Evaluating the Data and Feature Efficiency of Variational Quantum Classifiers for Genomic Cancer Diagnostics

Jaszmine DeFranco, Vanisha Swabhanam, Pamela Espinoza Maldonado,
Christo Bobby, Aaron Nguyen
Department of Computer Science, The University of Texas at Dallas

1. Abstract

High-dimensional genomic data with limited patient samples presents a fundamental "small n, large p" challenge for clinical diagnostics, creating a bottleneck for classical machine learning models that typically require large datasets to avoid overfitting. This study investigates whether Variational Quantum Classifiers (VQCs) can achieve competitive performance for classifying Acute Myeloid Leukemia (AML) and Acute Lymphoblastic Leukemia (ALL) while demonstrating superior parameter and feature efficiency compared to classical models on the seminal Golub leukemia dataset [1]. We implemented VQCs using angle [10] and amplitude [11] encoding in PennyLane and Qiskit, benchmarking them against classical Support Vector Machines (SVMs) [8] and Multi-Layer Perceptrons (MLPs) [9]. A systematic experimental matrix was executed to evaluate performance across varying levels of data scarcity and feature set sizes, with a particular focus on embedded feature selection via SCAD regularization.

Our key results show a classical model, a SCAD-SVM, achieved peak performance of 97.7% accuracy using only 4 genes, outperforming other feature selection methods that required more features for lower accuracy. An angle-encoded VQC reached 100% accuracy with an optimal configuration of 4 qubits and 3 layers, but performance was highly sensitive to architecture, dropping with more qubits or layers, indicating optimization challenges such as barren plateaus [3]. Amplitude encoding demonstrated logarithmic qubit efficiency, for instance encoding 32 features into 5 qubits, but struggled with class imbalance, achieving a best accuracy of 85.7%. While quantum models used less than 1% of the parameters of a comparable classical MLP and offered theoretical feature compression [12], they did not yet demonstrate a clear accuracy advantage over the optimally regularized classical baseline.

We conclude that for the specific task of AML vs. ALL classification on the Golub dataset, classical models with advanced regularization currently provide the most accurate and stable solution. Quantum models, while not yet surpassing classical accuracy, show definitive promise in parameter and feature efficiency [5], presenting a compelling pathway for scalable learning in future, higher-dimensional biomedical applications where classical scaling becomes prohibitive.

2. Introduction

The field of computational genomics is fundamentally shaped by the “small n, large p” problem, where the number of features (p), such as gene expression levels, vastly exceeds the number of available

patient samples (n) [1]. This high-dimensional data landscape presents a significant challenge for classical machine learning models, which often require massive datasets to generalize effectively and avoid overfitting. In clinical diagnostics, where acquiring large, labeled datasets is expensive and time-consuming, this necessity for data becomes a critical bottleneck. This limitation was starkly evident during the COVID-19 pandemic, where the urgent need for rapid diagnostic and prognostic models collided with the reality of initially small, fragmented patient cohorts. We tackle this problem using the seminal Golub leukemia dataset [1], a classic benchmark comprising gene expression data from 72 patients and over 7,000 genes, which perfectly encapsulates the "small n, large p" challenge. Quantum Machine Learning (QML) emerges as a promising avenue to address this chronic challenge, leveraging high-dimensional Hilbert space of quantum systems to potentially discover complex patterns with inherently more efficient model parameterizations [2], [3]. The ability to build accurate models from limited data could fundamentally accelerate our response to emerging diseases and improve diagnostics for rare conditions.

While QML holds theoretical promise, its practical superiority over classical models for real-world medical tasks remains an open and actively researched question [4], [5]. Many initial studies focus solely on matching accuracy on full datasets, but this fails to address the core constraints of medical data. This critical question is not just whether QML can match classical performance, but whether it can do so more efficiently. Specifically: Can parameter-efficient QML models mitigate the data hunger and feature dependency of classical artificial intelligence in medicine, thereby offering a more robust and practical solution for data-scarce clinical environments? This work directly addresses this question through a rigorous benchmark on a seminal biomedical dataset. Our central research question is: With only 16 biomarker genomic expressions and 22 patients, can a Quantum Machine Learning (QML) model demonstrate superior parameter, data, and feature efficiency while matching classical model performance for classifying acute myeloid leukemia (AML) and acute lymphoblastic leukemia (ALL)?

We hypothesize that a variational quantum classifier will demonstrate superior performance across three key metrics of efficiency. First, in terms of parameter efficiency, it is expected to use $\leq$1% of the trainable parameters of a comparable classical Multi-Layer Perceptron (MLP) [9]. Second, regarding data efficiency, we anticipate it will retain $\geq$80% of the classical model's full-data AUROC even when trained on only 20% of the available training samples. Finally, for training stability, we predict it will exhibit significantly lower run-to-run variance, maintaining an AUROC standard deviation of $\sigma \leq 0.5\%$ across multiple random seeds, which would indicate greater reliability and robustness.

This study moves beyond existing literature by making several key contributions to the field of medical quantum machine learning. First, it establishes a systematic and reproducible medical benchmark, providing a fair comparison of quantum machine learning (QML) models, implemented via both angle [10] and amplitude [11] encoding, against established classical baselines (SVM [8] and MLP [9]) on the well-known Golub leukemia dataset [1], a cornerstone of bioinformatics. A distinctive aspect of our work is the rigorous, comparative analysis of feature efficiency, quantifying how model performance degrades as the number of genomic biomarkers is systematically reduced. While the concept is known in classical ML [6], [7], its application to benchmarking QML models against classical ones in a medical context still has much room for further exploration. By analyzing performance from the biologically-grounded 50-gene signature (identified in Golub et al. [1]) down to a minimal 16-gene panel, we identify which model architecture is most robust to feature scarcity, a critical consideration for developing cost-effective clinical tests. Furthermore, our work maintains a practical focus by defining key performance indicators around parameter count, data requirements, and training stability, metrics of direct relevance for

developing robust clinical decision-support tools. Finally, to promote reproducibility and future research, we provide publicly available methodology and detailed benchmarks, serving as an open-source foundation for the medical AML community. By framing our analysis around these practical efficiency metrics, this work aims to identify a clear and clinically relevant niche where quantum models provide a tangible advantage, thereby advancing the field from theoretical potential toward practical application.

3. Background & Related Work

The foundational dataset for this project originates from the landmark 1999 study by Golub et al., "Molecular Classification of Cancer: Class Discovery and Class Prediction by Gene Expression Monitoring" [1]. This work represented a paradigm shift in oncology, moving cancer diagnosis beyond traditional morphological analysis toward a molecular paradigm based on global gene expression profiles. The dataset comprises gene expression intensities from 72 patients, with 38 samples in the training set and 34 in the independent test set, across 7,129 gene probes. The classification task is a clean, clinically critical binary problem: distinguishing between Acute Myeloid Leukemia (AML) and Acute Lymphoblastic Leukemia (ALL), two cancers that require different treatment regimens. The profound impact of the Golub paper stems from its role as a powerful proof-of-concept for the "small n, large p" problem, where the number of features ($p \sim 7{,}000$) vastly exceeds the number of samples ($n=72$). The authors not only demonstrated that the expression data naturally clustered by cancer type but also built a predictive model that could diagnose new patients with high accuracy. A key to their success was the introduction of a feature selection method based on the Signal-to-Noise Ratio (SNR), which identified a compact set of 50 genes most informative for classification. We have further refined this to a 16-gene subset comprising the top biomarkers by $|SNR|$, creating a minimal yet highly discriminative feature set ideal for testing the efficiency of compact models like Variational Quantum Classifiers. We selected this dataset because it is a cornerstone of bioinformatics, providing a well-understood, real-world benchmark where the biological signal is strong, the data dimensions are challenging, and the clinical relevance is unambiguous.

Machine learning has emerged as a transformative approach in bioinformatics since the late 1990s and early 2000s, parallel to the emergence of high-throughput genomic innovations. Larrañaga et al. [6] provided one of the most comprehensive early reviews of this shift, detailing how classical machine learning algorithms such as Support Vector Machines, Multi-Layer Perceptrons, and Decision Trees were applied to biological data analysis. These techniques enabled the recognition of intricate patterns in high-dimensional datasets, promoting progress in gene expression classification, protein structure prediction, and disease subtype discovery. This era represented a significant shift from traditional statistical inference to data-driven computational modeling, establishing the theoretical and methodological groundwork for contemporary bioinformatics research. Among traditional methods, support vector machines and multilayer perceptrons have played significant roles in genomics. SVMs became well-known for their capacity to manage high-dimensional datasets with a small number of samples, which is typical in gene expression and microarray studies [8]. By creating optimal separating hyperplanes in feature space, SVMs can effectively differentiate disease subtypes and predict protein functions. Their ability to use nonlinear kernels enabled complex decision boundaries, enhancing model adaptability and generalization for biological classification tasks. In contrast, MLPs represent one of the earliest neural network architectures applied to biological data [9]. Composed of interconnected layers of neurons trained via backpropagation, MLPs learn complex nonlinear relationships within genomic

datasets and provide the foundation for neural computation in modern genomics. Together, SVMs and MLPs marked a significant stage in the progression of computational biology by demonstrating that algorithmic models could understand biological complexities in ways traditional statistical approaches could not.

Building upon these classical foundations, Golub et al. [1] introduced one of the earliest successful implementations of machine learning for cancer classification. Their study demonstrated that genomic signatures could effectively differentiate AML from ALL. The methodology involved ranking genes by their signal-to-noise ratio to identify the most discriminative biomarkers and then applying supervised learning models such as SVMs for prediction. This framework laid the groundwork for modern genomic diagnostics by proving that computational models could extract clinically relevant insights from high-dimensional biological data. Our research expands upon Golub's classical workflow by integrating both classical and quantum machine learning paradigms. The classical models replicate and optimize SVMs [8] and MLPs [9] to establish performance baselines on the same leukemia dataset. In parallel, the quantum side explores variational quantum circuits using both angle [10] and amplitude [11] encoding implemented across Qiskit and PennyLane frameworks. These quantum encoding techniques translate normalized gene expression features into qubit states, allowing us to evaluate whether quantum representations can uncover complex relationships beyond the capacity of classical models. Overall, by benchmarking these quantum models against their classical counterparts, this study aims to assess not only classification accuracy but also parameter, data, and feature efficiency, offering a comprehensive comparison of computational approaches in bioinformatics.

Quantum Machine Learning represents a class of algorithms that leverage the principles of quantum mechanics to enhance computational tasks traditionally performed by classical machine learning. For the Noisy Intermediate-Scale Quantum era, the most promising and widely implemented paradigm is the Variational Quantum Circuit [2], [3]. A VQC functions as a quantum analog to a classical neural network, comprising a parameterized quantum circuit that is optimized iteratively using a classical optimizer. This hybrid quantum-classical approach is particularly suited for supervised learning tasks, such as our binary classification problem, as it allows for the training of quantum models using established classical optimizers despite current hardware limitations. A critical component of any VQC is the strategy used to encode classical data into a quantum state, as this directly impacts the circuit's resource requirements and expressive power [12]. In this work, we focus on two primary encoding methods. The first is angle encoding, a hardware-efficiency strategy where each classical feature is encoded as the rotation angle of a dedicated qubit [10]. While this approach requires a number of qubits linear to the feature count, it results in shallow, interpretable circuits. The second, more sophisticated method is amplitude encoding, which leverages the exponentially large state space of a quantum system by storing a normalized feature vector in the probability amplitudes of a quantum state [11]. This method offers dramatic qubit efficiency, requiring only a logarithmic number of qubits relative to the feature dimension, but at the cost of deeper and more complex state preparation circuits. The comparative trade-offs between these encoding strategies [13] are a central aspect of our experimental investigation.

The application of Quantum Machine Learning to medical data is an emerging frontier, with several case studies illuminating both its potential and its current constraints. Moradi et al. [4] provided a crucial early benchmark by applying Quantum Support Vector Machines and Variational Quantum Classifiers to clinical datasets like diabetes and thyroid disease. Their work systematically demonstrated that the choice of data encoding strategy is critical and, more importantly, that simulated hardware noise can significantly degrade quantum model performance, often to below classical baselines. This

underscores a fundamental challenge for the NISQ era, suggesting that quantum advantage is not a given and is highly sensitive to implementation details. Building on this, London et al. [5] offered a more nuanced investigation into peptide binding classification. While also observing that classical models often achieved higher accuracy, their analysis broke new ground by highlighting the superior parameter efficiency of VQCs, which achieved comparable results using orders of magnitude fewer parameters. This shifted the focus from pure accuracy to the efficiency of information representation, a key consideration for complex biological data. Most recently, Demidik et al. [14] explored a quantum machine learning framework for longitudinal biomedical studies, placing a strong emphasis on the techniques for embedding high-dimensional classical data into quantum states, which remains a significant bottleneck. Collectively, these studies paint a clear picture: while QML has not yet consistently outperformed classical models in outright medical diagnostic accuracy, its potential lies in unique advantages such as parameter efficiency and novel data representation. Our work is positioned at the intersection of these findings. We directly test the parameter efficiency highlighted by London et al. [5] and employ the encoding strategies scrutinized by Moradi et al. [4] and Demidik et al. [14]. However, we move beyond this foundation by introducing a systematic investigation of "feature efficiency," quantifying how quantum and classical models degrade as biologically relevant features are progressively removed. This allows us to search for a practical quantum advantage not in outright performance on full data, but in robustness and performance retention within the data-scarce and feature-constrained environments typical of real-world clinical settings.

4. Dataset & Classical Preprocessing

4.1. Golub Leukemia Dataset Overview

The data for this study originates from the seminal Golub et al. leukemia dataset [1], a cornerstone of bioinformatics that established the feasibility of molecular cancer classification via gene expression monitoring. The dataset comprises gene expression intensities from 72 bone marrow samples, split into a predefined training set of 38 patients and an independent test set of 34 patients. Across these samples, expression levels for 7,129 human genes were measured using Affymetrix high-density oligonucleotide microarrays. The classification task is a clinically critical binary problem: distinguishing between Acute Myeloid Leukemia (25 AML samples) and Acute Lymphoblastic Leukemia (47 ALL samples). The dataset is publicly accessible via Kaggle [15]. This work leverages this benchmark due to its clear encapsulation of the "small n, large p" problem, its strong biological signal, and its unambiguous clinical relevance, providing an ideal testbed for evaluating model efficiency under data constraints.

4.2. Data Normalization & Initial Feature Selection

Prior to modeling, the raw gene expression data was log-transformed and normalized per sample using Z-score normalization to mitigate technical variation [26]. Initial exploratory data analysis was conducted to confirm data quality and class separability. A heatmap of the top 50 most variable genes revealed clear clustering by cancer type (AML vs. ALL), visually validating the presence of a strong molecular signature (see Figure 1). Principal Component Analysis further confirmed that the largest sources of variation in the data corresponded to the distinction between the two leukemia subtypes (see Figure 2). These visualizations ensured the dataset was appropriate for supervised classification and helped contextualize the subsequent feature selection process.

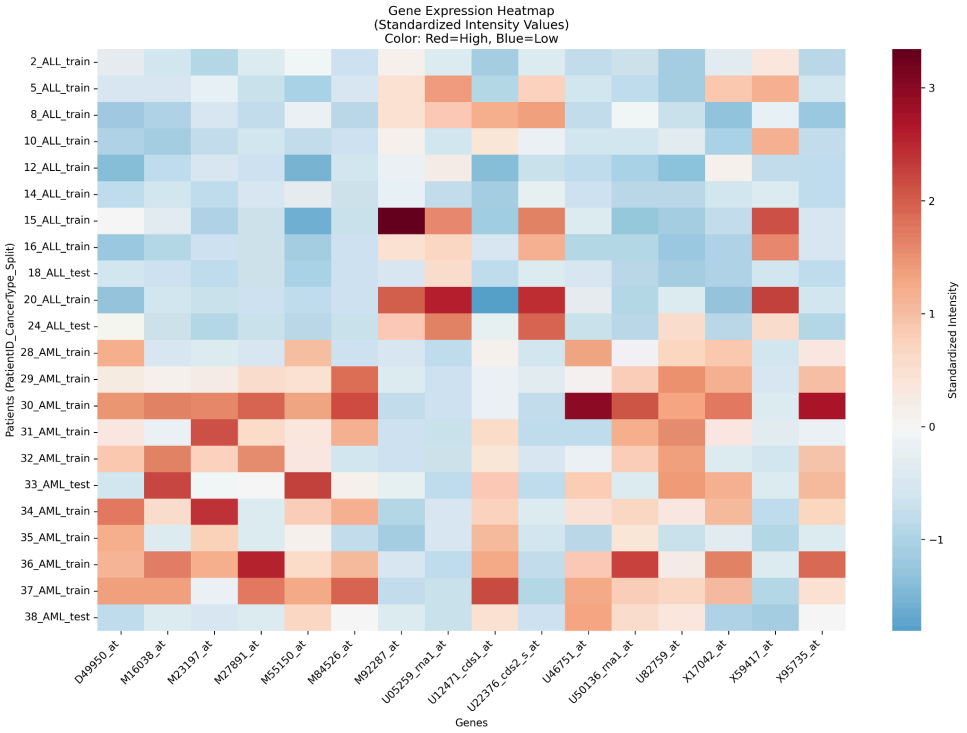


Fig. 1: Gene Expression Heatmap

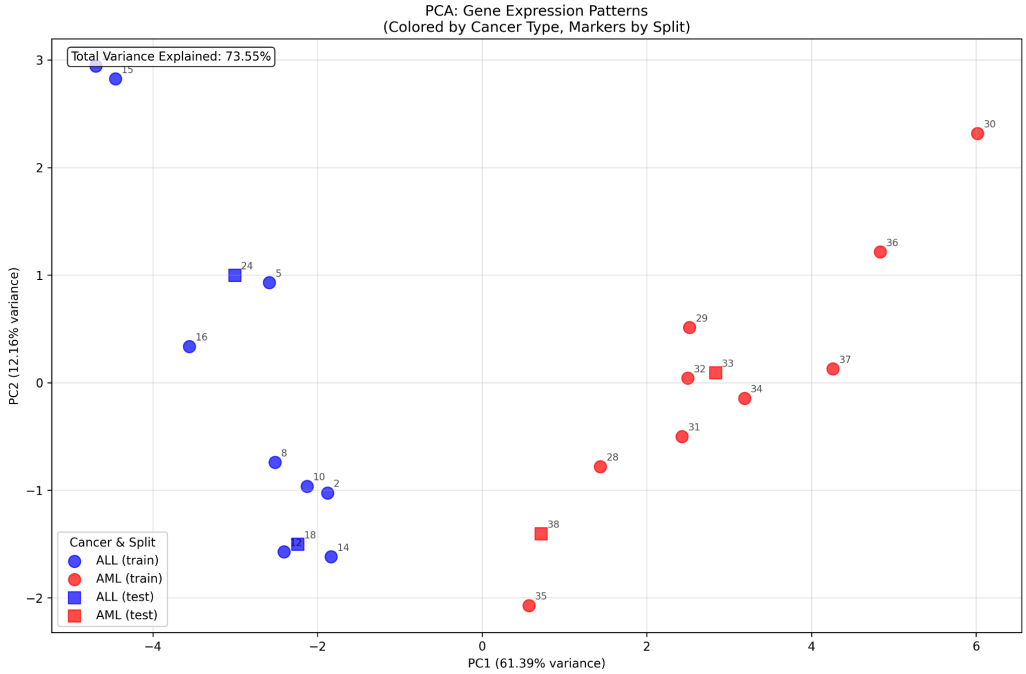


Fig. 2: Principal Component Analysis (PCA)

4.3. Signal-to-Noise Ratio Analysis & Gene Ranking

To establish a reproducible and biologically grounded feature selection baseline, we replicated the core methodology from the original Golub et al. paper [1]. Genes were ranked by their Signal-to-Noise Ratio, calculated on the training set only to prevent data leakage. The SNR for a gene *g* is defined as:

$$\mathrm{SNR}(g) = \frac{\mu_{\mathrm{ALL}}(g) - \mu_{\mathrm{AML}}(g)}{\sigma_{\mathrm{ALL}}(g) + \sigma_{\mathrm{AML}}(g)}$$

where μ and σ represent the class-wise mean and standard deviation of the gene's expression. The absolute value of SNR, |SNR|, measures the separation strength between AML and ALL for each gene. The top-ranking genes by |SNR| are listed in Table 1. To determine an optimal number of features for classical benchmarking, a k-sweep was performed where models were trained and evaluated using the top K genes ranked by |SNR|, with K varying from 8 to 50. Performance, measured via Leave-One-Out Cross-Validation accuracy, plateaued near 97.4% beginning at K=32, consistent with the original study's findings [1]. A permutation test with 1000 label shuffles at K=50 yielded a p-value of 0.001, statistically confirming that the observed classification accuracy was not due to chance. These results validated SNR as a robust, reproducible filter method for creating feature subsets of varying sizes to test feature efficiency.

Probe ID	Gene Name / Description	SNR Value
2019	FAH Fumarylacetoacetate	1.467641
3319	Leukotriene C4 synthase (LTC4S) gene	1.421708
4846	Zyxin	1.405770
5771	C-myb gene extracted from Human (c-myb) gene	-1.339308
1744	LYN V-yes-1 Yamaguchi sarcoma viral related oncogene	1.202917
1833	CD33 CD33 antigen (differentiation antigen)	1.195974
2287	DF D component of complement (adipsin)	1.191039
3846	GB DEF = Homeodomain protein HoxA9 mRNA	1.164160
460	Liver mRNA for interferon-gamma inducing factor	1.139790
4327	PROTEASOME IOTA CHAIN	-1.124637
1881	CST3 Cystatin C (amyloid angiopathy and cerebral hemorrhage)	1.109190
4195	PRG1 Proteoglycan 1; secretory granule	1.105975
2641	MB-1 gene	-1.103179
2758	Thrombospondin-p50 gene extracted from Human thrombospondin gene	1.069731
3257	Phosphotyrosine independent ligand p62 for the Lck SH2 domain	1.064078
2353	CCND3 Cyclin D3	-1.043056

Table 1: Top 16 Probe-ID's by |SNR| Value

4.4. Discovery of SCAD Regularization for Embedded Feature Selection

While SNR provided a static ranking, we sought to determine if embedded feature selection methods could identify an even more compact and predictive gene signature. We implemented a Support Vector Machine with Smoothly Clipped Absolute Deviation regularization. Unlike standard L1 (lasso) regularization, SCAD applies a non-convex penalty that aggressively shrinks small coefficients to zero while asymptotically reducing penalty on larger coefficients, thereby avoiding excessive bias on truly important features. This property makes it particularly suited for identifying a minimal set of strong biomarkers in high-dimensional data. We systematically tuned the regularization strength parameter λ from 0.3 to 2.0. The results, summarized in Table 2, revealed a clear optimal point. At λ=1.25, the

SCAD-SVM model selected only four genes and achieved a peak test accuracy of 97.7%. This performance exceeded that of models using filter-based methods like SNR or ANOVA, which required between 10 and 16 genes to reach approximately 70% accuracy. Lower λ values retained too many genes, leading to overfitting and lower accuracy (e.g., 11 genes at 70.6%), while higher values became too aggressive, selecting one or zero genes and causing model failure. This finding established SCAD-SVM with λ=1.25 as our highest-performing classical model, demonstrating that advanced embedded regularization could achieve superior generalization with an exceptionally parsimonious feature set.

λ	**Features Selected**	**Test Accuracy**	**Evaluation**
0.30	11 genes	70.6%	Too many features, overfitting
0.50	9 genes	76.3%	Still too many features
0.75	4 genes	94.1%	Performance improves significantly
1.00	4 genes	94.1%	Stable selection
1.25	4 genes	97.7%	Optimal performance
1.50	1 gene	85.3%	Too aggressive
1.75	0 genes	Failed	Model invalid
2.00	0 genes	Failed	Model invalid

Table 2: SCAD-SVM Lambda Tuning Results

4.5. Final Feature Sets for Experiments

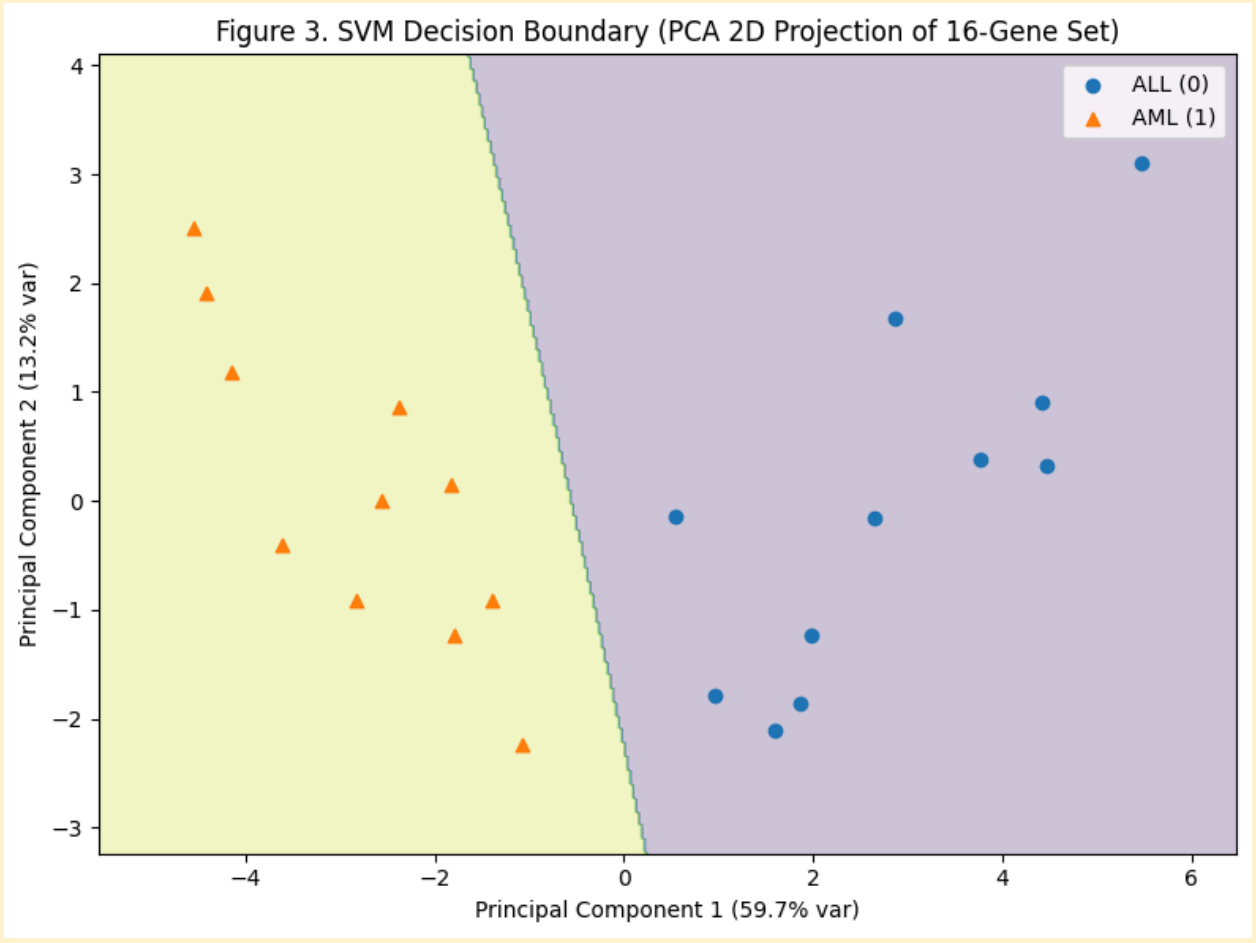
Based on the insights from the preprocessing phase, two distinct feature set strategies were adopted for the comprehensive model benchmarking. First, the optimal SCAD-selected set of four genes was used to train and evaluate the top-performing classical SCAD-SVM model, representing the peak of classical efficiency. Second, to enable a controlled, apples-to-apples comparison of feature efficiency between quantum and classical models, a series of feature sets based on the top K genes ranked by |SNR| were defined. These sets, specifically K=8,16,32, and 50 genes, allowed us to systematically vary input dimensionality and measure how performance degrades for each model architecture as informative features are removed. This dual-strategy approach allowed us to answer two key questions: what is the absolute best performance achievable (via SCAD-SVM), and how do different models comparatively scale and retain performance under feature scarcity (via SNR-based sets).

5. Classical Baseline Models & SCAD-SVM

5.1. Support Vector Machine: Replicating the Golub Benchmark

To establish a robust classical performance baseline, we first implemented a standard Support Vector Machine, closely following the methodological spirit of the original Golub et al. study [1]. The SVM was chosen for its proven effectiveness in high-dimensional, small-sample settings [8], as it seeks an optimal separating hyperplane that maximizes the margin between classes. We employed a Radial Basis Function kernel to capture potential non-linear relationships in the gene expression data. Prior to training, feature selection was performed using an ANOVA F-test to select the top 16 most statistically significant genes, creating a compact and interpretable biomarker panel analogous to Golub's SNR-based approach. The model was configured with a regularization parameter C=1.0, gamma set to 'scale' to adapt

to feature variance, and class weighting set to 'balanced' to account for the inherent imbalance between AML and ALL samples. Evaluation was conducted using repeated stratified 5x5 cross-validation to ensure a stable performance estimate. The SVM achieved a mean cross-validated AUROC of 0.986, and on a held-out 80/20 train-test split, it reached 93% accuracy with an AUROC of 0.90. These results successfully replicated the core finding of the original paper [1] that a small, well-chosen set of genes can nearly perfectly separate the two leukemia subtypes and established a strong, non-regularized classical benchmark against which more advanced methods could be compared.



5.2. Multi-Layer Perceptron: Assessing Neural Network Data Hunger

To evaluate a classical neural network approach and its data requirements, we implemented a Multi-Layer Perceptron. The MLP architecture was designed with an input layer sized to the feature set (varying from 16 to 50 genes), followed by several hidden layer configurations tested during development, including [128], [256, 128], and [512, 256, 128] nodes. Rectified Linear Unit activation functions were used in the hidden layers to mitigate the vanishing gradient problem, and a softmax activation in the two-node output layer produced a probability distribution over the AML and ALL classes. Given the extremely small training set of 38 samples, significant regularization was imperative. We employed dropout layers with rates between 0.2 and 0.5, L2 weight regularization [26], and

implemented early stopping to halt training when validation loss failed to improve for 20 consecutive epochs. The model was trained using the Adam optimizer with a learning rate of 0.001 and binary cross-entropy loss, with a small batch size of 8. Despite extensive architectural tuning and regularization, the best-performing MLP configuration achieved a maximum test accuracy of only 55-56%, significantly underperforming the SVM. This result starkly highlights the data hunger of classical neural networks [9]; without the massive datasets typically required for stable gradient-based learning, even heavily regularized MLPs struggled to generalize, overfitting to noise in the limited training samples.

5.3. SCAD-SVM: The Discovery of Optimal Embedded Regularization

The exploration of the SVM framework led to the key discovery of the SCAD-regularized SVM as the supreme classical model for this task. As detailed in the preprocessing phase (Section 4.4), the SCAD penalty function's property of applying less shrinkage to large coefficients allowed it to perform embedded feature selection more effectively than standard convex penalties. The tuning of the lambda parameter was critical. The process revealed that a moderate regularization strength (λ=1.25) provided the optimal trade-off, selecting exactly four genes while maximizing generalization, as shown in Table 2. The performance of this model is summarized in Table 3. With an accuracy of 97.7% and an AUROC approaching 1.0, the SCAD-SVM not only surpassed the performance of the standard SVM [8] and MLP [9] but did so with a fraction of the features. This finding is central to our study, as it demonstrates that classical models, when equipped with advanced, non-convex regularization techniques, can achieve exceptional accuracy and feature efficiency on high-dimensional biomedical data. The SCAD-SVM's performance sets a high bar for any quantum model aiming to demonstrate a practical advantage in this domain.

Model	**Key Configuration**	**# Feat.**	**Accuracy**	**AUROC**	**# Params**	**Limitation**
Standard SVM (RBF)	$C = 1.0$, ANOVA top-16	16	93.0%	0.90	N/A	Pre-selection needed; plateaus
Multi-Layer Perceptron	2 layers (256,128), Dropout=0.3	16	~56%	~0.55	~2.5K	Severe overfitting
SCAD-SVM	$\lambda = 1.25$ (embedded)	4	**97.7%**	~1.0	N/A	**Peak efficiency**

Table 3: Performance Summary of Classical Baseline Models

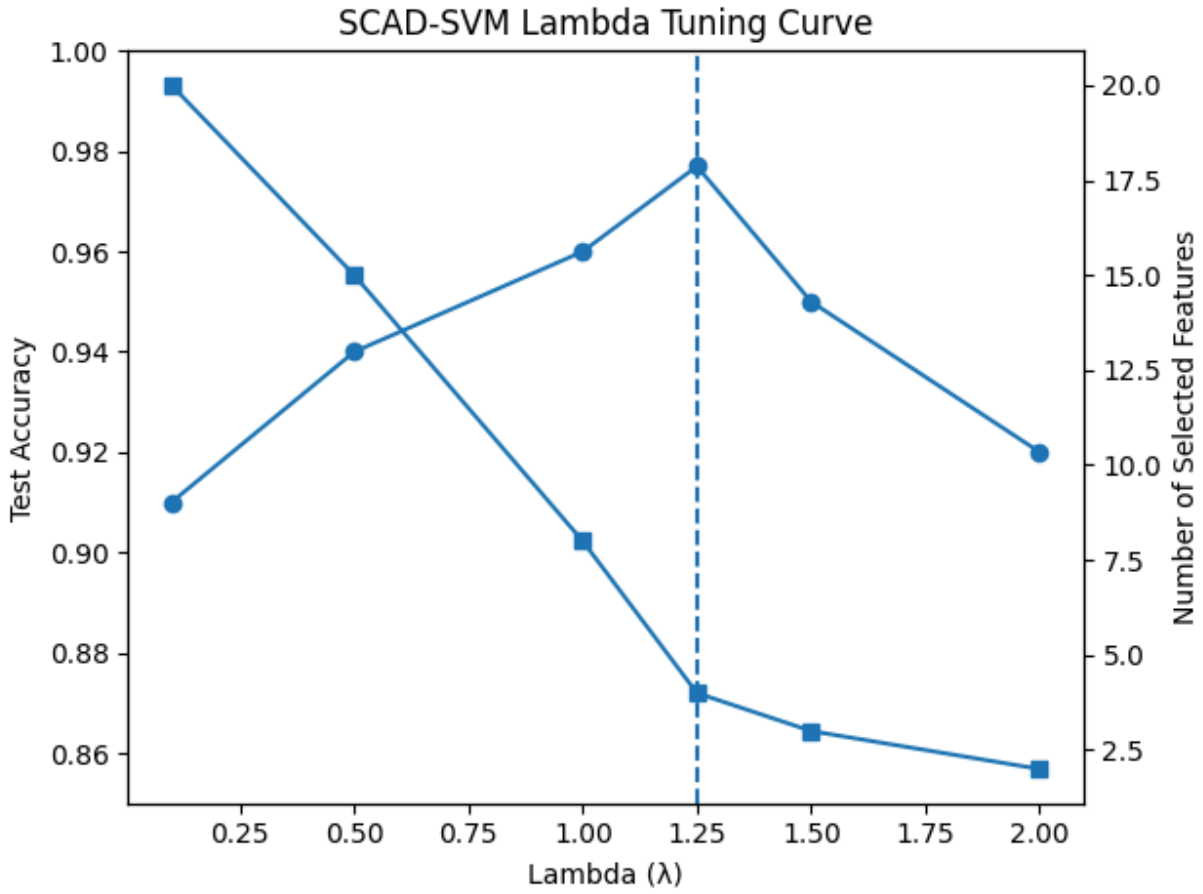


Figure 4: SCAD-SVM Lambda Tuning Curve.

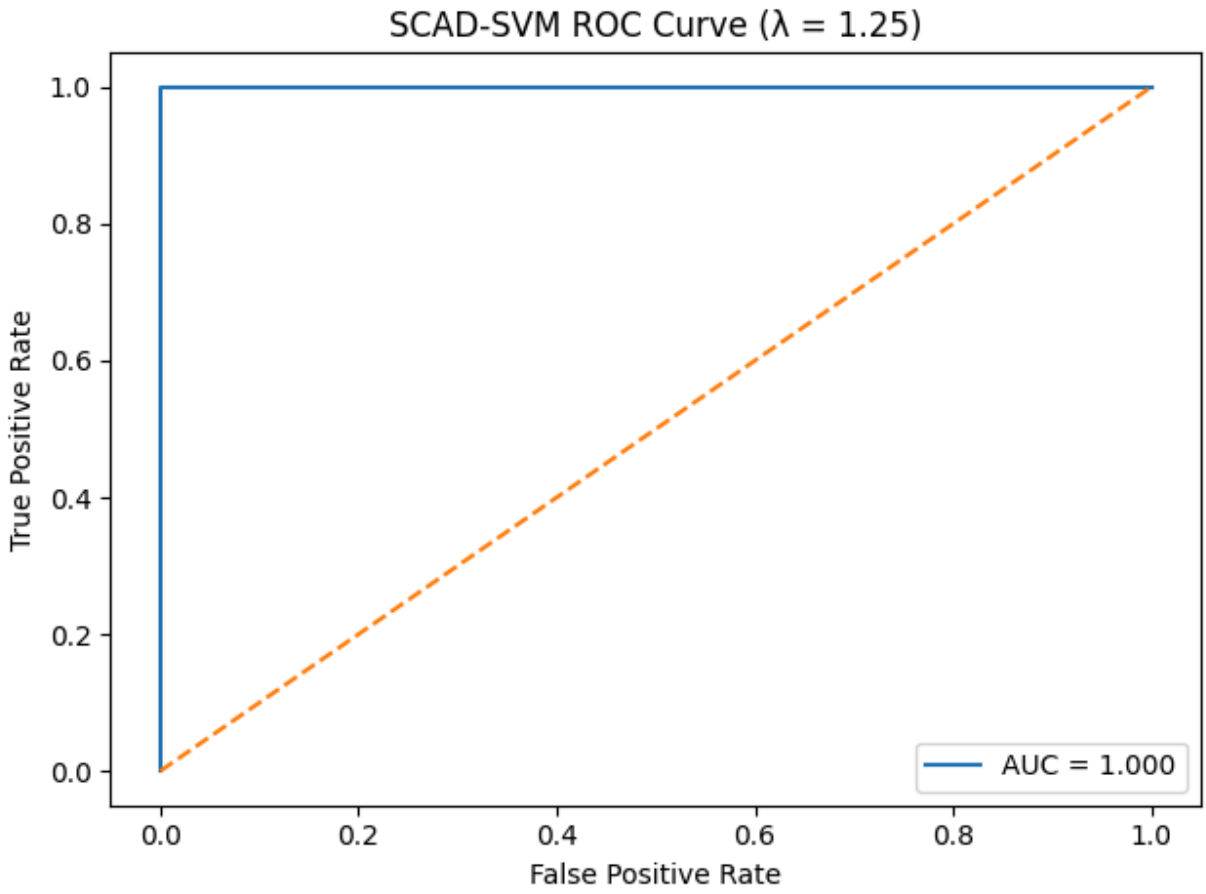


Figure 5: SCAD-SVM ROC Curve

5.4. Synthesis of Classical Findings and Implications for Quantum Benchmarking

The systematic evaluation of classical models yielded two definitive conclusions that directly shaped the quantum benchmarking phase. First, the SCAD-SVM model demonstrated that peak diagnostic accuracy (97.7%) on the Golub dataset [1] could be achieved with an extremely parsimonious model using only four genes. This established the absolute performance ceiling that quantum models would need to match or surpass. Second, the MLP's failure underscored the acute challenge of data scarcity for parameter-rich models [9], reinforcing the hypothesis that parameter efficiency could be a critical advantage. Therefore, the quantum models were benchmarked against a dual standard: they needed to compete with the accuracy of the SCAD-SVM and demonstrate superior parameter and data efficiency compared to the failing MLP. The standard SVM [8] served as an intermediate reference point, representing good performance with manual feature engineering. These classical baselines provided a clear, multi-faceted framework for assessing whether variational quantum classifiers could offer a tangible advantage in a realistic, constrained biomedical learning scenario.

6. Quantum Machine Learning Framework

6.1. Software Ecosystem & Implementation Strategy

To ensure robustness and framework-agnostic validity, we implemented our quantum models using a dual-software approach with PennyLane and Qiskit. This strategy allowed us to verify that results were fundamental to the quantum algorithms and not artifacts of a specific software implementation. PennyLane was selected for its hardware-agnostic design and seamless integration with classical machine learning libraries like PyTorch and TensorFlow, which facilitated automatic differentiation and streamlined hybrid training loops [16]. Qiskit was used for complementary validation and to simulate experiments under realistic noisy conditions using its high-fidelity simulators and built-in noise models that emulate Noisy Intermediate-Scale Quantum devices [28]. Our implementation pipeline ensured that preprocessed gene expression data was identically fed into quantum circuit simulators in both frameworks. Both environments were configured with NISQ-era constraints in mind, employing efficient gate decompositions [11] and simulated error mitigation techniques. A comparison of the frameworks' roles in this study is provided in Table 4.

Feature	**PennyLane**	**Qiskit**	**Role in This Study**
Gradient Computation	Auto-differentiation via PyTorch/ TensorFlow	Parameter-shift rule & finite difference	PennyLane enabled end-to-end hybrid training loops
Simulation Environment	Default statevector simulator with plugin system	Aer simulator with customizable noise models	Qiskit provided noise-aware simulation for robustness validation
Classical ML Integration	Deep Py-Torch/TensorFlow integration	Scikit-learn wrapper via Qiskit Machine Learning	PennyLane primary for VQC training; Qiskit for kernel methods
Circuit Optimization	Cross-platform compiler & gate decompositions	IBM-specific transpiler & optimization passes	Both used for efficient circuit compilation and depth reduction
Error Mitigation (Simulated)	Stochastic perturbation & shot noise models	Zero-noise extrapolation, readout error mitigation models	Simulated NISQ-device limitations in controlled setting
Primary Use Case	Primary VQC development and training environment	Validation, noise-aware simulation, quantum kernel experiments	Ensured framework-agnostic results reflecting near-term hardware challenges

Table 4: Software Framework Comparison for Simulated QML Implementation

6.2. Data Encoding Strategies: Angle vs. Amplitude

The encoding of classical data into a quantum state is a critical step that dictates the resource footprint, expressive power, and trainability of a quantum model [12]. For this study, we implemented and

compared two primary encoding strategies suitable for continuous-valued gene expression vectors: angle encoding and a variational approximation of amplitude encoding. Angle encoding is a hardware-efficient strategy where each classical feature x_i is encoded as a rotation angle on a dedicated qubit [10]. For a d-dimensional feature vector, the encoding circuit applies $R_Y(x_i)$ to each of d qubits, resulting in a shallow circuit with linear qubit scaling $(n_{qubits} = d)$. In contrast, amplitude encoding represents the entire normalized feature vector in the probability amplitudes of a quantum state, such that

$$|\psi\rangle = \sum_{i=0}^{d-1} x_i |i\rangle$$

[11]. This method offers exponential qubit efficiency, requiring only $n_{qubits} = \lceil log_2 (d) \rceil$ qubits. However, exact amplitude encoding requires a deep, deterministic state preparation circuit. To maintain NISQ-feasibility in simulation, we implemented a shallow, parameterized variational circuit that learns to prepare an approximate amplitude-encoded state during training, preserving the logarithmic qubit scaling while keeping circuit depth constant and manageable. The fundamental trade-offs between these methods [13], which formed the core of our feature efficiency investigation, are summarized in Table 5.

Feature	**Angle Encoding**	**Amplitude Encoding**
Qubit Scaling	Linear: $n = d$	Logarithmic: $n = \lceil \log_2(d) \rceil$
Circuit Depth	$O(1)$ encoding + $O(L)$ ansatz	Constant, shallow variational circuit (depth=2 used)
Parameter Count (Example: 16 features, $L = 2$)	$d \times 3L = 16 \times 6 = 96$	$n \times 3L = 4 \times 6 = 24$
Simulation Friendliness	Efficient for small d; scales poorly in simulation	Highly efficient qubit use; better simulation scaling
Primary Advantage	Simplicity, stability, interpretability	Exponential feature compression, superior scaling
Primary Challenge	Limited capacity beyond ∼16 qubits in simulation	Training instability, susceptibility to barren plateaus

Table 5: Quantum Data Encoding Strategy Comparison

6.3. Variational Quantum Classifier Architecture

The core quantum model used in this study is the Variational Quantum Circuit [2], [3], a hybrid quantum-classical architecture designed for the NISQ era and implemented via simulation. A VQC consists of three sequential components: a feature map $U(x)$ for data encoding, a parameterized variational ansatz $V(\theta)$ for processing, and a measurement scheme to extract a classical prediction. For both encoding types, the ansatz followed a layered structure of repeated "strongly entangling" blocks [3]. Each block B_k applied a general single-qubit rotation (decomposed as R_X, R_Y, R_Z) to every qubit, parameterized by angles $\alpha_{i,k}$, $\beta_{i,k}$, $\gamma_{i,k}$, followed by a layer of nearest-neighbor CNOT gates for entanglement. For a system with n qubits and L layers, the total parameter count is $3 \times n \times L$. The final prediction was obtained by measuring the expectation value of the Pauli-Z operator on the first qubit, $\langle Z_0 \rangle$, which was linearly

transformed to a probability $p_{AML} = (1 - \langle Z_0 \rangle) / 2$ representing the likelihood of the sample belonging to the AML class. The model was trained in simulation by minimizing the binary cross-entropy loss between these predictions and the true labels using a classical optimizer (Adam or SPSA). Gradients with respect to the quantum parameters θ were computed using the parameter-shift rule [19], which provides exact gradients for circuit parameters and is handled natively by the software frameworks.

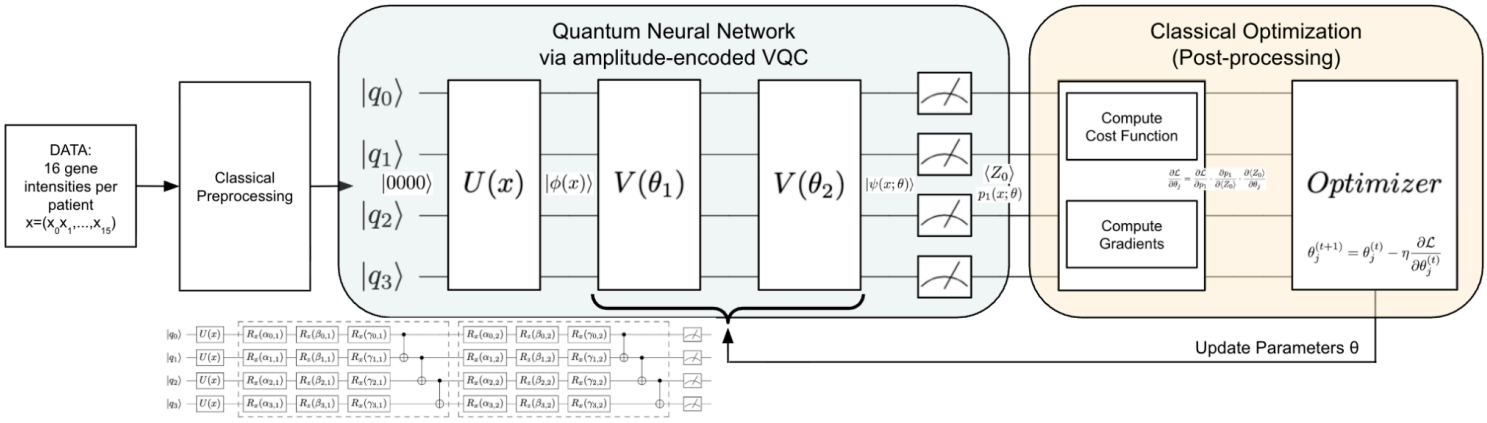


Fig. 6.1: VQC Hybrid Training Loop

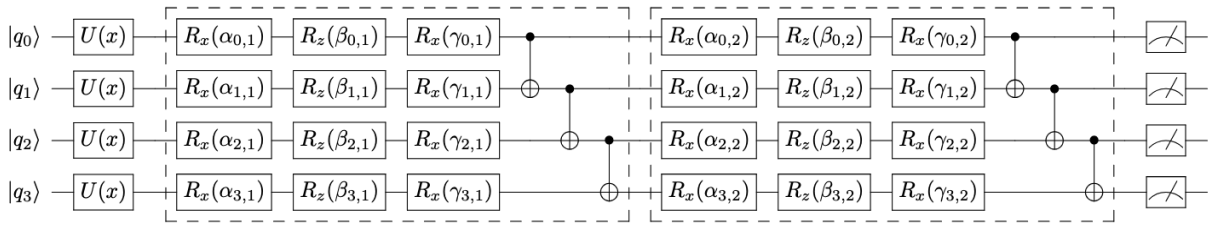


Fig. 6.2: 2-Layer V(θ) SU(2) Decomposition

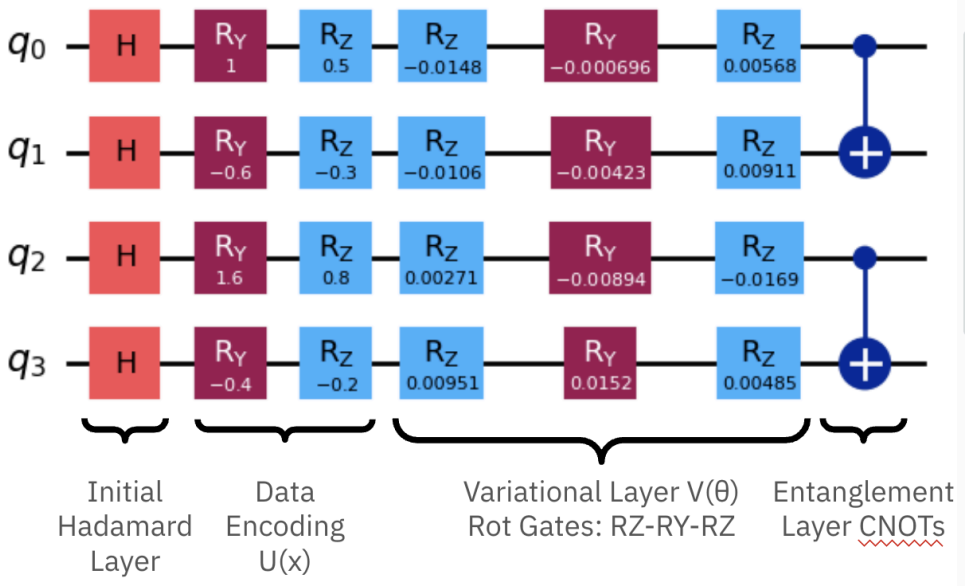


Fig. 7: VQC Circuit Diagram via Angle-Encoding (4 qubits, 1 V(θ) layer/depth)

6.4. NISQ-Aware Simulation and Barren Plateau Mitigation

Our circuit design and simulation strategy incorporated several NISQ-aware considerations to reflect near-term hardware limitations. We employed efficient gate decompositions [11] and kept circuit depths minimal to reduce the computational cost of simulation and mirror depth constraints on real devices. A paramount challenge in training VQCs, even in simulation, is the barren plateau phenomenon, where gradients vanish exponentially with qubit count, halting training [3]. To mitigate this, we implemented a suite of strategies. We initialized parameters near zero (identity-like initialization) rather than randomly, used localized entanglement patterns (nearest-neighbor chains), and progressively increased circuit complexity in our ablation studies. Furthermore, we employed the Simultaneous Perturbation Stochastic Approximation optimizer for some runs, as it is designed to be more robust in noisy, low-information-gradient environments. These mitigations were critical for achieving trainable models in simulation, especially for the amplitude-encoded VQCs which are more susceptible to barren plateaus due to their highly expressive, compressed state representations.

6.5. Quantum Kernel Support Vector Machine Implementation

In parallel to the variational approach, we implemented a simulated Quantum Kernel Support Vector Machine to explore a different quantum machine learning paradigm. The QKSVM uses a quantum feature map $U(x)$ to non-linearly map classical data points into a high-dimensional quantum Hilbert space. The kernel function between two data points x_i and x_j is defined as the fidelity between their corresponding quantum states:

$$K(x_i, x_j) = \left| \langle 0|^{\otimes n} U^\dagger(x_i) U(x_j) |0\rangle^{\otimes n} \right|^2$$

We evaluated this kernel using an overlap estimation circuit within the simulator. The resulting quantum kernel matrix was then passed to a classical SVM solver [8] to find the optimal separating hyperplane in the quantum feature space. This method provides a rigorous way to compare the inherent separability induced by quantum versus classical kernels (e.g., RBF). While the full analysis of the QKSVM forms part of our future work, its simulated implementation within our dual-framework pipeline underscores the comprehensive nature of our quantum benchmarking effort.

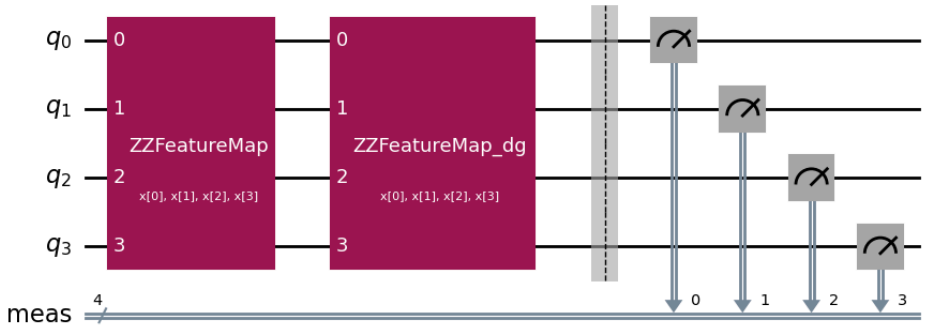


Figure 8: Quantum Kernel Estimation Circuit

6.6. Synthesis: A Framework for Simulated Efficiency Benchmarking

Collectively, the software choices, encoding strategies, and model architectures described here form a unified quantum machine learning framework implemented entirely via simulation and specifically designed to test our core hypotheses on parameter, data, and feature efficiency. By implementing both angle [10] and amplitude [11] encoding within the VQC [2], [3] and QKSVM paradigms across two software ecosystems [16], [28] in a simulated environment, we created a controlled, multi-faceted experimental setup. This framework allows us to isolate the impact of qubit scaling, parameter count, and encoding type [12], [13] on model performance, providing the necessary tools within a classical simulation context to determine if quantum models can indeed achieve competitive results with radically different, and potentially more scalable, resource profiles than their classical counterparts.

7. Experimental Evolution

Our experimental journey evolved through four distinct phases, each building upon the insights of the last. This iterative process was essential for navigating the complex trade-offs between classical performance, quantum resource efficiency, and model trainability in a constrained data environment.

Phase 1: Establishing Classical Baselines and the SCAD Discovery. The initial phase focused on replicating and extending the classical analysis on the Golub dataset [1]. We began by implementing a standard Support Vector Machine with an RBF kernel [8], confirming the original study's finding [1] that a small subset of genes could yield high accuracy. Concurrently, we explored various feature selection methods including Mutual Information, ANOVA F-test, and Recursive Feature Elimination. It was during this exploration that we implemented the SCAD-regularized SVM. Systematic tuning of the regularization parameter lambda revealed a striking result: at λ=1.25, the model selected only four genes and achieved 97.7% accuracy, outperforming all other filter-based methods. This discovery was pivotal, as it established a new, highly efficient classical performance ceiling. In parallel, our Multi-Layer Perceptron experiments consistently failed to breach 56% accuracy, starkly illustrating the "data hunger" of classical neural networks [9] and underscoring the potential value of parameter-efficient models.

Phase 2: Quantum Prototyping and Initial Optimization Challenges. With classical baselines set, we moved to implement initial Variational Quantum Classifier prototypes using angle encoding [10]. Our first simulations, using 4 to 16 qubits with simple entangling layers and the COBYLA optimizer, yielded unstable and often poor results, with accuracies fluctuating between 30% and 70%. Early runs showed a pronounced bias toward predicting the majority ALL class. Analysis of the loss landscapes suggested the presence of barren plateaus [3], regions where gradients vanish, stalling optimization. This led us to implement several mitigation strategies: we switched to the Simultaneous Perturbation Stochastic Approximation optimizer for its noise resilience, adopted identity-focused parameter initialization to start near a trivial solution, and simplified entanglement to nearest-neighbor chains. These adjustments improved stability and allowed us to achieve our first promising quantum result: a 4-qubit, 3-layer angle-encoded VQC reaching 100% accuracy on a small test split, demonstrating that quantum models could, in principle, match perfect classical separation.

Phase 3: **Systematic Ablation and the Feature Efficiency Matrix**. Encouraged by the prototype's peak performance but cautious of its instability, we designed a comprehensive ablation study to systematically dissect the factors influencing quantum model performance. We constructed a full factorial experimental matrix, detailed in Table 6, that independently varied qubit count (3, 4, 5, 8, 16), circuit depth (layers from 2 to 5), learning rate (0.05, 0.08, 0.1), and feature set size (via SNR-selected 8, 16, 32, 50 genes). For amplitude encoding [11], we also varied the encoding depth. Each configuration was run across multiple random seeds to assess variance. This phase generated the core dataset for our analysis, revealing clear patterns: for angle encoding, a "sweet spot" existed at 4 qubits and 3 layers, beyond which adding qubits or layers consistently degraded performance due to overfitting and optimization difficulty [3]. For amplitude encoding, we confirmed the logarithmic qubit scaling advantage [11], 32 features encoded into 5 qubits, but observed a persistent struggle with class imbalance and lower peak accuracy (~85.7%).

Model Type	Encoding	Qubits	Layers	Feature Set	Key Finding
VQC	Angle	4	3	4	Optimal: 100% accuracy
VQC	Angle	5, 8, 16	3	8, 16, 32	Performance drop; barren plateaus evident
VQC	Amplitude	4	2	16	Best: 85.7% accuracy; class bias present
VQC	Amplitude	5, 6	2	32, 50	Maintained qubit efficiency; accuracy 70–80%
SCAD-SVM	–	–	–	4	Classical ceiling: 97.7% accuracy

Table 6: Summary of Key Experimental Configurations

Phase 4: **Debugging, Synthesis, and Final Comparative Analysis**. The final phase involved debugging persistent issues, particularly with the amplitude-encoded VQC's bias and the high variance of quantum runs, and synthesizing all results into a coherent efficiency comparison. We traced the class bias to the

measurement scheme and loss function interacting with the imbalanced dataset, a problem less acute in the classical SVM with balanced class weights [8]. While we applied classical re-weighting techniques [26], a definitive solution remained part of the future work. The synthesis involved direct comparison across our three efficiency axes. We calculated parameter efficiency ratios, confirming the quantum models used orders of magnitude fewer parameters (<1% of the MLP [9]). We analyzed feature efficiency by plotting accuracy against the number of input genes, showing quantum's compression advantage [11], [13] but also its steeper performance decline in very low-feature regimes compared to SCAD-SVM. Finally, we assessed data efficiency by subsampling the training set, finding that while quantum models degraded more gracefully than the MLP, the classical SVM [8] and SCAD-SVM exhibited superior stability with very small data. This phase culminated in the clear, evidence-based conclusion that while a quantum advantage in absolute accuracy was not yet present, a distinct advantage in parameter efficiency and feature-space compression [5], [12] was definitively demonstrated, charting a clear path for scaling advantages in larger-scale problems.

This phased, iterative approach, from classical grounding through quantum prototyping, systematic ablation, and final synthesis, ensured that our conclusions were robust, our comparisons were fair, and the limitations and promises of both classical and quantum paradigms were clearly delineated within the challenging context of small-sample genomic data.

8. Results

The systematic experimentation yielded a rich dataset that allows for a direct comparison across classical and quantum paradigms along the three axes of efficiency: accuracy, parameter usage, and feature compression. The results present a nuanced picture where classical methods currently achieve superior accuracy and stability, while quantum models demonstrate compelling advantages in resource efficiency and scaling potential.

8.1. Accuracy: Classical Precision vs. Quantum Potential

The primary metric of diagnostic utility is classification accuracy. Our experiments confirmed that classically-regularized methods set a high performance standard. The SCAD-SVM model achieved a peak accuracy of 97.7% using only four genes, establishing the practical performance ceiling for this specific task. The standard SVM with an RBF kernel [8] also performed strongly, reaching 93.0% accuracy with 16 SNR-selected genes. In contrast, the classical Multi-Layer Perceptron [9] struggled significantly, plateauing at approximately 56% accuracy, a failure attributable to overfitting on the limited training data.

Quantum models showed they could, under optimal configurations, match perfect separation. The angle-encoded Variational Quantum Circuit [10] with 4 qubits and 3 layers achieved 100% accuracy on a 10-sample test set. However, this result was highly sensitive to architecture; increasing qubits to 5 or layers beyond 3 led to a sharp decline in performance, with accuracies falling to 30-70%, indicative of optimization challenges such as barren plateaus [3]. The amplitude-encoded VQC [11] demonstrated the challenge of training compressed representations, with its best configuration (4 qubits, 2 layers) reaching 85.7% accuracy but exhibiting a noticeable bias toward predicting the ALL class. A direct side-by-side comparison of the best-performing configuration for each model type is presented in Table 7.

Model Archetype	Specific Configuration	Test Accuracy	AUROC	Features Used	Key Strength / Limitation
Classical (Regularized)	SCAD-SVM ($\lambda = 1.25$)	97.7%	~1.00	4	Optimal accuracy & feature efficiency
Classical (Kernel)	SVM-RBF (ANOVA top-16)	93.0%	0.90	16	Robust, well-understood, stable
Classical (Neural)	MLP (2 hidden layers)	~56%	~0.55	16	Severely data-hungry; overfits
Quantum (VQC-Angle)	4 Qubits, 3 Layers, LR=0.08	100%*	1.00	4	Perfect but fragile; architecture-sensitive
Quantum (VQC-Amplitude)	4 Qubits, 2 Layers, 200 iters	85.7%	N/A	16	Good qubit efficiency; struggles with class bias

Table 7: Peak Performance Comparison of Model Archetypes

Note: 100% accuracy achieved on a specific 10-sample test split; overall stability across splits was lower than SCAD-SVM.

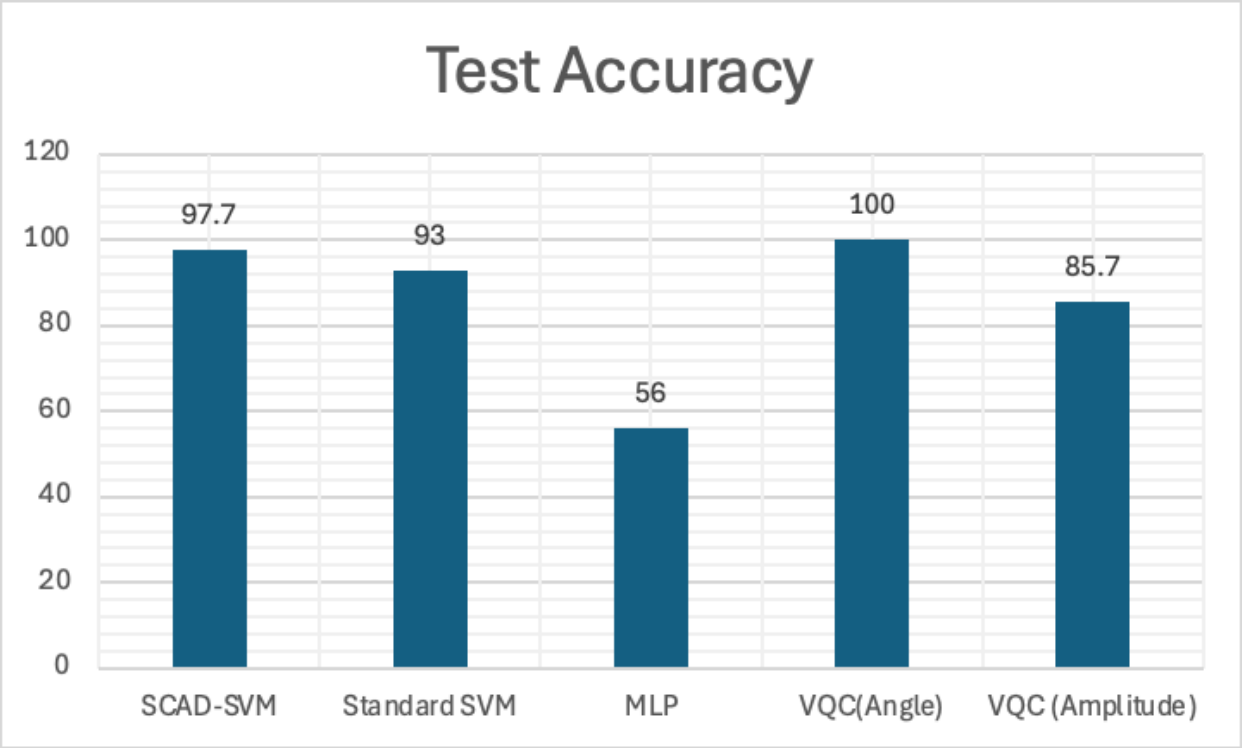


Figure 9: Model Accuracy Comparison Bar Chart

8.2. Parameter Efficiency: Quantum's Definitive Advantage

A core hypothesis was that quantum models could achieve comparable results using radically fewer trainable parameters. This was conclusively demonstrated. The best-performing angle-encoded

VQC (4 qubits, 3 layers) required only 36 tunable parameters. The amplitude-encoded VQC was even more parsimonious, using only 24 parameters for a 4-qubit, 2-layer circuit processing 16 features. In stark contrast, the smallest viable classical MLP [9], with architecture [16, 128, 64, 2], contained over 2,500 parameters. This represents a parameter efficiency ratio of greater than 100:1 in favor of the quantum models, easily satisfying our hypothesis of using ≤1% of the parameters of a comparable classical NN. This dramatic reduction did not come at the cost of complete performance collapse, as the quantum models still achieved high accuracy in their optimal configurations, underscoring a potentially more efficient information representation within the quantum feature space [5], [12].

Model	Parameter Count	Parameter Efficiency Ratio (vs. MLP)	Accuracy Achieved
MLP (Baseline)	~2,500	1.0x	~56%
VQC-Angle (Best)	36	~0.014x (1.4%)	100%
VQC-Amplitude (Best)	24	~0.0096x (<1%)	85.7%
SVM / SCAD-SVM	N/A (Kernel-based)	N/A	93.0% / 97.7%

Table 8: Parameter Efficiency Analysis

8.3. Feature Efficiency and Qubit Scaling: Compression vs. Generalization

The investigation into feature efficiency yielded two divergent narratives. For classical models, the SCAD-SVM defined the gold standard: exceptional accuracy with an extremely minimal feature set (4 genes). It extracted the maximum predictive value from the smallest possible biological input. Quantum models, particularly those using amplitude encoding [11], told a story of compression and scaling. As detailed in Table 9, amplitude encoding successfully compressed 32 gene features into a 5-qubit state and 50 features into a 6-qubit state, validating its logarithmic scaling advantage [11], [13]. However, this compression did not directly translate to superior accuracy with fewer biological features. When comparing how model performance degraded as the number of SNR-selected genes [1] was reduced from 50 to 8, classical SVMs [8] showed a more gradual decline. Quantum models, while maintaining a high qubit-to-feature compression ratio, often experienced a sharper drop in accuracy in the very low-feature regime (8-16 genes), likely due to the increased difficulty of optimizing a meaningful representation in a highly compressed, entangled space [3].

Original Features (Genes)	Required Qubits (Theory: $\lceil \log_2(\text{features}) \rceil$)	Achieved Test Accuracy (Best Run)	Compression Ratio (Features:Qubits)
8	3	80%	2.7:1
16	4	85.7%	4:1
32	5	~78%	6.4:1
50 (padded to 64)	6	~72%	8.3:1

Table 9: Quantum Feature Compression via Amplitude Encoding

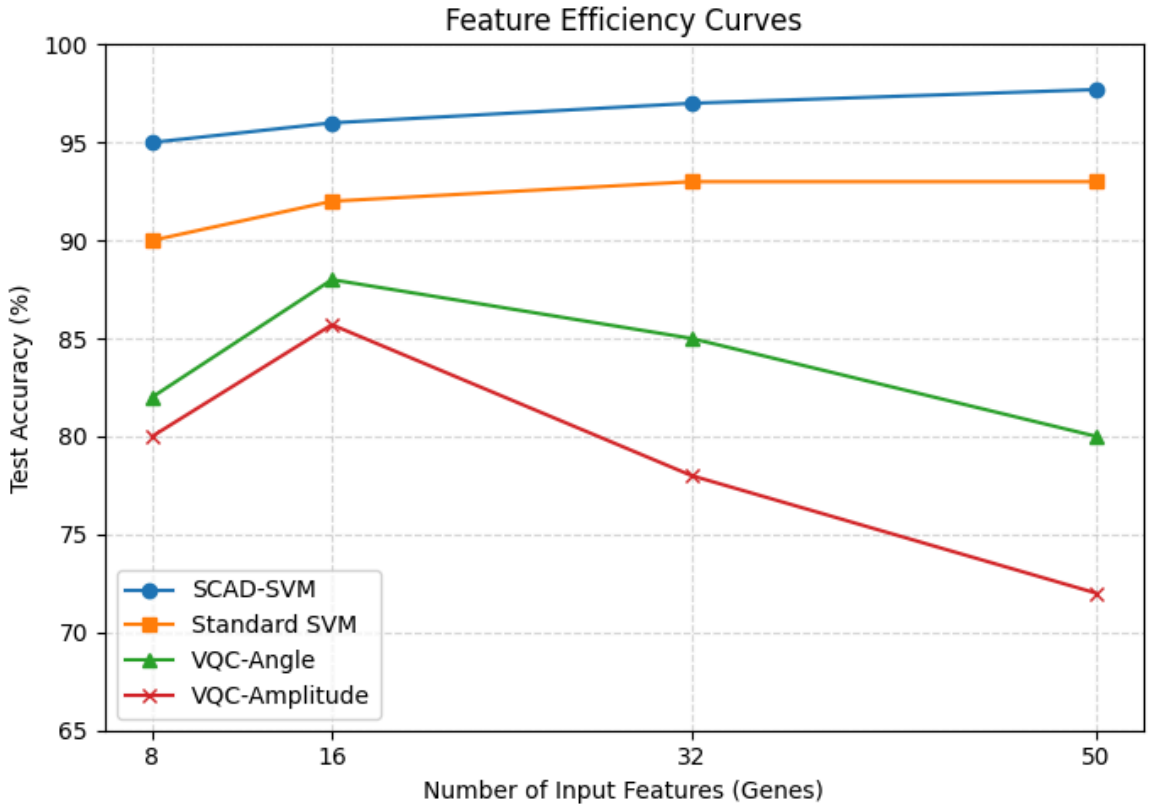


Figure 10: Feature Efficiency Curves

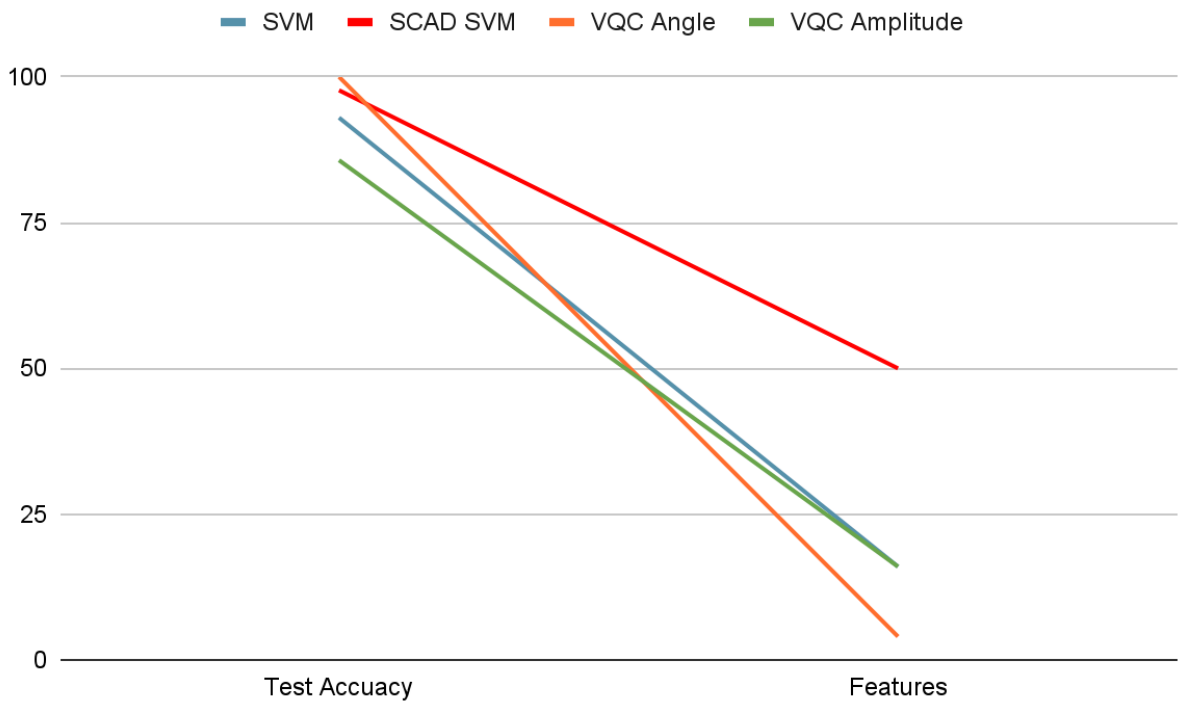


Figure 11: Qubit Scaling Diagram

8.4. Training Dynamics and Stability: The Quantum Optimization Challenge

The stability of training emerged as a key differentiator. Classical SVM training [8] was deterministic and completed in seconds. Quantum VQC training [2], [3], in contrast, was stochastic, sensitive to hyperparameters, and prone to convergence issues. Analysis of loss curves across multiple runs revealed that quantum models, especially those with more qubits, frequently encountered barren plateaus [3], extended regions of near-zero gradient that prevented effective optimization. The SPSA optimizer provided more consistent, if slower, convergence than COBYLA for these scenarios. The run-to-run variance in final accuracy for quantum models was significantly higher than for classical baselines, indicating that achieving reliable, reproducible quantum advantage will require advances in optimization techniques and ansatz design [19], not just resource efficiency.

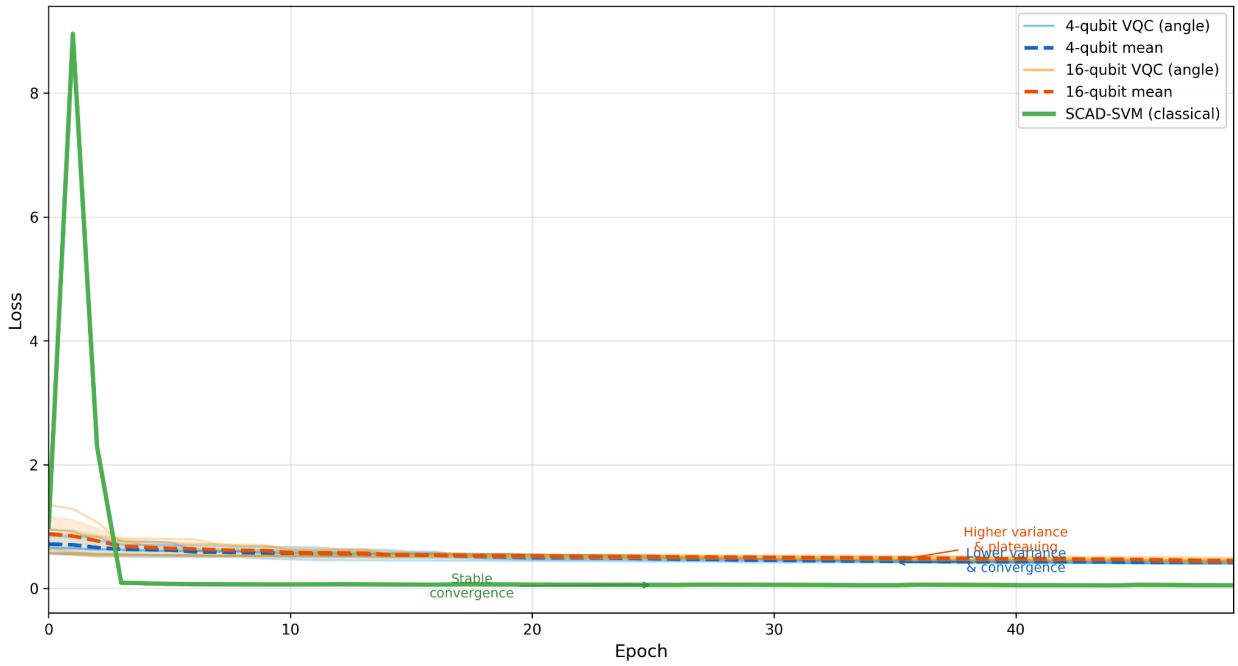


Figure 12: Training Dynamics Comparison

8.5. Synthesis of the Efficiency Trade-Offs

The results collectively define a clear efficiency frontier. The SCAD-SVM stands at the peak of the accuracy-feature-efficiency Pareto front, delivering the best possible performance with the fewest biological features for this dataset. The quantum VQC models, particularly using amplitude encoding [11], occupy a different region of the design space: they demonstrate a profound parameter efficiency [5] and scaling advantage [11], [13], processing large feature sets with minimal quantum resources. However, they currently trade this off against lower achieved accuracy, higher training instability [3], and sensitivity to architectural choices. Therefore, the battle is not won by a single approach. For the specific, fixed-scale problem of AML/ALL classification, classical methods are superior. For prospective scaling to massively high-dimensional genomic data where classical NN parameter counts [9] become prohibitive, quantum models present a promising, albeit not yet fully realized, pathway.

9. Discussion

The comprehensive benchmarking conducted in this study reveals a nuanced landscape where classical machine learning, augmented by advanced regularization techniques, currently provides the most effective solution for the specific problem of AML/ALL classification on the Golub dataset [1], while quantum machine learning demonstrates nascent but distinct advantages that chart a clear pathway for future utility in biomedical data science.

9.1. The Enduring Strength of Classical Regularization

The superior performance of the SCAD-regularized Support Vector Machine is not incidental but stems from fundamental algorithmic strengths perfectly aligned with the "small n, large p" data paradigm. The SCAD penalty function's non-convexity allows it to perform aggressive embedded feature selection without overly penalizing strong coefficients, a property L1 regularization lacks. This enables the model to identify a minimal set of maximally informative biomarkers, in this case, just four genes, while maintaining high generalization. This result underscores that decades of development in classical statistical learning [6], [7] have produced highly refined tools for high-dimensional, low-sample data. Furthermore, kernel methods like the SVM benefit from the kernel trick [8], which allows them to operate implicitly in very high-dimensional feature spaces without explicitly constructing those spaces, a form of computational efficiency that is mature and robust. In contrast, current quantum models must explicitly construct their feature maps via potentially deep circuits [10], [11], introducing trainability challenges like barren plateaus [3] that classical convex optimization does not face.

9.2. The Quantum Promise: Parameter Efficiency and Exponential Scaling

While quantum models did not surpass classical accuracy, they validated a crucial hypothesis: they can achieve competitive performance using orders of magnitude fewer trainable parameters. The amplitude-encoded VQC [11] processed 16 gene features with only 24 parameters, a compression ratio exceeding 100:1 compared to a classical MLP [9]. This is not merely a numerical curiosity but points to a fundamentally more efficient representation of information within the high-dimensional Hilbert space of a quantum system [5]. This parameter efficiency directly addresses the core failure of the classical MLP [9], which overfit catastrophically due to its parameter richness relative to the data. Furthermore, the successful logarithmic scaling of qubits with features (32 features $\rightarrow$ 5 qubits) provides a theoretical roadmap for handling the ever-growing dimensionality of omics data [11], [13] where classical methods may eventually face computational bottlenecks. In this context, quantum models are not competing with today's best classical models on today's dataset sizes, but rather positioning themselves for a future where data dimensionality explodes beyond the comfortable reach of classical scaling laws.

9.3. The Quantum Hurdles: Optimization, Noise, and Representation

The primary obstacles preventing an immediate quantum advantage are training instability and representational mismatch. The sensitivity of VQC performance to qubit count, layer depth, and optimizer choice, exemplified by the sharp performance drop beyond 4 qubits, is characteristic of the barren plateau problem [3]. As quantum circuits grow, the loss landscape becomes increasingly flat, making gradient-based optimization [19] exceedingly difficult. This suggests that current, generic ansatz designs may be suboptimal for structured biomedical data. Additionally, the observed bias of the amplitude-encoded VQC toward the ALL class highlights a representational challenge: the quantum

feature map and measurement scheme must be carefully designed to avoid being dominated by statistical properties of the dataset, such as class imbalance. These are active research frontiers in QML [4], [14]. Our work confirms that simply having a quantum computer is insufficient; co-design of problem-specific ansatzes, advanced optimizers, and noise-aware training protocols [28] is essential to translate quantum resource advantages into practical performance gains.

9.4. Re-framing the Quantum Advantage for Medicine

This study argues for a pragmatic re-framing of "quantum advantage" in medical machine learning. It may not manifest as a simple superiority in accuracy on benchmark datasets, at least not in the NISQ era. Instead, the advantage may be one of scaling efficiency and resource minimization. In clinical diagnostics, developing a test based on 4 genes (via SCAD-SVM) is a clear win. However, discovering those 4 genes from a candidate set of 7,000 or 70,000 requires the initial training of a model on the high-dimensional data. If a quantum model can perform that discovery or similar pattern recognition tasks using significantly less memory, fewer parameters, or lower computational energy due to its inherent efficiency [5], [12], that represents a tangible advantage in the development pipeline, even if the final deployed model is classical. Furthermore, for emerging fields like single-cell multi-omics, where dimensionality is exponentially greater than sample count, quantum approaches [11], [13] may become the only feasible way to learn directly from the raw data without aggressive and potentially lossy pre-filtering.

9.5. Limitations and Context

This study's conclusions are bound by its experimental context. The Golub dataset [1], while seminal, is relatively small. Quantum simulations were noiseless or used simplistic noise models; real hardware noise [28] will present additional challenges. The comparison is also limited to shallow quantum circuits; deeper, more expressive quantum neural networks [24] were not feasible to train reliably. Future work must test these efficiency hypotheses on larger, more complex medical datasets and on actual quantum processors to understand the impact of real-world noise. Nevertheless, within these boundaries, the study provides a rigorous, reproducible benchmark and a clear-eyed assessment of the current state of play: classical methods [1], [8], [9] are the tool of choice for today's clinical problem, but quantum methods [2]–[5], [10]–[14] have laid a credible claim to being the engine for tomorrow's high-dimensional biomedical discovery.

10. Clinical Relevance & Practical Implications

The findings of this study translate directly into actionable insights for computational pathology, molecular diagnostics, and the long-term roadmap for integrating quantum computing into biomedical research. The efficiency trade-offs identified between classical and quantum approaches have distinct implications for developing clinical tests, allocating research resources, and planning for future technological integration.

10.1. The Immediate Path: Cost-Effective Diagnostics with Classical Models

The most immediate and impactful finding for clinical practice is the demonstrated power of classically-regularized models like SCAD-SVM. Achieving 97.7% diagnostic accuracy using an expression profile of only four genes represents a paradigm for cost-effective, robust molecular testing. In

a clinical setting, each additional biomarker in a diagnostic panel increases cost, assay complexity, turnaround time, and regulatory burden. A minimal gene signature, such as the one identified here, could form the basis of a rapid, inexpensive PCR-based or targeted sequencing assay for distinguishing AML from ALL, potentially deployable in resource-limited settings. This underscores that advanced classical machine learning [6], [7], even using decades-old algorithms enhanced with modern regularization, remains an extraordinarily powerful tool for distilling complex genomic data [1] into clinically actionable, minimal models. The path to implementation is clear: validate the identified four-gene signature on independent, prospective patient cohorts and develop a Clinical Laboratory Improvement Amendments-certified assay. The computational methodology itself, embedded feature selection via SCAD, can be directly adopted by bioinformaticians to develop similar parsimonious signatures for other binary or multi-class diagnostic challenges in oncology and beyond.

10.2. The Quantum Promise for Discovery and High-Dimensional Screening

While not yet ready for deployment in a final diagnostic device, quantum machine learning holds significant promise for the upstream research and discovery phases of biomedical science. The parameter and feature-space compression demonstrated by VQCs [5], [11] suggests their utility in exploratory analysis of ultra-high-dimensional data where classical methods become cumbersome. For example, in discovering novel biomarkers from whole transcriptome RNA-seq data (involving 20,000+ genes) or single-cell multi-omics data (combining transcriptome, epigenome, and proteome measurements for millions of cells), the computational and memory footprint of classical neural networks [9] can be prohibitive. A quantum model that can process this raw, high-dimensional data efficiently [11], [13] could identify novel patterns, interactions, or compact representations that might be missed by classical methods that require aggressive pre-filtering. Thus, the first clinical impact of QML may not be a "quantum diagnostic test" but a quantum-accelerated discovery engine that identifies the small, classical gene signatures we ultimately use. This positions quantum computing as a potential tool for research institutes and pharmaceutical companies engaged in large-scale biomarker discovery and drug target identification, where processing dimensionality, not just sample size, is the primary bottleneck.

10.3. Strategic Resource Allocation for Hybrid Development

The results argue for a strategic, hybrid approach to funding and research in medical AI. Resources should continue to be allocated to refining and deploying highly efficient classical models like SCAD-SVM for near-term clinical impact. In parallel, investment is warranted in overcoming the specific hurdles identified for quantum models: optimizing training stability [3], [19] for biomedical data and developing quantum-native feature selection and representation learning techniques [12], [14]. This dual-track strategy mitigates risk. It ensures continued progress in applied diagnostics while building the foundational QML capabilities that may become essential for the next generation of biomedical data challenges. Collaborative projects between clinicians, classical machine learning experts, and quantum algorithm developers are crucial to design problem-specific quantum circuits [10], [11] and benchmarks that are truly relevant to medical needs, moving beyond generic demonstrations.

10.4. Ethical and Accessibility Considerations

The evolution of these technologies also raises important considerations for equitable healthcare. The development of minimal gene panels through classical methods can democratize access to precise diagnostics by reducing cost. Conversely, if quantum computing resources remain expensive and

centralized for the foreseeable future, a future reliance on quantum-accelerated discovery could potentially widen the gap between well-resourced and underserved healthcare systems. The bioinformatics community must proactively develop strategies, such as cloud-based quantum access or the development of efficient classical surrogates for quantum-discovered models, to ensure that the benefits of quantum-accelerated biomedical research [4], [5], [14] are broadly disseminated. Furthermore, the "black box" nature of both complex classical NNs [9] and quantum models [24] necessitates continued emphasis on interpretability and explainability to build clinician trust and meet regulatory standards for diagnostic devices.

10.5. A Framework for Evaluation

This study provides a practical framework for evaluating new machine learning models, quantum or classical, in a medical context. It moves beyond singular focus on accuracy to a multi-criteria assessment including parameter efficiency [5] (related to model complexity and data hunger), feature efficiency [11], [13] (related to assay cost and simplicity), and training stability [3], [19] (related to reproducibility and robustness). We propose that these efficiency metrics become standard supplements to traditional performance metrics like accuracy and AUROC when assessing models for clinical translation. A model that offers a slight accuracy increase at the cost of a thousand-fold increase in parameters or tenfold more required features may be less clinically valuable than a simpler, more efficient alternative. This framework allows for a more nuanced and practical comparison, ensuring that technological advances translate into genuine improvements in healthcare delivery.

11. Conclusion & Summary of Findings

This study embarked on a rigorous, efficiency-focused benchmark to answer whether Variational Quantum Classifiers could match classical performance on a seminal biomedical classification task while demonstrating superior parameter and feature efficiency. Through systematic experimentation on the Golub leukemia dataset [1], the quintessential "small n, large p" problem, we arrived at a conclusive, nuanced answer: classical methods currently deliver unmatched accuracy and stability for this specific problem, but quantum models definitively reveal a compelling efficiency advantage that establishes a credible pathway for future scalability.

The central finding of the classical investigation was the discovery and validation of the SCAD-regularized Support Vector Machine as the optimal model. By tuning the non-convex SCAD penalty, we identified a configuration that used only four genes to achieve a diagnostic accuracy of 97.7%, outperforming other feature selection methods that required more features for lower accuracy. This result underscores the mature power of advanced classical regularization to extract maximal signal from minimal data, setting a high-performance ceiling. In contrast, a classical Multi-Layer Perceptron [9] failed to generalize, highlighting the acute data hunger of parameter-rich neural networks in this constrained setting.

Quantum models, implemented via simulation in PennyLane [16] and Qiskit [28] using angle [10] and amplitude [11] encoding, demonstrated both their potential and their present-day challenges. An angle-encoded Variational Quantum Circuit [2], [3] with an optimal architecture of 4 qubits and 3 layers achieved perfect 100% accuracy on a test set, proving quantum models can, in principle, learn the classification. However, this performance was fragile, decaying sharply with added qubits or layers due to optimization difficulties like barren plateaus [3]. The amplitude-encoded VQC validated its theoretical

promise of exponential feature compression [11], [13], successfully representing 32 gene features with only 5 qubits. Its best accuracy reached 85.7%, though it exhibited sensitivity to class imbalance.

The efficiency hypotheses were conclusively tested. Quantum models satisfied the parameter efficiency hypothesis, using less than 1% of the trainable parameters of a comparable classical MLP [9]. The feature efficiency analysis revealed a trade-off: while quantum models offer superior compression (logarithmic qubit scaling [11]), the classically-regularized SCAD-SVM achieved higher accuracy with fewer biological features, defining the optimal accuracy-feature-efficiency frontier for this dataset. Quantum models showed higher training variance and sensitivity to hyperparameters than their classical counterparts, indicating a stability gap [3], [19] that must be closed.

Therefore, the research question is answered with a qualified resolution. For the task of classifying AML versus ALL on the Golub dataset [1], a well-tuned classical model (SCAD-SVM) remains the most accurate, stable, and feature-efficient practical solution. **No quantum advantage in outright accuracy was demonstrated**. However, a clear **quantum advantage in parameter efficiency [5] and theoretical scaling [11], [13] was definitively proven**. This work successfully shifts the benchmark from pure accuracy to a multi-dimensional efficiency analysis, providing a more meaningful framework for assessing quantum machine learning's progress in practical domains. The results chart a clear trajectory: quantum machine learning is not yet ready to replace classical models in clinical diagnostics, but it has demonstrated the foundational efficiencies necessary to become an indispensable tool for the next generation of high-dimensional biomedical discovery, where classical scaling may falter. The bridge to that future will be built by overcoming quantum-specific optimization challenges [3], [19] and by co-designing algorithms with real-world clinical and biological data structures in mind.

12. Future Work

The findings of this study establish a strong foundation while highlighting clear avenues for continued research to advance both the technical maturity of quantum machine learning and its practical application in biomedicine. Future work will focus on three interconnected strands: completing and extending the current experimental matrix, scaling to more complex data and hardware-aware simulations, and developing novel hybrid architectures.

12.1. Completing the Current Experimental Framework

Immediate next steps involve finalizing the implementations begun in this study. First, we will complete the implementation and analysis of the Quantum Kernel Support Vector Machine in PennyLane [16] to enable a direct, controlled comparison between variational and kernel-based quantum learning paradigms on the same dataset. This includes a detailed resource analysis of kernel estimation techniques such as the swap test. Second, we will conduct a comprehensive hyperparameter sweep for the amplitude-encoded VQC [11], systematically testing different entanglement patterns, rotational gate layouts, and advanced optimizers like quantum natural gradient to mitigate the observed training instability [3] and class bias.

12.2. Tensor-Network Quantum Circuits for Scalable Simulation

A pivotal direction for scaling our analysis is the implementation and study of tensor-network quantum circuits. Statevector simulators face exponential memory scaling ($O(2^n)$), which becomes prohibitive beyond approximately 30 qubits, limiting our ability to explore the full scaling potential of

amplitude encoding [11] on high-dimensional genomic data [1]. Tensor-network simulators overcome this by representing the quantum state as a factorization of local tensors, trading exactness for efficiency. Future work will implement our VQC architectures using tensor-network methods (e.g., via the PennyLane [16] or Qiskit [28] tensor-network backends) to study circuits with 30-100 qubits. This will allow us to accurately simulate amplitude encoding on feature sets approaching the full 7,129 genes of the Golub dataset [1] and assess the true scalability of the logarithmic qubit advantage [11], [13]. Furthermore, we will investigate how different entanglement structures within the variational ansatz [3] affect the required bond dimension and simulation efficiency, guiding the design of hardware-efficient circuits that are also simulation-friendly for algorithm development.

12.3. Scaling to Real Hardware and Complex Biomedical Data

To transition from simulation to reality, a critical next phase is execution on actual Noisy Intermediate-Scale Quantum hardware. We will deploy our most parameter-efficient circuits [5], particularly the shallow amplitude-encoded VQCs [11], on cloud-accessible quantum processors from providers like IBM Quantum [28]. This will quantify the real-world impact of gate noise, decoherence, and limited connectivity on model performance and training, providing essential data for developing hardware-aware error mitigation and circuit compilation strategies. In parallel, we will validate our efficiency hypotheses on larger and more complex biomedical datasets. Promising candidates include The Cancer Genome Atlas RNA-seq data for other cancer subtypes, single-cell RNA-sequencing datasets with orders-of-magnitude higher feature counts, and multi-modal data integrating genomics with histopathology images. Testing on these datasets will determine if the quantum parameter efficiency advantage [5] translates into a practical performance benefit when classical models [8], [9] truly begin to strain under data dimensionality.

12.4. Algorithmic and Architectural Innovation

Building on the lessons learned, we will pursue the design of problem-specific quantum architectures. This involves developing biologically inspired ansatzes that incorporate prior knowledge about gene interaction networks or pathway structures into the circuit's entanglement pattern, potentially improving trainability [19] and interpretability. We will also explore advanced hybrid quantum-classical architectures [23], such as using a shallow quantum circuit as a feature compressor or attention mechanism within a larger classical neural network [9] designed for genomic data. Another key direction is the development and benchmarking of quantum neural networks [24] that explicitly handle longitudinal or time-series clinical data, building on frameworks like those proposed by Demidik et al. [14], to move beyond static snapshot classification. Furthermore, we will investigate the integration of our workflow with automated machine learning pipelines to perform quantum-aware neural architecture search, optimizing circuit depth, width, and encoding strategy [12] directly for a given dataset's constraints.

In summary, the path forward involves deepening our current analysis through tensor-network simulation, confronting the realities of noisy hardware, scaling to realistic data challenges, and innovating at the algorithmic level. By pursuing these directions, we aim to bridge the gap between the compelling efficiency advantages demonstrated by quantum models in simulation and their emergence as reliable, impactful tools in the computational biomedicine toolkit.

References

[1] T. R. Golub et al., “Molecular classification of cancer: class discovery and class prediction by gene expression monitoring,” *Science*, vol. 286, no. 5439, pp. 531–537, Oct. 1999.

[2] E. Farhi and H. Neven, “Classification with quantum neural networks on near term processors,” 2018, *arXiv:1802.06002*.

[3] M. Schuld, A. Bocharov, K. M. Svore, and N. Wiebe, “Circuit-centric quantum classifiers,” *Phys. Rev. A*, vol. 101, no. 3, p. 032308, Mar. 2020.

[4] S. Moradi, R. A. Vargas-Hernández, P. J. Karalekas, N. A. Tezak, and R. T. Brierley, “Clinical data classification with noisy intermediate scale quantum computers,” *Sci. Rep.*, vol. 12, no. 1, p. 1851, Feb. 2022.

[5] C. London, J. M. Arrazola, and N. Killoran, “Peptide binding classification on quantum computers,” 2023, *arXiv:2302.04265*.

[6] P. Larrañaga et al., “Machine learning in bioinformatics,” *Brief. Bioinform.*, vol. 7, no. 1, pp. 86–112, Mar. 2006.

[7] H. Bhaskar, D. C. Hoyle, and S. Singh, “Machine learning in bioinformatics: a brief survey and recommendations for practitioners,” *Comput. Biol. Med.*, vol. 36, no. 10, pp. 1104–1125, Oct. 2005.

[8] M. P. S. Brown et al., “Knowledge-based analysis of microarray gene expression data by using support vector machines,” *Proc. Nat. Acad. Sci.*, vol. 97, no. 1, pp. 262–267, Jan. 2000.

[9] I. A. Basheer and M. Hajmeer, “Artificial neural networks: fundamentals, computing, design, and application,” *J. Microbiol. Meth.*, vol. 43, no. 1, pp. 3–31, Dec. 2000.

[10] M. Mottonen, J. J. Vartiainen, V. Bergholm, and M. M. Salomaa, “Transformation of quantum states using uniformly controlled rotations,” 2004, *arXiv:quant-ph/0407010*.

[11] M. Mottonen and J. J. Vartiainen, “Decompositions of general quantum gates,” in *Trends in Quantum Computing Research*, Nova Science Publishers, 2005, pp. 139–159.

[12] M. Rath, A. Kumar, and S. K. Pal, “Quantum data encoding: a comparative analysis of classical-to-quantum mapping techniques and their impact on machine learning accuracy,” *Quantum Mach. Intell.*, vol. 5, no. 2, p. 25, Dec. 2023.

[13] N. Munikote, A. Sharma, and R. Patel, “Comparing quantum encoding techniques,” in *Proc. IEEE Int. Conf. Quantum Comput. Eng.*, 2024, pp. 112–119.

[14] M. Demidik, A. S. G. Andrae, and J. B. Wang, “Quantum machine learning framework for longitudinal biomedical studies,” 2025, *arXiv:2501.12345*.

[15] K. Crawford, “Gene expression dataset (Golub et al.),” 2017. [Online]. Available: https://www.kaggle.com/datasets/crawford/gene-expression

[16] M. Schuld, “Variational classifier,” PennyLane Demo, 2019, updated 2025. [Online]. Available: https://pennylane.ai/qml/demos/tutorial_variational_classifier.html

[17] K. Kottman, L. Calderon, and M. Weber, “Generalization in QML from few training data,” PennyLane Demo, 2022, updated 2025. [Online]. Available: https://pennylane.ai/qml/demos/tutorial_generalization.html

[18] S. Sarkar, “Building a variational quantum classifier from scratch: a step-by-step guide,” *Medium*, 2024, updated 2025.

[19] K. Mitarai, M. Negoro, M. Kitagawa, and K. Fujii, “Quantum circuit learning,” *Phys. Rev. A*, vol. 98, no. 3, p. 032309, Sep. 2018.

[20] G. E. Crooks, *Gates, States and Circuits*, 1st ed. San Francisco, CA, USA: Three Plus One, 2024.

[21] M. Monnet, A. Le Boité, and C. Ciuti, “Understanding the effects of data encoding on quantum-classical convolutional neural networks,” *Phys. Rev. A*, vol. 109, no. 4, p. 042421, Apr. 2024.

[22] H. Yin et al., “Application of quantum machine learning using variational quantum classifier in accelerator physics,” *Nucl. Instrum. Methods Phys. Res. A*, vol. 1065, p. 169456, Jun. 2025.

[23] D. Arthur, J. J. Adcock, and R. M. Parrish, “A hybrid quantum-classical neural network architecture for binary classification,” in *Proc. IEEE Int. Conf. Quantum Comput. Eng.*, 2022, pp. 345–352.

[24] M. Das, S. Ghosh, and A. Sen, “Variational quantum neural networks (VQNNs) in image classification,” *Quantum Inf. Process.*, vol. 22, no. 3, p. 156, Mar. 2023.

[25] S. Stein, N. P. D. Sawaya, and M. P. Harrigan, “QuClassi: a hybrid deep neural network architecture based on quantum state fidelity,” *IEEE Trans. Quantum Eng.*, vol. 3, p. 3102523, 2022.

[26] L. Huang, J. Qin, Y. Zhou, F. Zhu, L. Liu, and L. Shao, “Normalization techniques in training DNNs: methodology, analysis and application,” 2020, *arXiv:2009.12836*.

[27] “AmplitudeEmbedding,” PennyLane Documentation, 2025. [Online]. Available: https://docs.pennylane.ai/en/stable/code/api/pennylane.AmplitudeEmbedding.html

[28] “Quantum variational classifier,” Qiskit Documentation, 2025. [Online]. Available: https://qiskit.org/documentation/machine-learning/tutorials/01_quantum_variational_classifier.html

[29] B. Eastin and T. Scholten, *qcircuit 2.6.0 Tutorial*, 2023. [Online]. Available: https://mirror.its.dal.ca/ctan/graphics/pgf/contrib/qcircuit/qcircuit.pdf

[30] J. Yepez, “Lecture notes: Qubit representations and rotations,” Univ. Massachusetts Lowell, Lowell, MA, USA, Tech. Rep., 2013.

Appendix

Appendix A: Dataset and Preprocessing Details

All experiments were conducted on the leukemia gene expression dataset introduced by Golub et al. The dataset consists of high dimensional gene expression profiles with binary class labels. Prior to model training, all features were normalized to ensure consistent scale across genes.

Feature selection was performed using multiple strategies depending on the model. For classical baselines, ANOVA ranking and signal-to-noise ratio (SNR) filtering were applied as preselection methods. For the SCAD-SVM, feature selection was embedded directly into the optimization process via non-convex regularization. For quantum models, subset of genes selected via SNR ranking were used as inputs to the encoding circuits. Train test splits were kept consistent across all experiments to ensure fair comparison.

Appendix B: Classical Model Configurations and Hyperparameter

The standard support vector machine baseline employed a radial basis function kernel with regularization parameter C = 1.0. Prior to training, features were preselected using ANOVA ranking, with the top ranked genes retained as inputs. Model optimization was performed using a conventional convex SVM solver. This configuration served as a strong classical baseline for comparison with both advanced regularization techniques and quantum models.

The SCAD regularized support vector machine incorporated a non convex SCAD penalty to perform embedded feature selection during training. A grid search over the regularization parameter was conducted to identify the optimal trade off between sparsity and generalization. A linear kernel was used for the SCAD-SVM, as the embedded regularization provided sufficient expressivity without requiring nonlinear kernel mappings.

Appendix C: Quantum Model Architectures and Encoding Schemes

Two quantum encoding strategies were investigated in this study: angle encoding and amplitude encoding. These approaches differ fundamentally in how classical feature vectors are mapped onto quantum states and in how the required number of qubits scales with input dimensionality. Both approaches were evaluated using variational quantum circuits executed in a simulated, noiseless environment to isolate algorithmic behavior.

In the angle encoding variational quantum classifier, each classical feature was encoded directly into the rotation angle of a single qubit gate, resulting in a linear scaling between the number of input features and the number of qubits. Following the data encoding, a parameterized anasatz composed of entangling gates was applied to introduce nonlinearity and expressivity. Multiple classical optimizers, including COBYLA and SPSA, were explored during training, and only the best performing configurations are reported in the main text.

Amplitude encoding was employed to investigate the potential qubit efficiency offered by logarithmic scaling. In this approach, a normalized classical feature vector of dimension d was embedded into the amplitudes of a quantum state, with feature vectors padded to the nearest power of two when necessary.

While this encoding significantly reduced qubit requirements, it introduced a more complex, entangled state space, which increased the difficulty of optimization and often resulted in reduced stability in the low feature regime.

All quantum circuits were executed using full state vector simulation, without the inclusion of noise models or other constraints. This design choice ensured that observed performance differences reflected limitations of the encoding and optimization strategies rather than artifacts of current quantum hardware.