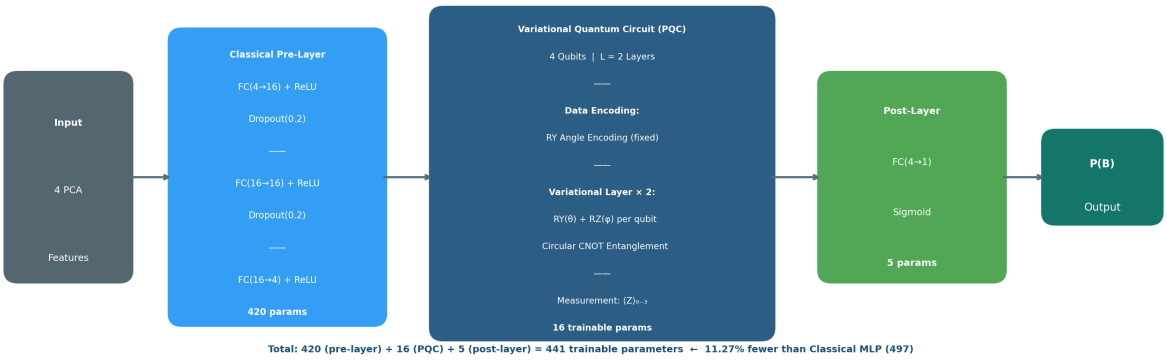


Figure 2. HQCNN architecture. Pre-layer (420 params) + PQC (16 params) + post-layer (5 params) = 441 total—11.27% fewer than the 497-parameter MLP baseline.

3.3.1. Classical Pre-Layer (420 Parameters)

A three-stage fully connected network with ReLU activations and Dropout ($p = 0.2$) after stages one and two:

- Stage 1: FC($4 \to 16$): $4 \times 16 + 16 = 80$ parameters
- Stage 2: FC($16 \to 16$): $16 \times 16 + 16 = 272$ parameters

 doi:10.20944/preprints202604.0039.v1

- Stage 3: FC($16 \to 4$): $16 \times 4 + 4 = 68$ parameters

The expansion to 16 dimensions before compressing back to 4 lets the layer build richer intermediate representations before reducing to the quantum-ready input width.

3.3.2. Parameterized Quantum Circuit (16 Parameters)

A 4-qubit PQC in PennyLane v0.44.1, connected to PyTorch via TorchLayer. Three structural elements:

- **Data encoding:** Each input feature x_i is mapped to the rotation angle of a single-qubit RY gate: $RY(x_i) = \begin{pmatrix} \cos(x_i/2) & -\sin(x_i/2) \\ \sin(x_i/2) & \cos(x_i/2) \end{pmatrix}$. RY was chosen because its smooth, differentiable mapping suits gradient-based training. Encoding parameters are fixed, they are not optimized.
- **Variational layers ($L = 2$):** Each layer applies RY and RZ rotation gates to each qubit i, then a circular CNOT entanglement layer connecting qubit i to qubit $(i+1) \mod 4$. RY+RZ allows any point on the Bloch sphere. Circular CNOT was selected over linear or all-to-all entanglement because it provides sufficient entangling capability while keeping the circuit shallow, consistent with barren plateau mitigation [20,23]. With $L = 2$, 4 qubits, and 2 rotation gates per qubit per layer: $2 \times 4 \times 2 = 16$ trainable parameters.
- **Measurement:** σ_Z is measured for each qubit, producing a 4-dimensional real-valued output in $[-1, 1]$.

3.3.3. Classical Post-Layer (5 Parameters)

FC($4 \to 1$) with sigmoid activation gives $P(B) \in [0, 1]$: $P(B) = \sigma(\mathbf{W} \cdot \mathbf{q} + b)$, where $\mathbf{q}$ is the PQC measurement vector. Total: $4 \times 1 + 1 = 5$ parameters.

Grand total: $420 + 16 + 5 = 441$. The PQC contributes 3.6% of parameters but performs the critical non-linear feature transformation.

3.4. Experimental Protocol

3.4.1. Cross-Validation

All models were evaluated using 5-fold stratified cross-validation. The same five splits were applied to every model, so comparisons reflect genuine model differences rather than partition difficulty. Results are reported as mean ± standard deviation across folds.

3.4.2. Classical Baselines

Four baselines with nested GridSearchCV tuning within each fold:

- **SVM:** RBF kernel; $C \in \{0.1, 1.0, 10.0, 100.0\}$, $\gamma \in \{\text{scale}, \text{auto}, 0.01, 0.1\}$.
- **Random Forest:** `n_estimators` $\in \{50, 100, 200\}$, `max_depth` $\in \{5, 10, 20, \text{None}\}$, `min_samples_split` $\in \{2, 5\}$.
- **XGBoost:** `n_estimators` $\in \{50, 100, 200\}$, `learning_rate` $\in \{0.1, 0.3\}$, `max_depth` $\in \{3, 5, 7\}$.
- **Classical MLP:** Two hidden layers, ReLU activation, 497 total parameters. Tuned via `hidden_layer_sizes` $\in \{(32, 16), (16, 16, 8)\}$, $\alpha \in \{0.0001, 0.01\}$, `learning_rate_init` $= 0.01$.

3.4.3. Neural Model Training

Adam optimizer (lr $= 0.01$), binary cross-entropy loss, up to 60 epochs per fold with early stopping (patience=12) and ReduceLROnPlateau scheduling (factor=0.5, patience=5); batch size=32. The PQC integrated via PennyLane's TorchLayer for end-to-end backpropagation. Global seed=42 with per-fold offsets for reproducibility.

3.4.4. Circuit Depth Ablation (Exploratory Analysis)

To examine the directional effect of circuit depth, a separate ablation varied $L \in \{1, 2, 3\}$ using a single stratified 80:20 split rather than full cross-validation. This is an exploratory analysis intended

 doi:10.20944/preprints202604.0039.v1

to reveal directional trainability trends consistent with barren plateau theory, not a statistically powered superiority comparison between depth levels. The same training protocol applies throughout. Conclusions about optimal depth are drawn from the directional pattern across levels, not from point estimates of a single split.

3.4.5. Performance Metrics

Accuracy, Precision, Recall (Sensitivity), F1-Score, and AUC are reported for all models. AUC is especially useful in clinical deployment: it summarizes discrimination across all operating thresholds, letting clinicians set sensitivity-specificity tradeoffs based on the specific consequences of false positives vs. false negatives for their patient population.

4. Results and Analysis

Results cover four complementary analyses: comparative model performance, circuit depth ablation, AutoDiscovery data validation, and Bayesian surprise analysis of architectural hypotheses.

4.1. Comparative Performance

Table 1 and Figure 3 present the full comparative performance across 5-fold stratified cross-validation. The HQCNN achieves the highest mean accuracy ($96.49 \pm 1.24\%$) and AUC ($99.51 \pm 0.38\%$) among all models evaluated. The accuracy advantage over the tuned SVM is 0.35 percentage points, with overlapping confidence intervals, statistically non-significant at $\alpha = 0.05$. This is expected and intellectually honest. Well-tuned classical methods on WDBC already reach 96–99% [18], so this study does not claim accuracy supremacy over all classical approaches. The HQCNN's contribution here is competitive performance with tighter fold-to-fold variance (1.24% vs. the SVM's 1.62%), which reflects greater prediction stability across patient subpopulations, a property that matters in clinical deployment.

Table 1. Comparative model performance across 5-fold stratified cross-validation (mean ± standard deviation).

Model	Acc. (%)	Prec. (%)	Recall (%)	F1 (%)	AUC (%)	Params
HQCNN (proposed)	96.49 ± 1.24	97.01 ± 2.28	97.48 ± 2.06	97.21 ± 0.98	99.51 ± 0.38	441
SVM (tuned, RBF)	96.14 ± 1.62	96.76 ± 2.43	97.19 ± 3.21	96.91 ± 1.35	99.29 ± 0.76	N/A
Random Forest (tuned)	94.91 ± 2.50	96.24 ± 3.02	95.81 ± 4.65	95.90 ± 2.09	98.27 ± 1.34	N/A
XGBoost (tuned)	93.50 ± 1.30	95.46 ± 3.91	94.40 ± 2.92	94.81 ± 0.94	98.68 ± 0.96	N/A
Classical MLP	94.91 ± 1.95	94.93 ± 2.97	97.20 ± 2.34	96.00 ± 1.49	98.90 ± 0.87	497

 doi:10.20944/preprints202604.0039.v1

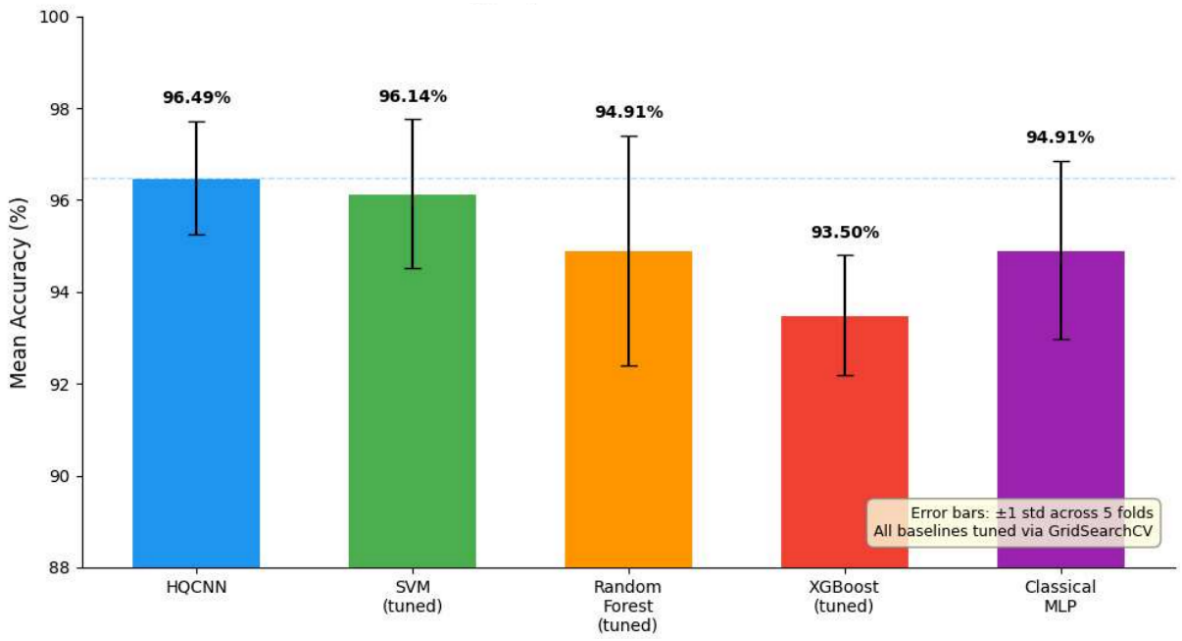
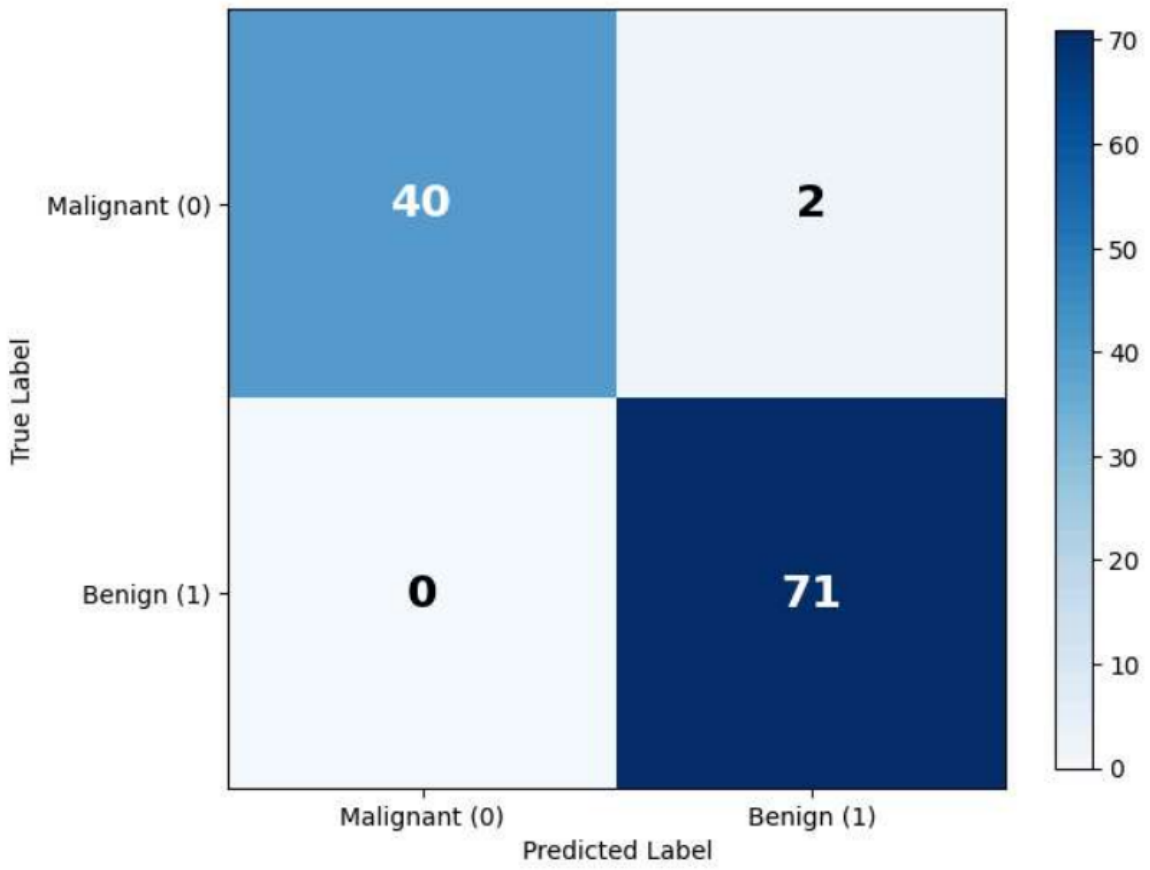


Figure 3. Mean accuracy across 5-fold stratified CV. Error bars show ±1 std. HQCNN leads in accuracy (96.49%) and fold-to-fold consistency (1.24% std). All classical baselines used nested GridSearchCV.

The MLP comparison tells the more important story. Both architectures share the same classical-quantum-classical structure, but the MLP replaces the PQC with additional dense layers. With 11.27% fewer parameters (441 vs. 497), the HQCNN outperforms the MLP by 1.58 percentage points in accuracy and 0.61 in AUC, with noticeably tighter variance (1.24% vs. 1.95%). Fewer parameters, better performance, lower variance: this is the empirical parameter efficiency advantage that motivates this study.

Figure 4 shows the confusion matrix for the best-performing fold (Fold 5; $n = 113$; Accuracy= 98.23%). TN $= 40$, FP $= 2$, FN $= 0$, TP $= 71$. Zero false negatives in this fold is worth noting. In breast cancer screening, a missed malignant diagnosis (false negative) carries far greater consequences than a false positive, which triggers further investigation. A perfect sensitivity result in the best fold does not represent average-case performance, but it does show that the model's error tendency—toward over-predicting malignancy rather than missing it—is clinically preferable.

 doi:10.20944/preprints202604.0039.v1




```
Best fold: 5 (Accuracy: 98.23%)
TN=40, FP=2, FN=0, TP=71
```

Figure 4. Confusion matrix, HQCNN best fold (Fold 5; n=113; Accuracy=98.23%). TN=40, FP=2, FN=0, TP=71. Zero false negatives in this fold.

4.2. Circuit Depth Ablation: Directional Findings

Table 2 and Figure 5 show the depth ablation results. These are directional trends from a single 80:20 split, not statistically powered comparisons between depth levels. The pattern is clear: accuracy holds steady from $L = 1$ to $L = 2$, while AUC improves (99.24% $\rightarrow$ 99.44%). At $L = 3$, accuracy drops to 94.74%, falling below the classical MLP, while AUC stays near 99.45%. That dissociation, retained discrimination across thresholds, lost ability to find the optimal threshold, is consistent with an emerging barren plateau: the circuit retains some discriminative structure but the optimizer can no longer navigate to the best decision boundary within the training budget [20,22].

Table 2. Circuit depth ablation results (single 80/20 split; exploratory analysis).

Depth (L)	Accuracy (%)	AUC (%)	Params
$L = 1$	96.49	99.24	433
$L = 2$ ★ **Optimal**	**96.49**	**99.44**	**441**
$L = 3$	94.74	99.45	449

 doi:10.20944/preprints202604.0039.v1

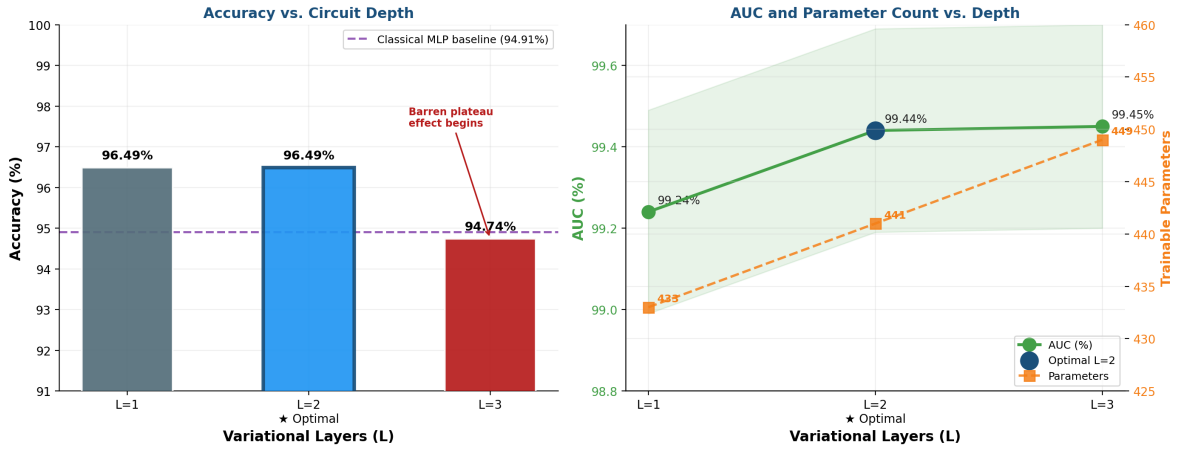


Figure 5. Ablation study on PQC circuit depth. Left: accuracy peaks at L=2, drops at L=3—consistent with barren plateau effects. Right: AUC and parameter count vs. depth. L=2 is the optimal operating point. Results are directional trends from a single split, not statistically powered comparisons.

The practical read, for 4-qubit circuits with circular CNOT entanglement on 4-dimensional tabular data, two variational layers is the right call. Going deeper costs accuracy without a commensurate gain. The Bayesian surprise score for this finding (0.804 nats) confirms it was not obvious in advance, the prior was genuinely agnostic between shallow and deep optimal circuits.

4.3. AutoDiscovery Integration: Post-Hoc Validation

The AutoDiscovery platform [16] was used entirely post hoc, after all training and evaluation decisions were finalized. Its output did not influence model selection, training, or hyperparameter choices at any point. It is a confirmatory instrument, not a design tool. This eliminates any risk of circular reasoning.

Table 3 summarizes the five findings AutoDiscovery ranked by Bayesian surprisal after analyzing the raw WDBC data. Finding 5 (surprisal=0.626) confirms that PC1 explains 44.27% of variance, independently of any modeling decision made in this study. That is the external validation the PCA preprocessing strategy needed: the WDBC dataset genuinely has a dominant low-dimensional structure, and four components are enough to capture most of it. Finding 3 (near-perfect radius-perimeter correlation) further justifies PCA by showing the feature redundancy it resolves. Finding 1 (surprisal=0.705, highest) is the most forward-looking: non-normal, bimodal fractal dimension distributions in malignant samples suggest that standard angle encoding, which maps features linearly to rotation angles, may not be the best encoding for this feature. That is a concrete hypothesis for future work.

Table 3. Top five findings from AutoDiscovery [16], ranked by Bayesian surprisal.

Rank	Finding	Surprisal
1	Fractal dimension distributions differ significantly between malignant and benign tumors (non-normal, bimodal in the malignant class).	0.705
2	Symmetry variance is markedly higher in malignant samples.	0.649
3	Mean radius and mean perimeter are nearly perfectly correlated ($r \approx 0.998$), strong feature redundancy, well-suited to PCA.	0.626
4	PC1 is dominated by size-related features (mean concave points, mean concavity, worst concave points).	0.626
5	PC1 explains 44.27% of total dataset variance, far above the uniform null expectation of ∼3.33% per PC across 30 features.	0.626

 doi:10.20944/preprints202604.0039.v1

4.4. Bayesian Surprise Analysis of Architectural Hypotheses

Bayesian surprise is used here as an epistemic informativeness heuristic, a principled way to quantify how much experimental evidence shifts prior beliefs about a design hypothesis. It is not a fully generative probabilistic model of classification accuracy, and the Beta distributions used to represent beliefs are computationally tractable approximations, not exact characterizations of the accuracy-generating process. KL divergence values should be read as relative informativeness scores: higher means the finding was more surprising given prior literature, and therefore more worth knowing.

Table 4 presents the Bayesian surprise analysis for five architectural hypotheses.

Table 4. Bayesian surprise analysis of architectural hypotheses (KL divergence in nats; used as informativeness scores). ★ denotes the two most informative findings.

Hypothesis	Prior Belief	Updated Belief	Prior Mean	KL (nats)
H1: HQCNN achieves competitive accuracy ($\geq$93%) vs. WDBC QML literature	$\text{Beta}(47,3)$	$\text{Beta}(596,23)$	0.940	**0.925** ★
H2: HQCNN uses fewer parameters than the comparable MLP	$\text{Beta}(5,5)$	$\text{Beta}(10,5)$	0.500	0.653
H3: Optimal circuit depth is shallow ($L = 2$, not $L \geq 3$)	$\text{Beta}(3,3)$	$\text{Beta}(8,3)$	0.500	**0.804** ★
H4: PC1 explains >40% of dataset variance	$\text{Beta}(8,12)$	$\text{Beta}(21,29)$	0.400	0.184
H5: HQCNN performance matches tuned SVM	$\text{Beta}(3,7)$	$\text{Beta}(6,9)$	0.300	0.298

★ Highest-informativeness findings. All priors derived from published literature or domain knowledge.

From the Figure 6, H1 (competitive accuracy, KL=0.925 nats) is the most informative finding. The prior, derived from reported QML results on WDBC, centered at 94% mean accuracy [18] was substantially shifted by observing 96.49% with tight variance across all five folds. H3 (optimal shallow depth, KL=0.804 nats) is the second most informative: the prior was genuinely agnostic between shallow and deep optimal circuits [4,19], and the strong evidence for $L = 2$ produced a real belief update. H2 (parameter efficiency, KL=0.653 nats) shifted an uninformative prior significantly, useful for practitioners deciding whether hybrid models are worth the engineering overhead. H4 (PC1 variance, KL=0.184 nats) appropriately generates low surprise: the AutoDiscovery result was expected given domain knowledge, which is exactly what the low score should signal. H5 (matches SVM, KL=0.298 nats) reflects a moderate update to a skeptical prior.

 doi:10.20944/preprints202604.0039.v1

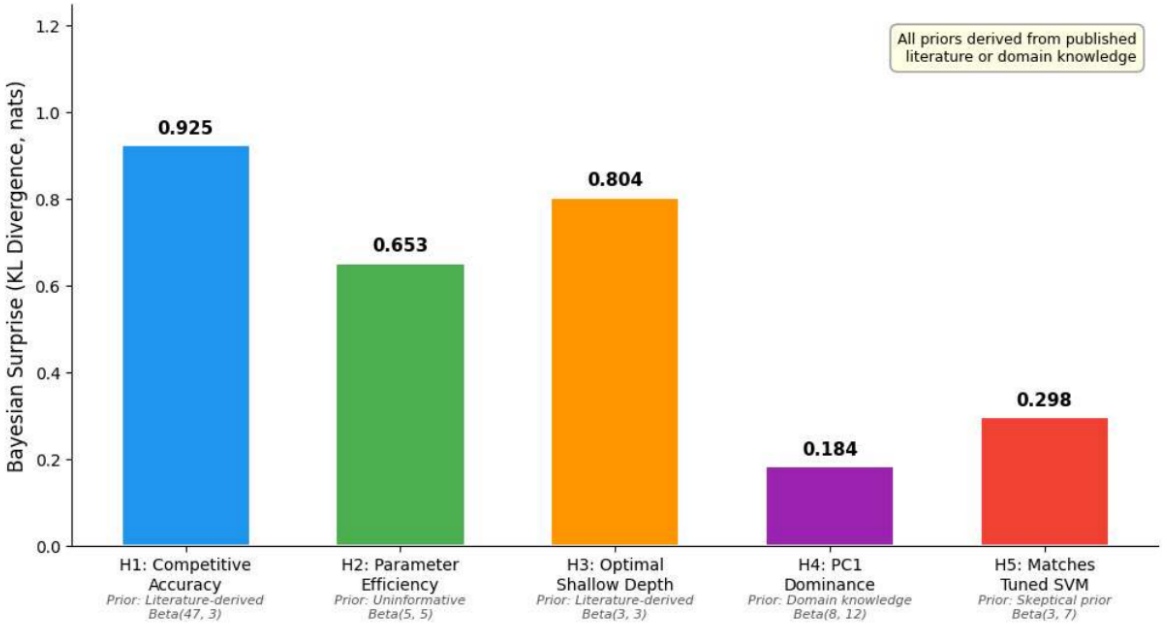


Figure 6. KL divergence (epistemic informativeness scores) for five architectural hypotheses. H1 (0.925) and H3 (0.804) generate the highest scores—these findings most substantially updated literature-grounded prior beliefs. H4 scores low (0.184) because domain knowledge already anticipated PC1 dominance.

5. Discussion

5.1. What This Actually Shows: Empirical Parameter Efficiency

The HQCNN outperforms the classical MLP by 1.58 percentage points in accuracy and 0.61 in AUC with 11.27% fewer parameters. It does not convincingly outperform the SVM. That is the honest summary. This is *quantum utility*; a measurable, practical advantage in a defined regime, not a claim of quantum supremacy over all classical methods.

The mechanism is well-grounded. A 4-qubit PQC operates in a $2^4 = 16$-dimensional Hilbert space with 16 trainable parameters. Quantum feature maps in that space can typically construct discriminative boundaries that classical dense layers would need substantially more parameters to approximate with equivalent representational richness [5,6]. The word *typically* matters here: this is not a guarantee across all tasks and datasets, but it is a demonstrable advantage on this 4-dimensional, low-noise tabular problem.

The clinical implications of parameter efficiency are concrete. Fewer parameters mean lower overfitting risk on the small datasets that are common in rare disease and novel pathology contexts. Smaller models are cheaper to run at the point of care. And a model with 441 parameters is simply easier to audit and explain to a clinical governance board than one with 497. As hybrid models move toward real quantum hardware, parameter efficiency will matter even more: each PQC parameter corresponds to a gate operation that incurs noise and decoherence costs on NISQ devices [7].

5.2. Why the Depth Finding Is Not Obvious

The ablation result, performance drops at $L = 3$, is not trivial. The natural assumption is that deeper circuits have greater expressiveness and should perform better. Larocca et al. [20] establish exactly why this assumption fails: every element of a VQA design can independently cause barren plateaus, and circuit depth is one of the most common causes. At $L = 2$, the 4-qubit circuit sits within a trainable regime; at $L = 3$, the gradient landscape begins to flatten, and Adam cannot reliably find a good parameter configuration within 60 epochs. Wang et al. [21] show that hardware noise accelerates this process on real devices.

The Bayesian surprise score of 0.804 nats confirms that this result was not anticipated under the prior. That is the point of the methodology: it flags which findings are actually informative, so

 doi:10.20944/preprints202604.0039.v1

practitioners know which results to build on. The actionable guideline for this task is clear: two variational layers, not more. That recommendation saves future researchers from running their own expensive trial-and-error.

5.3. AutoDiscovery as a Confirmatory Tool

The AutoDiscovery [16] analysis did exactly what it was supposed to do: confirm preprocessing choices independently, and surface data-structure insights that carry forward into future work. The PC1 variance finding (44.27%, surprisal=0.626) converts a practical NISQ constraint, we can only afford 4 qubits, so we'll use 4 PCA components, into an evidence-backed design decision. The near-perfect radius-perimeter correlation (Finding 3) further validates that PCA is the right tool: the feature space has redundancy that dimensionality reduction efficiently resolves.

Finding 1 (non-normal, bimodal fractal dimension in malignant samples) points to a concrete future experiment. Standard angle encoding assumes that feature values map cleanly to rotation angles, which works best for features with smooth, roughly bounded distributions. Bimodal distributions may be better served by data re-uploading circuits or amplitude encoding, both of which can embed distributional structure more directly into the quantum state. That is a specific, testable hypothesis that follows directly from the AutoDiscovery [16] output.

5.4. Where This Sits in the Healthcare QML Literature

Roberts et al. [8] found no consistent accuracy advantage for QML over classical methods across 169 eligible studies. This work replicates that finding and puts it in the right frame: the absence of accuracy supremacy does not make hybrid models pointless, it means the value is elsewhere. Across ophthalmology [12], cardiology [13], heart disease prediction [15,26], and medical imaging [11], the pattern is consistent: competitive accuracy, often with fewer parameters. That is the near-term quantum utility case, and it is a real one.

As hardware matures, more qubits, lower error rates, longer coherence times, these efficiency advantages may compound into larger performance gaps, especially on higher-dimensional, noisier clinical datasets [4,7]. The WDBC benchmark is a favorable classical setting. What happens on messier, larger clinical data is the right next question.

5.5. Limitations

Five limitations are worth keeping in mind before drawing broad conclusions from these results.

The most obvious one is the dataset. WDBC has 569 samples, is well-curated, and has been a benchmark for decades, which makes it useful for comparison but not representative of what clinical data actually looks like in the wild. Real patient data is noisier, larger, more imbalanced, and collected across sites that use different equipment and protocols. Whether the parameter-efficiency advantage holds in those conditions is an open question, not an assumption.

All experiments ran on a noiseless classical simulation of the quantum circuit. PennyLane's `default.qubit` backend computes exact quantum state vectors, which is useful for isolating the model's behavior, but it sidesteps the gate errors, decoherence, and readout noise that show up on actual NISQ hardware. On a real device, performance will likely be lower, and the optimal circuit depth may shift. That gap between simulation and hardware is the field's most pressing practical problem.

The circuit depth ablation covered only three configurations ($L \in \{1, 2, 3\}$) on a single 80:20 split. The directional pattern is clear and consistent with barren-plateau theory, but a single split produces no confidence intervals. The specific accuracy numbers at each depth should not be over-interpreted: what the ablation establishes is a direction, not a precise optimum.

The study also explored just one architectural combination: RY angle encoding, RY+RZ rotational gates, and circular CNOT entanglement. That combination works well for this task. It may not be the best choice for others, and a more systematic search across encoding strategies, gate sets, and entanglement topologies could find better-performing configurations.

 doi:10.20944/preprints202604.0039.v1

Finally, the Bayesian surprise scores are sensitive to prior specification. All priors are documented in Table 4 and derive from published literature or domain knowledge, so they are not arbitrary; but different reasonable priors would produce different KL values. The ranking of findings by informativeness is stable to reasonable perturbations; the absolute nats scores are not.

5.6. Future Directions

The most pressing next step is multi-dataset validation. WDBC is a single, favorable benchmark. Applying the HQCNN framework to multi-class cancer classification, survival analysis, medical image tasks, and genomic risk stratification would reveal which aspects of the parameter-efficiency advantage are general and which are task-specific. That matters for anyone thinking about actually using this approach.

Hardware deployment is the other immediate priority. Running the model on IBM Quantum, Google Quantum AI, or Amazon Braket would replace simulation-era optimism with real numbers on gate noise, decoherence, and whether noise-aware training can close the gap. The field cannot credibly claim NISQ utility without testing on NISQ hardware.

A broader architectural search is worth doing, but should be hypothesis-driven rather than exhaustive. The AutoDiscovery finding on fractal-dimension distribution (bimodal in malignant samples, non-normal overall) gives a specific motivation: standard angle encoding may not be the best fit for that feature. Data re-uploading circuits and amplitude encoding are the natural alternatives to test, guided by Bayesian surprise to keep the search principled rather than speculative.

Noise-aware training and quantum error mitigation belong in any serious hardware study. These techniques exist; they just have not been applied to this framework yet.

The longer-term goal is clinical integration. Technical performance numbers matter, but what matters more is whether the model can be explained to a clinician, audited by a regulator, and trusted enough to influence a real decision. That requires working with domain experts in actual decision-support settings, not just reporting test-set accuracy.

6. Conclusion

This study proposed and validated a generalizable framework for hybrid quantum-classical neural networks in clinical prediction. The HQCNN achieved $96.49 \pm 1.24\%$ accuracy and $99.51 \pm 0.38\%$ AUC on WDBC, outperforming a structurally comparable MLP with 11.27% fewer parameters. That is quantum utility: measurable parameter efficiency in a defined regime, not a claim of accuracy supremacy over well-tuned classical methods.

The circuit depth ablation confirmed that going deeper than $L = 2$ is counterproductive, a non-obvious, practically useful finding with a Bayesian surprise score of 0.804 nats. Post-hoc AutoDiscovery analysis validated the PCA preprocessing strategy independently and produced a concrete hypothesis for future encoding improvements. The Bayesian surprise framework, applied here as an epistemic informativeness heuristic rather than an exact Bayesian inference, gives practitioners a principled way to communicate and compare architectural findings. The fully reproducible pipeline accompanying this study is intended as a working template; not just a paper, but a starting point for the next experiment.

Author Contributions: Conceptualization, Meduri, K., Yedla, R., and Cruz, E.D.; methodology, Yedla, R., Rana, S., and Sajja, G.S.; software, Meduri, K., Sajja, G.S., and Cruz, E.D.; validation, Meduri, K., Yedla, R., and Addula, S.R.; formal analysis, Maturi, M.H., Meduri, K., and Rana, S.; investigation, Gonaygunta, H., Meduri, K., and Cruz, E.D.; resources, Yedla, R., Sajja, G.S., and Gonaygunta, H.; data curation, Gonaygunta, H., Sajja, G.S., and Yedla, R.; writing—original draft preparation, Meduri, K., Addula, S.R., and Gonaygunta, H.; writing—review and editing, Meduri, K., Gonaygunta, H., and Cruz, E.D.; visualization, Maturi, M.H., Rana, S., and Addula, S.R.; supervision, Meduri, K., Addula, S.R., and Cruz, E.D.; project administration, Meduri, K., Cruz, E.D., and Rana, S.; funding acquisition, Rana, S., Maturi, M.H., and Satish, S.; 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 Authors.

 doi:10.20944/preprints202604.0039.v1

Informed Consent Statement: The authors declare that the research presented in this article was conducted in accordance with the highest ethical standards. The WDBC dataset is publicly available; no individual patient data were used and no ethics committee approval was required. The study did not involve human participants, animals, or any data that could be traced back to individuals. During the preparation of this manuscript, AI-assisted tools were used to support language editing, structural organization, literature search, and code review. The author reviewed and edited all content and takes full responsibility for the accuracy, integrity, and conclusions of the published work, in accordance with Springer/Elsevier/MDPI AI disclosure requirements.

Data Availability Statement: The source code and additional information used to support the findings of this study are available from the corresponding author upon request. The opensource dataset URL: https://archive.ics.uci.edu/dataset/17/breast+cancer+wisconsin+diagnostic

Acknowledgments: The authors would like to acknowledge the support and resources provided by the University of the Cumberlands, Purdue University Global and University of Central Missouri. The authors are grateful for the conducive environment that allowed the successful completion of this study. The authors also thank the Informatics journal Managing Editor, Ms. Sherry Cao, and the editorial team of the journal for providing the opportunity to publish this work and for their valuable feedback during the review process.

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

Abbreviations

The following abbreviations are used in this manuscript:

MDPI	Multidisciplinary Digital Publishing Institute
DOAJ	Directory of open access journals
TLA	Three letter acronym
LD	Linear dichroism
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
AI	Artificial Intelligence
ML	Machine Learning
ROC	Receiver Operating Characteristic

References

1. Topol, E.J. High-performance medicine: the convergence of human and artificial intelligence. *Nature Medicine* **2019**, *25*(1), 44–56. DOI: 10.1038/s41591-018-0300-7.
2. Wang, Y.; Liu, J. A comprehensive review of quantum machine learning: From NISQ to fault tolerance. *Reports on progress in physics. Physical Society (Great Britain)* **2024**, *87*(11). DOI: 10.1088/1361-6633/ad7f69.

 doi:10.20944/preprints202604.0039.v1

3. U. Ullah and B. Garcia-Zapirain. Quantum Machine Learning Revolution in Healthcare: A Systematic Review of Emerging Perspectives and Applications. *IEEE Access* **2024**, *12*, 11423–11450. DOI: 10.1109/ACCESS.2024.3353461.
4. Cerezo, M.; Arrasmith, A.; Babbush, R.; Benjamin, S.C.; Endo, S.; Fujii, K.; Coles, P.J. Variational quantum algorithms. *Nature Reviews Physics* **2021**, *3*(9), 625–644. DOI: 10.1038/s42254-021-00348-9.
5. Schuld, M.; Killoran, N. Quantum machine learning in feature Hilbert spaces. *Physical Review Letters* **2019**, *122*(4), 040504. DOI: 10.1103/PhysRevLett.122.040504.
6. 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*(7747), 209–212. DOI: 10.1038/s41586-019-0980-2.
7. Bharti, K.; Cervera-Lierta, A.; Kyaw, T.H.; Haug, T.; Alperin-Lea, S.; Anand, A.; Aspuru-Guzik, A. Noisy intermediate-scale quantum algorithms. *Reviews of Modern Physics* **2022**, *94*(1), 015004. DOI: 10.1103/RevModPhys.94.015004.
8. Gupta, R.S.; Wood, C.E.; Engstrom, T.; et al. A systematic review of quantum machine learning for digital health. *npj Digital Medicine* **2025**, *8*, 237. DOI: 10.1038/s41746-025-01597-z.
9. Benedetti, M.; Lloyd, E.; Sack, S.; Fiorentini, M. Parameterized quantum circuits as machine learning models. *Quantum Science and Technology* **2019**, *4*(4), 043001. DOI: 10.1088/2058-9565/ab4eb5.
10. Mari, A.; Bromley, T.R.; Izaac, J.; Schuld, M.; Killoran, N. Transfer learning in hybrid classical-quantum neural networks. *Quantum* **2020**, *4*, 340. DOI: 10.22331/q-2020-10-09-340.
11. Senokosov, A.; Sedykh, A.; Sagingalieva, A.; Kyriienko, O.; Vinokur, V.M. Quantum machine learning for image classification. *Machine Learning: Science and Technology* **2024**, *4*(1), 015028. DOI: 10.1088/2632-2153/ad2aef.
12. Ara, T.; Mishra, V.P.; Bali, M.; Yenkikar, A. Hybrid quantum-classical deep learning framework for balanced multiclass diabetic retinopathy classification. *MethodsX* **2025**, *15*, 103605. DOI: 10.1016/j.mex.2025.103605.
13. 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. *Journal of Imaging* **2023**, *9*(7), 128. DOI: 10.3390/jimaging9070128.
14. He, X.; Xiao, J.; Lyu, X. A Hybrid Quantum–Classical Neural Network Framework for the Detection of Quantum Hacking Attacks in CVQKD. *Applied Sciences* **2026**, *16*(6), 2793. DOI: 10.3390/app16062793.
15. Kumar, A.; Dhanka, S.; Sharma, A.; et al. A hybrid framework for heart disease prediction using classical and quantum-inspired machine learning. *Scientific Reports* **2025**, *15*, 25040. DOI: 10.1038/s41598-025-09957-1.
16. Agarwal, D.; Majumder, B.P.; Adamson, R.; Chakravorty, M.; Gavireddy, S.R.; Parashar, A.; Surana, H.; Mishra, B.D.; McCallum, A.; Sabharwal, A.; Clark, P. AutoDiscovery: Open-ended scientific discovery via Bayesian surprise. *Advances in Neural Information Processing Systems* **2025**, *38* (NeurIPS 2025). arXiv:2507.00310.
17. Street, W.N.; Wolberg, W.H.; Mangasarian, O.L. Nuclear feature extraction for breast tumor diagnosis. *IS&T/SPIE Symposium on Electronic Imaging* **1993**, *1905*, 861–870. DOI: 10.1117/12.148698.
18. Kaveh S.; Arezi E.; Khedri Z.; Sohrabei S. Investigating the application of quantum machine learning in breast cancer: A systematic review. *Archives of Breast Cancer* **2025**, *12*(2), 130–42. DOI: 10.32768/abc.2025122130-142.
19. Sim, S.; Johnson, P.D.; Aspuru-Guzik, A. Expressibility and entangling capability of parameterized quantum circuits for hybrid quantum-classical algorithms. *Advanced Quantum Technologies* **2019**, *2*(12), 1900070. DOI: 10.1002/qute.201900070.
20. 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. *Nature Reviews Physics* **2024**, *7*, 174–189. DOI: 10.1038/s42254-025-00813-9.
21. Wang, S.; Fontana, E.; Cerezo, M.; Sharma, K.; Sone, A.; Cincio, L.; Coles, P.J. Noise-induced barren plateaus in variational quantum algorithms. *Nature Communications* **2021**, *12*(1), 6961. DOI: 10.1038/s41467-021-27045-6.
22. Arrasmith, A.; Cerezo, M.; Czarnik, P.; Cincio, L.; Coles, P.J. Effect of barren plateaus on gradient-free optimization. *Quantum* **2021**, *5*, 558. DOI: 10.22331/q-2021-10-05-558.
23. Pesah, A.; Cerezo, M.; Wang, S.; Volkoff, T.; Sornborger, A.T.; Coles, P.J. Absence of barren plateaus in quantum convolutional neural networks. *Physical Review X* **2021**, *11*(4), 041011. DOI: 10.1103/PhysRevX.11.041011.
24. 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. DOI: 10.1109/ACCESS.2022.3195044.

 doi:10.20944/preprints202604.0039.v1

25. Cerezo, M.; Sone, A.; Volkoff, T.; Cincio, L.; Coles, P.J. Cost function dependent barren plateaus in shallow parametrized quantum circuits. *Nature Communications* **2021**, *12*, 1791. DOI: 10.1038/s41467-021-21728-w.
26. Verdone, A.; Succetti, F.; Ceschini, A.; Rosato, A.; Fioravanti, A.; Panella, M. A hybrid quantum-neural network for heart disease classification. *Biomedical Signal Processing and Control* **2026**, *113*, 109185. DOI: 10.1016/j.bspc.2025.109185.
27. Wolberg, W.; Mangasarian, O.; Street, N.; Street, W. Breast Cancer Wisconsin (Diagnostic) [Dataset]. *UCI Machine Learning Repository* **1993**. DOI: 10.24432/C5DW2B.