QuCNet: Quantum Deep Learning Driven Multi-Circuit Network for Remote Sensing Image Classification

Supplementary Material

1. Dataset Description

We evaluate QuCNET across seven publicly available remote sensing datasets spanning diverse tasks (scene classification, disaster recognition, environmental monitoring). The **AID** dataset contains 10,000 aerial RGB images across 30 scene classes, sourced from Google Earth. **AIDER** includes 6,433 disaster images over five categories, offering real-world variability. **UCM** provides 2,100 fine-grained land-use images across 21 classes. **IIITDMJ_Smoke** (21,281 images, four classes) and **USTC_SmokeRS** (6,225 images, six classes) are used for atmospheric event detection. **NWPU-RESISC45 (NPU-45)** comprises 31,500 images over 45 classes with high inter-class similarity. Finally, **EuroSAT** offers 27,000 multi-spectral images across 10 land-use categories. Our motivation for selecting these datasets lies in their complementary challenges ranging from high-resolution, class rich scenes (AID, NPU-45) to low-resolution, and domain specific tasks (AIDER, IIITDMJ_Smoke, USTC_ SmokeRS).

Table 1. Benchmark datasets considered for evaluation of the proposed QuCNET and existing approaches

Dataset	Classes	Samples	Image Size
USTC_SmokeRS	6	6225	256×256
IIITDMJ_Smoke	4	21281	512×512
AIDER	5	6433	400×600
UCM	21	2100	256×256
AID	30	10000	600×600
NPU-45	45	31500	256×256
EuroSAT	10	27000	64×64

These datasets, drawn from the literature, span diverse domains scene classification, disaster recognition, and environmental monitoring and present unique challenges in resolution, class granularity, and feature complexity (refer Figure 1). Our goal is to benchmark QuCNET's quantum enhanced feature extraction and classification capabilities under varied imaging conditions, leveraging the datasets' heterogeneity to probe the model's adaptability to real world remote sensing tasks. Table 1 summarizes the datasets, which vary widely in the number of classes, image resolutions, and sample sizes highlighting the broad applicability and robustness of QuCNET across diverse geospatial tasks.

2. TQC Depth and Complexity Analysis

Our proposed TQC implementation highlights the key practical considerations involved in translating logical quantum circuits into hardware-native operations. When decomposed into the IBM native basis set $[cx, x, rz, sx]$, the proposed 4-qubit TQC achieves a depth of 43 gates, representing a balanced and hardware-efficient design suitable for current NISQ devices. The relationship between qubit scalability and depth complexity for both amplitude and angle encoding is summarized in Table 2.

The decomposed circuit (Figure 2) illustrates the structural efficiency of our design. The logical circuit begins with Hadamard layers for superposition, followed by parameterized R_z and R_x rotations for state preparation and expressive manipulation (see Figure 2 of the main paper). Controlled operations introduce entanglement with minimal two-qubit overhead. The hardware-mapped depth of 43 corresponds to only 2.5-5 μs of execution time, roughly 1.5-5% of the available coherence budget. The regular pattern of single-qubit rotations interspersed with two-qubit entangling operations reflects a hardware-efficient design that balances expressiveness with implementability. The rotation angles (θ values) are optimized parameters that define the quantum state space exploration for our algorithm.

Table 2. Circuit depth comparison across different qubit sizes n for two encoding schemes across same ansatz. Depth values follow typical asymptotic scaling behavior for each method.

Method / Qubit Size (n)	**2**	**4**	**6**	**8**
Amplitude Encoding (approx)	12	64	256	1024
Angle Encoding (approx)				
TQC with Amplitude Encoding	46	107	307	1083
TQC with Angle Encoding	34	43	51	59

To further demonstrate scalability, we provide decomposed hardware circuits for 2-, 4-, 6-, and 8-qubit configurations (2). These visualizations clearly show that the proposed TQC does not exhibit exponential depth growth with increasing qubit count. Instead, the depth increases smoothly and remains well within the shallow-circuit regime required for reliable NISQ execution. The combined quantitative depth analysis and the hardware level circuit visualizations collectively demonstrate that the proposed TQC offers strong scalability, and hardware efficiency for real-device execution on contemporary NISQ platforms.

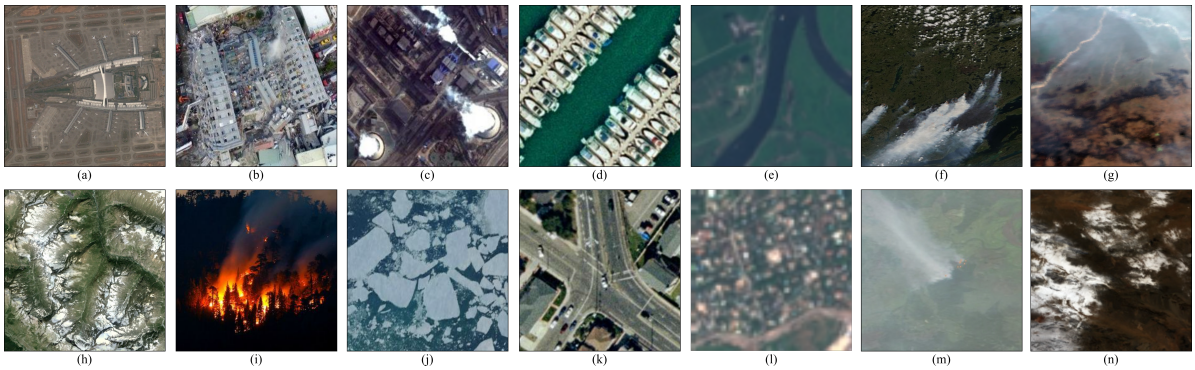


Figure 1. Sample images from each dataset illustrating visual diversity. (a, h) AID, (b, i) AIDER, (c, j) NWPU-RESISC45, (d, k) UC Merced, (e, l) EuroSAT, (f, m) FOG1, and (g, n) FOG2. These samples highlight variations in resolution, scene complexity, and environmental conditions across datasets.

2.1. Limitation

Our current implementation does not deploy amplitude encoding on real hardware, as the full state-preparation overhead remains prohibitive on NISQ devices. We also do not evaluate the end-to-end state-preparation cost of the complete model under amplitude encoding. The present work does not integrate error-mitigation techniques or address measurement noise.

3. Additional Ablation Experiments

3.1. Impact of Feature Vector Dimension (D)

We study the impact of increasing the length of the feature vector (D). Table 3 evaluates QuCNET with different dimensions of the feature vector $D \in \{32, 64, 128\}$ on the AID dataset. Increasing D improves performance in both classical only and hybrid configurations, as larger embeddings provide richer feature representations but at the cost of significantly more parameters in the classical layers.

Table 3. Performance comparison with and without TQC for different feature vector dimensions (D) on AID dataset. Params (Parameters), Acc (Accuracy), P (Precision), R (Recall), F1 (F1-Score)

D	TQC	Params	Acc	P	R	F1
32	w/i	0.023M	82.34	83.21	82.01	82.41
	w/o	0.021M	81.89	82.41	82.28	82.16
64	w/i	0.083M	88.84	89.60	89.30	89.35
	w/o	0.083M	86.26	86.92	86.95	86.84
128	w/i	0.21M	89.08	90.14	89.11	89.01
	w/o	0.30M	90.47	91.11	90.95	90.98

While the TQC consistently improves accuracy for a fixed D ($+3.00\%$ at $D = 64$), the relative quantum advantage diminishes as D increases. At $D = 128$, the classical only model already achieves 90.47%, leaving limited room for improvement by the TQC. This highlights that much of the performance gain at higher D values comes from the increased capacity of the classical encoder rather than quantum processing.

3.2. Impact of Qubit Count

We conducted comprehensive ablation studies to further validate our TQC design choices. First, we investigated the impact of varying qubit counts per TQC layer on both expressibility and classification performance using the AID dataset.

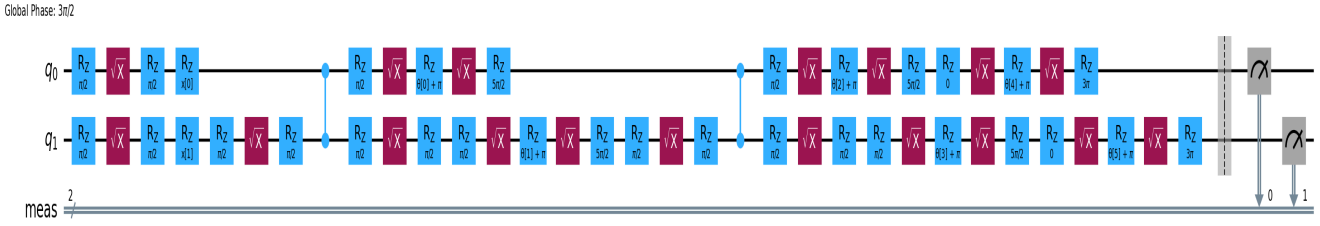
Table 4. Comparison results on the AID dataset evaluating the effect of varying the number of qubits per TQC and the number of parallel circuit layers. Acc (Accuracy)

Qubits	Layers	Expressibility	Acc
4	16	0.01	88.84
6	10	0.27	87.89
8	8	1.01	87.25

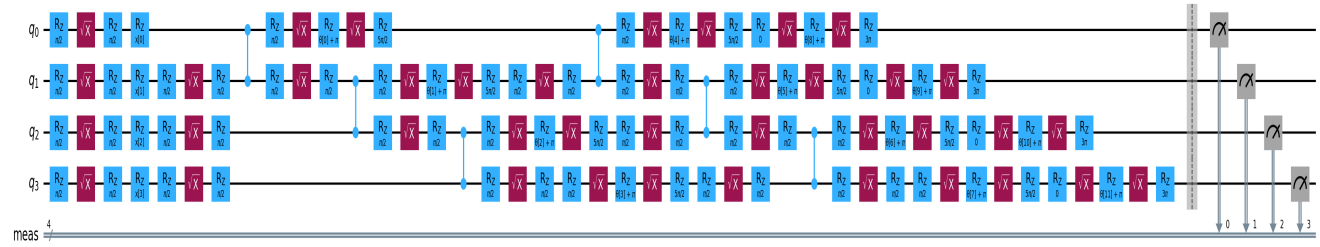
As shown in Table 4, our 4-qubit configuration achieves optimal performance (88.84% accuracy) with the highest expressibility (0.01 KL divergence), while 6-qubit and 8-qubit variants show degraded performance due to reduced expressibility and increased circuit complexity.

3.3. Impact of TQC Parameter Freezing

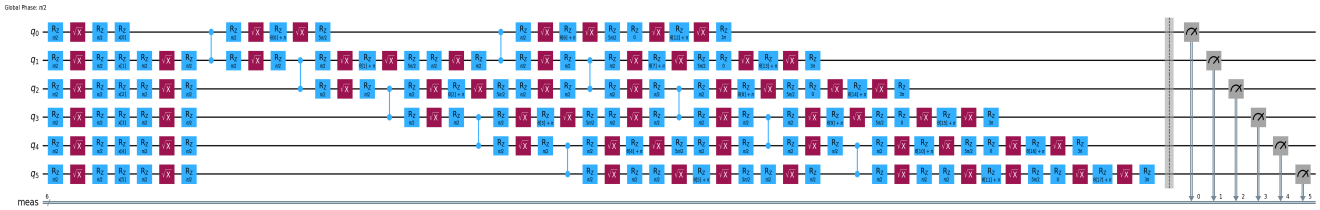
We analyzed potential overparameterization in the TQC by systematically freezing 25%, 50%, and 75% of quantum parameters during training. The results in Table 5 demonstrate performance degradation as more parameters are frozen, with accuracy dropping from 88.84% (fully trainable) to 86.45% (75% frozen). The results indicates that our TQC operates near optimal parameterization without significant redundancy, supporting our design choice of 64 trainable quantum parameters across 16 parallel circuits.



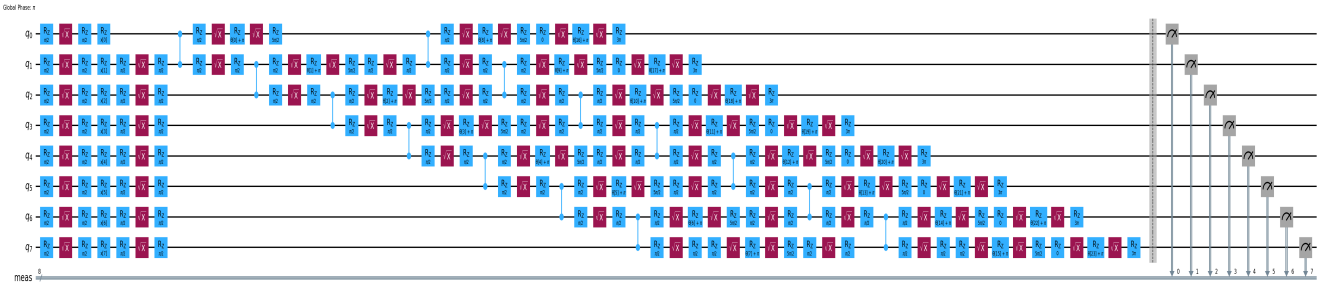
(a) Decomposed 2-Qubit TQC Architecture



(b) Decomposed 4-Qubit TQC Architecture



(c) Decomposed 6-Qubit TQC Architecture



(d) Decomposed 8-Qubit TQC Architecture

Figure 2. Decomposed/Transpiled TQC architectures into native gate sets

Table 5. Performance impact of freezing TQC parameters, used to assess over- and underparameterization. Acc (Accuracy), P (Precision), R (Recall), F1 (F1-Score)

Freeze Ratio (%)	Acc	P	R	F1
0	88.84	89.60	89.30	89.35
25	88.54	89.17	88.80	88.90
50	87.59	88.04	88.18	87.92
75	86.45	88.03	86.84	86.00

3.4. Comparison with Existing VQC Architectures

To isolate the contribution of the proposed TQC, we replace it with variational circuits inspired by prior hybrid quantum classical models while keeping the classical backbone, feature dimension, and training protocol fixed. For methods originally proposed with larger registers, we preserve the original rotation and entanglement pattern but compress the qubit count to four to ensure a fair comparison under identical resources. Table 6 summarizes the architectural complexity and performance of all compared quantum circuits, including the number of trainable quantum parameters, circuit depth, single- and two-qubit gate counts, and expressibility measured using D_{KL}. We additionally report classification accuracy on AID under same training schedules, evaluated with and without HCWS. Despite operating with a comparable gate budget, QuCNet achieves the lowest expressibility divergence and better accuracy across both settings. The perfomance results indicates that our proposed TQC of QuCNet maintains a more desirable balance between expressivity and trainability, avoiding the over-

Table 6. Circuit-only comparison of prior ansätze and the proposed QuCNet under a fixed classical backbone on the AID dataset. All circuits use the same quantum feature dimension. For each method, we report the circuit depth, number of trainable quantum parameters, single- and two-qubit gate counts, and expressibility score D_{KL} (lower is better), along with classification accuracy both without and with HCWS.

Circuit	**Dataset**	**#Classes**	Depth	# Params$_q$	# 1Q gates	# 2Q gates	D_{KL}	Acc (%) (w/o HCWS)	Acc (%) (w/i HCWS)
HQNNC	EuroSAT	Multi	8	4	8	6	0.2	84.92	87.35
HQTL_Resnet34	EuroSAT, AID	Multi	8	12	16	4	0.06	86.80	87.89
QTL_VGG16	EuroSAT, UCM	Binary	7	12	16	4	0.4	85.26	87.54
HybridQC	UCM, NPU-45, AID	Multi	9	8	12	6	0.2	84.67	87.74
QuCNet (ours)	EuroSAT, UCM, AID NPU-45, AIDER USTC_SmokeRS, IIITDMJ_Smoke	Multi	12	16	24	6	**0.01**	**87.59**	**88.84**

parameterization seen in prior designs while still achieving stronger generalization across multi-domain remote sensing datasets.

4. QuCNET Model Complexity

In this paper, we assess QuCNET's complexity by considering both classical and quantum resource requirements, rather than relying solely on traditional classical metrics. While traditional metrics such as the number of parameters and GFLOPs are used to assess the efficiency of classical architectures, hybrid quantum classical models require a more comprehensive analysis. Figure 3 provides the comparison of the proposed QuCNET with state-of-the-art (SOTA) methods in terms of GFLOPs for remote sensing applications. In addition to quantifying quantum components, we explicitly account for qubit count, circuit depth, gate operations, and entanglement topology. We compared our proposed TQC with other SOTA hybrid quantum architectures (see Table 6).

5. Architectural Parameter Analysis

To provide a comprehensive understanding of QuCNET's architectural efficiency, Table 7 presents a detailed layer-wise parameter breakdown for the model configured with feature dimension $D = 64$. The SFE comprises three classical blocks (CB1, CB2, CB3), each designed with residual connections using 1×1 and 3×3 kernels followed by Batch Normalization (BN) and ReLU activation. Pooling layers are employed after each block to downsample spatial resolution.

The TQC module operates on 16 parallel 4-qubit circuits using HCWS. Each circuit applies Z-angle embedding followed by a sequence of entangling (CNOT) and rotational (Rx, Ry) gates with shared weights across cyclic segments. The quantum module adds only 64 parameters. The QCM layer concatenates the Pauli-Z probability values from all

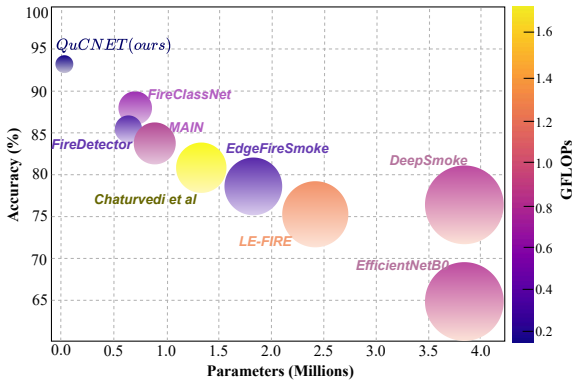


Figure 3. Graph highlights the Performance comparison of the proposed QuCNET with existing models in terms of trainable parameters (x-axis) and accuracy (y-axis) on the USTC_Smoke dataset. Here bubble size $\propto$ Parameters and the color intensity reflects GFLOPs

quantum circuits and feeds them into a fully connected layer for final classification. The structured design of QuCNET emphasizes both efficiency and trainability, ensuring its adaptability to diverse remote sensing domains.

6. QuCNET Interpretability

We investigate the interpretability of QuCNET through multiple analytical approaches. Feature space distributions are visualized using t-SNE to assess class-wise separability in the learned embeddings. Grad-CAM is applied to highlight salient input regions contributing to model predictions, providing spatial attribution for classification decisions (ref Figure 5). Furthermore, gradient variance analysis is conducted to examine learning dynamics and the stability of feature extraction across the hybrid architecture. These interpretability assessments offer insights into QuCNET's internal representations and decision-making behavior. Figure 4 visualizes the t-SNE plots of class-wise fea-

Table 7. Layer-wise parameter configuration of QuCNET (D = 64)

Module	Output Shape	Configuration Details	Params
CB1 (Input)	$16 \times 128 \times 128$	Conv($3 \to 16, 3 \times 3$) + BN + ReLU Conv($16 \to 16, 3 \times 3$) + BN Conv($16 \to 16, 1 \times 1$) + BN AvgPool(2×2) $\times$ 2	480 2,352 256 **Total: 3,088**
CB2	$32 \times 32 \times 32$	Conv($16 \to 32, 3 \times 3$) + BN + ReLU Conv($32 \to 32, 3 \times 3$) + BN Conv($32 \to 32, 1 \times 1$) + BN AvgPool(2×2) $\times$ 2	4,640 9,248 1,056 **Total: 14,944**
CB3	$64 \times 8 \times 8$	Conv($32 \to 64, 3 \times 3$) + BN + ReLU Conv($64 \to 64, 3 \times 3$) + BN Conv($64 \to 64, 1 \times 1$) + BN AvgPool(2×2) $\times$ 2	18,496 37,056 4,032 **Total: 59,584**
GAP + Vectorize	$64 \times 1 \times 1 \to 1 \times 64$	Global AvgPooling + Flatten	0
TQC (HCWS)	16×4 qubit blocks	Z-AngleEmbedding $\to$ RX($\theta_{1\to3}$) $\to$ CNOT($0 \to 2$) $\to$ RX(θ_4) $\to$ RX+RY($\theta_{5\to15}$) repeated cyclically across 16 splits	**64**
Quantum Output Merge	1×256	Pauli-Z measurements + Concatenation	0
Linear Classifier	$256 \to 30$	Fully Connected Layer (QCM Head)	**7,710**
Total Parameters			**86,206**

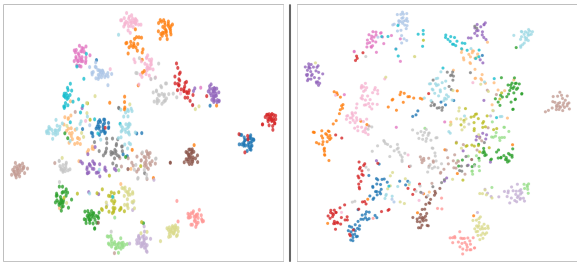


Figure 4. t-SNE visualization of feature embeddings with and without the proposed TQC module in QuCNET. The model with TQC exhibits better class separability and compact clustering, indicating improved discriminative representation learning.

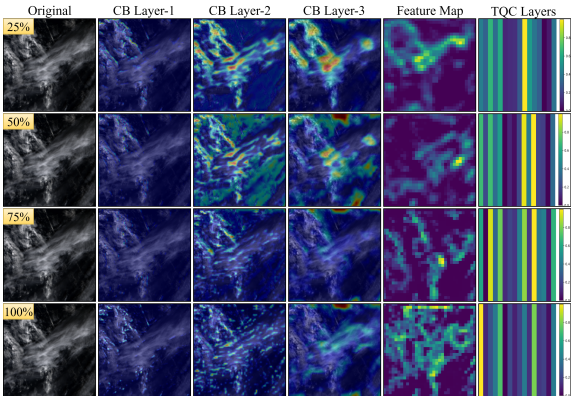


Figure 5. Visualization of feature in QuCNET at different stages of training. Each row corresponds to a checkpoint at 25%, 50%, 75%, and 100% training completion. From left to right: original input image, GradCAM heatmaps from CB Layer-1 to CB Layer-3, intermediate feature maps before the quantum module, and activation dynamics across the TQC layers.

ture embeddings produced by our quantum-enhanced encoder (TQC) and its classical counterpart (w/o TQC). Each color corresponds to a distinct class. Embeddings produced by the QuCNET model form well-separated clusters, indicating better inter-class separability and intra-class consistency. The visualization indicates that TQC produce more expressive feature learning compared to its classical counterpart.

Although the AID dataset is more diverse and challenging, containing 30 classes with high inter-class similarity. To assess the trainability and stability of the quantum layers in QuCNET, we conducted a comprehensive analysis of gradient norms throughout the training process. Figure 6 present the evolution of gradient norms for all quantum layers at intervals of 25 epochs, up to 200 epochs. The figure includes plots of (i) per-layer gradient norms over training steps and (ii) mean gradient norm over time. The gradient norms for all quantum layers remain consistently above 10^{-4}, and do not vanish or explode, indicating effective mitigation of barren plateaus and stable optimization throughout training. All four quantum layers exhibit similar gradient dynamics, showing balanced learning and no dominance or stagnation in any single layer.

Subplots of figure 6 illustrate the consistent and stable gradient behavior across the entire training trajectory, providing visualization for the trainability of the quantum layers in QuCNET.

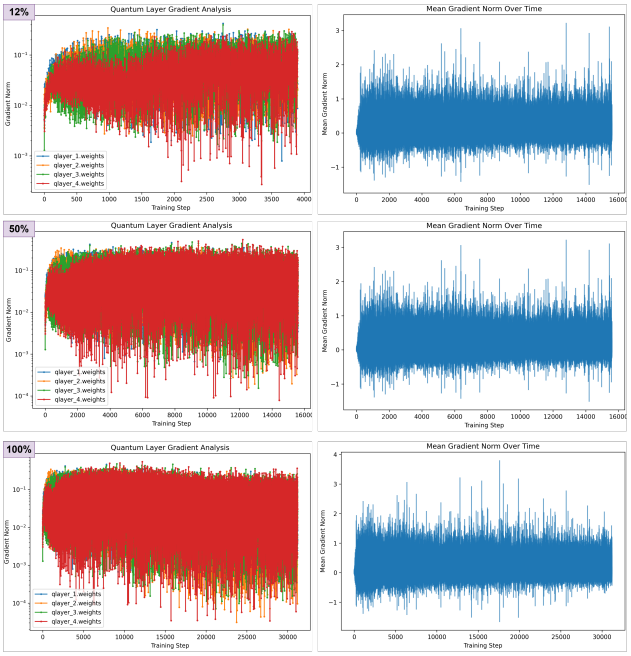


Figure 6. Quantum layer gradient norms and mean gradient norm over time for 12% (early), 50% (mid), and 100% (late) of the training data. Stable gradients are observed across all data regimes, indicating robust trainability of the quantum layer

7. Perfomance Evaluation of TQC Variants

The experimentation conducted on the different variants of quantum circuits (gate layout and depth) to signify the affect on model's ability to learn and generalize, as measured by both expressibility (via KL divergence) and classification accuracy. TQC-2, with an expressibility metric nearly similar to TQC, confirms the importance of selecting an optimal circuit architecture in quantum deep learning, as detailed in the main paper.

The figure 7 presents different TQC variants (TQC-1 to TQC-4) with their respective architectures, Bloch sphere distribution plots, and expressibility analysis using Kullback-Leibler (KL) divergence. A lower KL divergence indicates higher expressibility.

8. Parameter Scheduling and Gradient Dynamics for HCWS

Parameter sharing raises the concern of reduced expressibility. In QuCNet, the SFE module produces globally pooled features that induce correlation among input chunks:

$$\text{Corr}(f^{(i)}, f^{(i+4)}) \geq \rho > 0$$

For moderate correlation ($\rho \approx 0.3$-0.5 typical in remote sensing), the effective parameter dimension is:

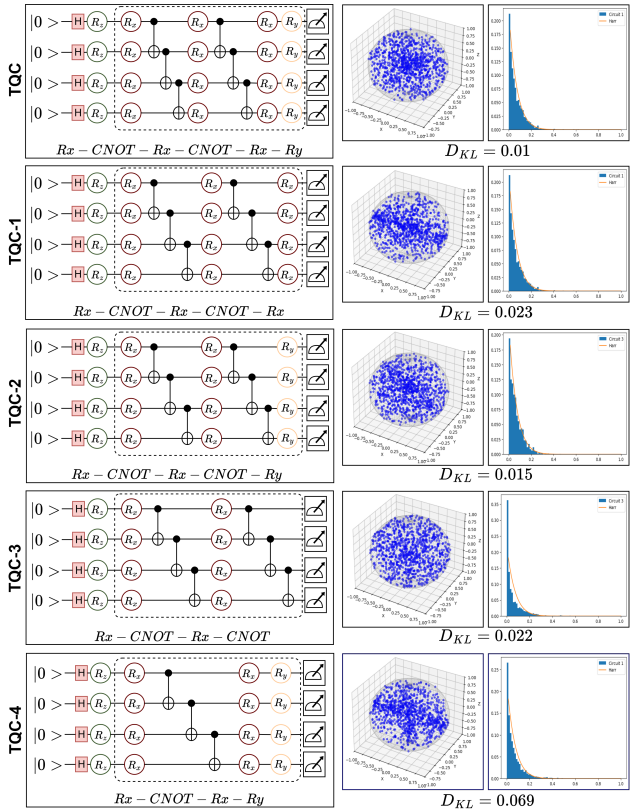


Figure 7. Comparison of expressibility across proposed TQC and its variants variants in QuCNET

$$d_{eff} \approx d_0 \cdot \frac{1 + (n-1)\rho}{1 + (n-1)\rho^2} \tag{1}$$

Under our configuration, $d_{eff} \approx$ 4-6 parameter blocks, validating our choice of cycle period $p = 4$.

Theorem 1 (Expressibility Preservation): Let $\mathcal{S}_{full}$ and $\mathcal{S}_{HCWS}$ denote reachable quantum state spaces under full and HCWS parameterizations. Under mild regularity assumptions:

$$D_{KL}(P_{HCWS} \| P_{Haar}) \leq D_{KL}(P_{full} \| P_{Haar}) + \mathcal{O}(1/p) \tag{2}$$

where P_{Haar} is the Haar-random distribution. The cyclic structure ensures each of p parameter blocks explores distinct regions of the unitary group, with entanglement operations creating non-local correlations that enhance state diversity.

Proof Sketch: Let $\mathcal{L}$ denote the training loss and θ the set of trainable parameters in the VQC. The gradient of the loss with respect to θ is given by:

$$\nabla_\theta \mathcal{L} = \frac{\partial \mathcal{L}}{\partial U} \cdot \frac{\partial U}{\partial \theta}$$

where U denotes the parameterized unitary transformation implemented by the quantum circuit.

In the proposed HCWS scheme, parameters are shared across multiple circuits. Let the full parameter set be partitioned into p parameter-sharing groups (cycle period, $p = 4$):

$$\theta = \{\theta_1, \theta_2, \ldots, \theta_p\}$$

where each θ_i is reused across a subset of quantum circuits.

We decompose the gradient vector as:

$$\nabla_\theta \mathcal{L} = [g_1, g_2, \ldots, g_p]$$

where $g_i = \nabla_{\theta_i} \mathcal{L}$ denotes the gradient corresponding to the i-th parameter group.

The variance of the aggregated gradient depends on both individual variances and cross-correlations:

$$\mathrm{Var}[\nabla_\theta \mathcal{L}] = \frac{1}{p^2} \left(\sum_{i=1}^{p} \mathrm{Var}(g_i) + \sum_{i \neq j} \mathrm{Cov}(g_i, g_j) \right)$$

In independently parameterized circuits, gradient components are approximately uncorrelated, i.e., $\mathrm{Cov}(g_i, g_j) \approx 0$, leading to a rapid decay of variance as system size increases. By enforcing shared parameters across multiple quantum circuits, HCWS reduces independent randomization and couples gradient updates. This structured dependency mitigates concentration effects in high-dimensional spaces, thereby preventing barren plateaus and enabling stable optimization.

8.1. Optimal Cycle Period Selection

We derive the choice of cycle period by analyzing the generalization error through a standard bias-variance decomposition:

$$\text{Error}(p) = \text{Bias}^2(p) + \text{Variance}(p) + \sigma^2_{noise}$$

where $\text{Bias}^2(p) \propto (n/p)^2$ captures underfitting and $\text{Variance}(p) \propto p$ captures overfitting. The optimal period minimizes total error:

$$p^* = \left(\frac{2\alpha n}{\beta} \right)^{1/3} \tag{3}$$

For our datasets ($n = 16$, moderate sample sizes), yields $p^* \approx 4$, consistent with our design.

Proposition 2 (Optimality of p=4): Under squared loss and Gaussian feature distributions, $p = 4$ is optimal within $\{1, 2, 4, 8, 16\}$ for feature dimensionality $d = 64$.

Our ablation results (see Table 4 in main paper) validate this: HCWS (88.84%) outperforms both all-shared (88.05%, high bias) and all-independent (87.59%, high variance).

9. Forward Evolution of the Trainable Quantum Circuit

For clarity of exposition, we present the full analytical derivation using a two-qubit instance of our circuit. This restriction does not limit generality. The proposed TQC consists of the sequence
Hadamard$\rightarrow$ R_z(angle encoding) $\rightarrow$ R_x $\rightarrow$ CNOT $\rightarrow$ R_x $\rightarrow$ CNOT $\rightarrow$ R_y $\rightarrow$ Measurement
The Hadamard gate:

$$H = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$$

Rotation gates:

$$R_x(\theta) = \begin{bmatrix} \cos\frac{\theta}{2} & -i\sin\frac{\theta}{2} \\ -i\sin\frac{\theta}{2} & \cos\frac{\theta}{2} \end{bmatrix}$$

$$R_y(\theta) = \begin{bmatrix} \cos\frac{\theta}{2} & -\sin\frac{\theta}{2} \\ \sin\frac{\theta}{2} & \cos\frac{\theta}{2} \end{bmatrix}$$

$$R_z(\theta) = \begin{bmatrix} e^{-i\theta/2} & 0 \\ 0 & e^{i\theta/2} \end{bmatrix}$$

The generalized representation for 2 qubits:

(a) Hadamard on Both Qubits $|00\rangle$:
Hadamard ($H^{\otimes 2}$) gate applied to generate uniform superposition over all basis states.

$$|\Psi\rangle_H = (H^{\otimes 2}) |00\rangle$$

Matrix form of the $H^{\otimes 2}$ is:

$$H^{\otimes 2} = \frac{1}{2} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix}$$

Since

$$H|0\rangle = \bigotimes_{j=1}^{n} \frac{|0_j\rangle + |1_j\rangle}{\sqrt{2^n}} \quad = \frac{1}{\sqrt{2^n}} \sum_{b_1, b_2 \in \{0,1\}} |b_1 b_2\rangle$$

Where the $\otimes$ is tensor product operator. we get

$$\begin{aligned} |\Psi\rangle_H &= \frac{|0\rangle + |1\rangle}{\sqrt{2}} \otimes \frac{|0\rangle + |1\rangle}{\sqrt{2}} \\ &= \frac{1}{2} \left(|00\rangle + |01\rangle + |10\rangle + |11\rangle \right) \end{aligned}$$

Matrix form:

$$|\Psi\rangle_H = \frac{1}{2} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

$$|\Psi\rangle_H = \frac{1}{2}\begin{bmatrix}1\\1\\1\\1\end{bmatrix}$$

Each computational basis state has

$$\text{Amplitude} = \frac{1}{\sqrt{2^n}}, \qquad \text{Probability} = \frac{1}{2^n}.$$

after the hadamard operation entire system is an equal superposition of the all possible basis states.

(b) Applying Rotation-Z After Hadamard

In our circuit, the R_z gate is used for angle encoding, where input features are mapped directly to phase rotations. As R_z is a diagonal operator, it modifies only the relative phases of the computational basis states without altering their amplitudes. The 2-qubit rotation-Z gate is

$$R_z^{\otimes 2}(\theta_1, \theta_2) = R_z(\theta_1) \otimes R_z(\theta_2).$$

Matrix form:

$$R_z^{\otimes 2} = \begin{bmatrix} e^{-i(\theta_1+\theta_2)/2} & 0 & 0 & 0 \\ 0 & e^{-i(\theta_1-\theta_2)/2} & 0 & 0 \\ 0 & 0 & e^{-i(-\theta_1+\theta_2)/2} & 0 \\ 0 & 0 & 0 & e^{-i(-\theta_1-\theta_2)/2} \end{bmatrix}$$

This matrix is diagonal because R_z only applies phase rotations.

The state after applying $R_z^{\otimes 2}(\theta)$ is

$$|\Psi_{R_z}\rangle = R_z^{\otimes 2}(\theta_i)\ |\Psi_H\rangle$$

It is a simple matrix multiplication of mat1 and mat2:

$$\begin{aligned}|\Psi_{R_z}\rangle = \frac{1}{2}\big(&e^{-i(\theta_1+\theta_2)/2}\,|00\rangle + e^{-i(\theta_1-\theta_2)/2}\,|01\rangle \\ &+ e^{-i(-\theta_1+\theta_2)/2}\,|10\rangle + e^{-i(-\theta_1-\theta_2)/2}\,|11\rangle\big)\end{aligned}$$

Let $\quad e^{-i(\theta_1+\theta_2)/2} = \alpha_1, \qquad e^{-i(\theta_1-\theta_2)/2} = \alpha_2,$
$e^{-i(-\theta_1+\theta_2)/2} = \alpha_3, \qquad e^{-i(-\theta_1-\theta_2)/2} = \alpha_4.$

(c) Apply Rotation-X Gate After Rotation-Z

R_x gate is applied on all qubits to introduce amplitude mixing. The $R_{x_1}^{\otimes 2}(\theta)$ matrix is:

$$R_{x_1}^{\otimes 2}(\theta_3, \theta_4) = R_z(\theta_3) \otimes R_z(\theta_4)$$

$$R_{x_1}^{\otimes 2}(\theta) = \begin{bmatrix} \cos\frac{\theta_3}{2}\cos\frac{\theta_4}{2} & -i\sin\frac{\theta_4}{2}\cos\frac{\theta_3}{2} & -i\sin\frac{\theta_3}{2}\cos\frac{\theta_4}{2} & -\sin\frac{\theta_3}{2}\sin\frac{\theta_4}{2} \\ -i\sin\frac{\theta_4}{2}\cos\frac{\theta_3}{2} & \cos\frac{\theta_3}{2}\cos\frac{\theta_4}{2} & -\sin\frac{\theta_3}{2}\sin\frac{\theta_4}{2} & -i\sin\frac{\theta_3}{2}\cos\frac{\theta_4}{2} \\ -i\sin\frac{\theta_3}{2}\cos\frac{\theta_4}{2} & -\sin\frac{\theta_3}{2}\sin\frac{\theta_4}{2} & \cos\frac{\theta_3}{2}\cos\frac{\theta_4}{2} & -i\sin\frac{\theta_4}{2}\cos\frac{\theta_3}{2} \\ -\sin\frac{\theta_3}{2}\sin\frac{\theta_4}{2} & -i\sin\frac{\theta_3}{2}\cos\frac{\theta_4}{2} & -i\sin\frac{\theta_4}{2}\cos\frac{\theta_3}{2} & \cos\frac{\theta_3}{2}\cos\frac{\theta_4}{2} \end{bmatrix}$$

(Substitution) Let

$$\cos\frac{\theta_3}{2} = \beta_1, \qquad \cos\frac{\theta_4}{2} = \beta_2,$$

$$\sin\frac{\theta_3}{2} = \delta_1, \qquad \sin\frac{\theta_4}{2} = \delta_2$$

Then

$$R_{x_1}^{\otimes 2}(\theta) = \begin{bmatrix} \beta_1\beta_2 & -i\beta_1\delta_2 & -i\delta_1\beta_2 & -\delta_1\delta_2 \\ -i\delta_2\beta_1 & \beta_1\beta_2 & -\delta_1\delta_2 & -i\beta_2\delta_1 \\ -i\beta_2\delta_1 & -\delta_1\delta_2 & \beta_1\beta_2 & -i\delta_2\beta_1 \\ -\delta_1\delta_2 & -i\beta_2\delta_1 & -i\delta_2\beta_1 & \beta_1\beta_2 \end{bmatrix}$$

Final State After Applying $R_{x_1}^{\otimes 2}$
Now:

$$|\Psi_{R_{x_1}}\rangle = R_{x_1}^{\otimes 2}(\theta_i)\ \ |\Psi_{R_z}\rangle\,.$$

$$= \frac{1}{2}\begin{bmatrix} \beta_1\beta_2 & -i\beta_1\delta_2 & -i\delta_1\beta_2 & -\delta_1\delta_2 \\ -i\delta_2\beta_1 & \beta_1\beta_2 & -\delta_1\delta_2 & -i\beta_2\delta_1 \\ -i\beta_2\delta_1 & -\delta_1\delta_2 & \beta_1\beta_2 & -i\delta_2\beta_1 \\ -\delta_1\delta_2 & -i\beta_2\delta_1 & -i\delta_2\beta_1 & \beta_1\beta_2 \end{bmatrix}\begin{bmatrix}\alpha_1\\\alpha_2\\\alpha_3\\\alpha_4\end{bmatrix}$$

State After Applying $R_{x_1}^{\otimes 2}$

$$\begin{aligned}|\Psi_{x_1}\rangle = \frac{1}{2}\Big[&(\beta_1\beta_2\alpha_1 - i\beta_1\delta_2\alpha_2 - i\delta_1\beta_2\alpha_3 - \delta_1\delta_2\alpha_4)\,|00\rangle \\ &+(-i\beta_1\delta_2\alpha_1 + \beta_1\beta_2\alpha_2 - \delta_1\delta_2\alpha_3 - i\beta_2\delta_2\alpha_4)\,|01\rangle \\ &+(-i\delta_1\beta_2\alpha_1 - \delta_1\delta_2\alpha_2 + \beta_1\beta_2\alpha_3 - i\beta_1\delta_2\alpha_4)\,|10\rangle \\ &+(-\delta_1\delta_2\alpha_1 - i\beta_2\delta_2\alpha_2 - i\delta_2\beta_1\alpha_3 + \beta_1\beta_2\alpha_4)\,|11\rangle\Big]\end{aligned}$$

The state now contains a mixture of phase-shifted and amplitude-mixed components. At this point, all qubits are still unentangled because operations were single qubit.

(d) Apply First CNOT (Control = Q0 → Target = Q1)

We applied the CNOT gate on $|\Psi_{R_{x_1}}\rangle$, the effect is simply interchanging the coefficients of $|10\rangle$ and $|11\rangle$.

$$\text{CNOT} = \begin{bmatrix}1&0&0&0\\0&1&0&0\\0&0&0&1\\0&0&1&0\end{bmatrix}$$

Thus:

$$\begin{aligned}|\Psi_{\text{CNOT}}\rangle = \frac{1}{2}\Big[&(\beta_1\beta_2\alpha_1 - i\beta_1\delta_2\alpha_2 - i\delta_1\beta_2\alpha_3 - \delta_1\delta_2\alpha_4)\,|00\rangle \\ &+(-i\beta_1\delta_2\alpha_1 + \beta_1\beta_2\alpha_2 - \delta_1\delta_2\alpha_3 - i\beta_2\delta_2\alpha_4)\,|01\rangle \\ &+(-\delta_1\delta_2\alpha_1 - i\beta_2\delta_2\alpha_2 - i\delta_2\beta_1\alpha_3 + \beta_1\beta_2\alpha_4)\,|10\rangle \\ &+(-i\delta_1\beta_2\alpha_1 - \delta_1\delta_2\alpha_2 + \beta_1\beta_2\alpha_3 - i\beta_1\delta_2\alpha_4)\,|11\rangle\Big]\end{aligned}$$

(Defining Coefficients) Let:

$$x_1 = (\beta_1\beta_2\alpha_1 - i\beta_1\delta_2\alpha_2 - i\delta_1\beta_2\alpha_3 - \delta_1\delta_2\alpha_4)$$

$$x_2 = (-i\beta_1\delta_2\alpha_1 + \beta_1\beta_2\alpha_2 - \delta_1\delta_2\alpha_3 - i\beta_2\delta_2\alpha_4)$$

$$x_3 = (-\delta_1\delta_2\alpha_1 - i\beta_2\delta_2\alpha_2 - i\delta_2\beta_1\alpha_3 + \beta_1\beta_2\alpha_4)$$

$$x_4 = (-i\delta_1\beta_2\alpha_1 - \delta_1\delta_2\alpha_2 + \beta_1\beta_2\alpha_3 - i\beta_1\delta_2\alpha_4)$$

(e) Apply Second Rotation-X Gate

$$|\Psi_{R_{x_2}}\rangle = R_{x_2}^{\otimes 2}(\theta_i)\,|\Psi_{\text{CNOT1}}\rangle$$

$$= \begin{bmatrix} \beta_1\beta_2 & -i\beta_1\delta_2 & -i\delta_1\beta_2 & -\delta_1\delta_2 \\ -i\delta_2\beta_1 & \beta_1\beta_2 & -\delta_1\delta_2 & -i\beta_2\delta_1 \\ -i\beta_2\delta_1 & -\delta_1\delta_2 & \beta_1\beta_2 & -i\delta_2\beta_1 \\ -\delta_1\delta_2 & -i\beta_2\delta_1 & -i\delta_2\beta_1 & \beta_1\beta_2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}$$

State After Second R_x Gate

$$|\Psi_{R_{x_2}}\rangle = [(\beta_1\beta_2x_1 - i\beta_1\delta_2x_2 - i\delta_1\beta_2x_3 - \delta_1\delta_2x_4)\,|00\rangle$$
$$+(-i\delta_2\beta_1x_1 + \beta_1\beta_2x_2 - \delta_1\delta_2x_3 - i\beta_2\delta_1x_4)\,|01\rangle$$
$$+(-i\beta_2\delta_1x_1 - \delta_1\delta_2x_2 + \beta_1\beta_2x_3 - i\delta_2\beta_1x_4)\,|10\rangle$$
$$+(-\delta_1\delta_2 - i\beta_2\delta_1x_2 - i\delta_2\beta_1x_3 + \beta_1\beta_2x_4)\,|11\rangle$$

(f) Applying the Second CNOT Layer

$$|\Psi_{\text{CNOT}_2}\rangle = \text{CNOT}\cdot|\Psi_{R_{x_2}}\rangle$$

CNOT swaps the amplitudes of $|10\rangle$ and $|11\rangle$, so:

$$|\Psi_{\text{CNOT}_2}\rangle = \Big[(\beta_1\beta_2x_1 - i\beta_1\delta_2x_2 - i\delta_1\beta_2x_3 - \delta_1\delta_2x_4)\,|00\rangle$$
$$+(-i\delta_2\beta_1x_1 + \beta_1\beta_2x_2 - \delta_1\delta_2x_3 - i\beta_2\delta_1x_4)\,|01\rangle$$
$$+(-\delta_1\delta_2 - i\beta_2\delta_1x_2 - i\delta_2\beta_1x_3 + \beta_1\beta_2x_4)\,|10\rangle$$
$$+(-i\beta_2\delta_1x_1 - \delta_1\delta_2x_2 + \beta_1\beta_2x_3 - i\delta_2\beta_1x_4)\,|11\rangle\Big]$$

(Defining Coefficients) Let:

$$\lambda_1 = (\beta_1\beta_2x_1 - i\beta_1\delta_2x_2 - i\delta_1\beta_2x_3 - \delta_1\delta_2x_4)$$
$$\lambda_2 = (-i\delta_2\beta_1x_1 + \beta_1\beta_2x_2 - \delta_1\delta_2x_3 - i\beta_2\delta_1x_4)$$
$$\lambda_3 = (-\delta_1\delta_2 - i\beta_2\delta_1x_2 - i\delta_2\beta_1x_3 + \beta_1\beta_2x_4)$$
$$\lambda_4 = (-i\beta_2\delta_1x_1 - \delta_1\delta_2x_2 + \beta_1\beta_2x_3 - i\delta_2\beta_1x_4)$$

In our circuit, an additional R_x rotation followed by a CNOT applied to form a standard variational entangling block. The rotational gate changes amplitudes, and the CNOT distributes correlations across qubits.

(g) Apply the Third Rotation-X Gate

Adding a final R_y rotation introduces another trainable axis of rotation. Since R_x, R_y, R_z span the full SU(2) space, this enables the circuit to reach arbitrary single-qubit transformations before measurement.

$$|\Psi_{R_{x_3}}\rangle = R_{x_3}^{\otimes 2}(\theta)\,|\Psi_{\text{CNOT}_2}\rangle$$

$$= \begin{bmatrix} \beta_1\beta_2 & -i\beta_1\delta_2 & -i\delta_1\beta_2 & -\delta_1\delta_2 \\ -i\delta_2\beta_1 & \beta_1\beta_2 & -\delta_1\delta_2 & -i\beta_2\delta_1 \\ -i\beta_2\delta_1 & -\delta_1\delta_2 & \beta_1\beta_2 & -i\delta_2\beta_1 \\ -\delta_1\delta_2 & -i\beta_2\delta_1 & -i\delta_2\beta_1 & \beta_1\beta_2 \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \\ \lambda_4 \end{bmatrix}$$

$$|\Psi_{R_{x_3}}\rangle = (\beta_1\beta_2\lambda_1 - i\beta_1\delta_2\lambda_2 - i\delta_1\beta_2\lambda_3 - \delta_1\delta_2\lambda_4)\,|00\rangle$$
$$+(-i\delta_2\beta_1\lambda_1 + \beta_1\beta_2\lambda_2 - \delta_1\delta_2\lambda_3 - i\beta_2\delta_1\lambda_4)\,|01\rangle$$
$$+(-i\beta_2\delta_1\lambda_1 - \delta_1\delta_2\lambda_2 + \beta_1\beta_2\lambda_3 - i\delta_2\beta_1\lambda_4)\,|10\rangle$$
$$+(-\delta_1\delta_2\lambda_1 - i\beta_2\delta_1\lambda_2 - i\delta_2\beta_1\lambda_3 + \beta_1\beta_2\lambda_4)\,|11\rangle\,.$$

(Defining Coefficients) Let:

$$\mu_1 = (\beta_1\beta_2\lambda_1 - i\beta_1\delta_2\lambda_2 - i\delta_1\beta_2\lambda_3 - \delta_1\delta_2\lambda_4)$$
$$\mu_2 = (-i\delta_2\beta_1\lambda_1 + \beta_1\beta_2\lambda_2 - \delta_1\delta_2\lambda_3 - i\beta_2\delta_1\lambda_4)$$
$$\mu_3 = (-is_1\beta_2\lambda_1 - s_1s_2\lambda_2 + \beta_1\beta_2\lambda_3 - i\beta_1s_2\lambda_4)$$
$$\mu_4 = (-s_1s_2\lambda_1 - i\beta_1s_2\lambda_2 - is_1\beta_2\lambda_3 + \beta_1\beta_2\lambda_4)$$

(h) Apply Rotational-Y After Third Rotational-X Gate

$$R_y^{\otimes 2}(\theta) = \begin{bmatrix} \cos\frac{\theta_5}{2}\cos\frac{\theta_6}{2} & -\sin\frac{\theta_6}{2}\cos\frac{\theta_5}{2} & -\sin\frac{\theta_5}{2}\cos\frac{\theta_6}{2} & \sin\frac{\theta_5}{2}\sin\frac{\theta_6}{2} \\ \sin\frac{\theta_6}{2}\cos\frac{\theta_5}{2} & \cos\frac{\theta_5}{2}\cos\frac{\theta_6}{2} & -\sin\frac{\theta_5}{2}\sin\frac{\theta_6}{2} & -\sin\frac{\theta_5}{2}\cos\frac{\theta_6}{2} \\ \sin\frac{\theta_5}{2}\cos\frac{\theta_6}{2} & -\sin\frac{\theta_5}{2}\sin\frac{\theta_6}{2} & \cos\frac{\theta_5}{2}\cos\frac{\theta_6}{2} & -\sin\frac{\theta_6}{2}\cos\frac{\theta_5}{2} \\ \sin\frac{\theta_5}{2}\sin\frac{\theta_6}{2} & \sin\frac{\theta_5}{2}\cos\frac{\theta_6}{2} & \cos\frac{\theta_5}{2}\sin\frac{\theta_6}{2} & \cos\frac{\theta_5}{2}\cos\frac{\theta_6}{2} \end{bmatrix}$$

(Substitution)

$$\cos\frac{\theta_5}{2} = p, \quad \sin\frac{\theta_5}{2} = q, \quad \cos\frac{\theta_6}{2} = r, \quad \sin\frac{\theta_6}{2} = s$$

Thus,

$$R_y^{\otimes 2}(\theta) = \begin{bmatrix} pr & -ps & -qr & qs \\ ps & pr & -qs & -qr \\ qr & -qs & pr & -ps \\ qs & qr & ps & pr \end{bmatrix}$$

Now apply R_y Gate to State $|\Psi_{R_{x_3}}\rangle$

$$|\Psi_{Ry}\rangle = R_y^{\otimes 2}(\theta)\,|\Psi_{R_{x_3}}\rangle$$

$$= \begin{bmatrix} pr & -ps & -qr & qs \\ ps & pr & -qs & -qr \\ qr & -qs & pr & -ps \\ qs & qr & ps & pr \end{bmatrix} \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \\ \mu_4 \end{bmatrix}$$

$$|\Psi_{Ry}\rangle = (pr\mu_1 - ps\mu_2 - qr\mu_3 + qs\mu_4)\,|00\rangle$$
$$+(ps\mu_1 + pr\mu_2 - qr\mu_3 + ps\mu_4)\,|01\rangle$$
$$+(qr\mu_1 - qs\mu_2 + pr\mu_3 - ps\mu_4)\,|10\rangle$$
$$+(qs\mu_1 + qr\mu_2 + ps\mu_3 + pr\mu_4)\,|11\rangle$$

At the final stage, a projective measurement is performed in the computational basis. Measurement collapses the quantum state onto one of the basis states, and the probability of obtaining outcome is given by

$$p(\phi) = |\langle\phi \mid \Psi_{\text{final}}\rangle|^2$$

Algorithm 1 Training Procedure for QuCNET

Require: Dataset $\mathcal{G} = \{(X_i, Y_i)\}_{i=1}^{N}$, learning rate η, epochs E

Initialize: CNN weights W, TQC parameters $\vec{\theta}$
for $e = 1$ to E **do**
  for $X_i, Y_i \in \mathcal{D}$ **do**
    Step 1: Preprocess X_i (normalize to $[0, 1]$)
    Step 2: Extract features via SFE
      $F_1 \leftarrow \text{LeakyReLU}(\text{BN}(\text{Conv}_{3\times3}(X_i)))$
      $F_1 \leftarrow \text{Pool}_{2\times2}(F_1))$
      $F_{1-2} \leftarrow \text{BN}(\text{Conv}_{3\times3}(F_1))$
      $F_2 \leftarrow \text{Pool}_{2\times2}(\text{BN}(\text{Conv}_{1\times1}(X_i)))$
      $F_2 \leftarrow \text{LeakyReLU}(F_{1-2} + F_2)$
      $f_c^{\text{final}} \leftarrow tanh(\frac{1}{H \cdot W} \sum_{h=1}^{H} \sum_{w=1}^{W} F_2(h, w, c))$
    Step 3: Encode into TQC
      Partition f_c^{final} into 16 segments
      $f^{(i)} = \{f^{(1)}, f^{(2)}, f^{(3)}, ..., f^{(16)}\}$

    for $f_c^{\text{final}} \in f^{(i)}$ **do** ▷ $f^{(1)} = \{f_0, f_1, f_2, f_3\}$
      $|\psi_{f^{(i)}}\rangle \leftarrow R_Z(x_l)H|0\rangle$ ▷ Encode per qubit

    Step 4: Process with TQC
      $|\Psi_{\text{final}}\rangle \leftarrow U'' U_{\text{CNOT}} U' U_{\text{CNOT}} U |\Psi_{f^{(i)}}\rangle$
      $P_k \leftarrow |\langle k|\Psi_{\text{final}}\rangle|^2, k = 1 \dots 16$
    Step 5: Classify with QCM
      $P_{\text{final}} \leftarrow \text{concat}(P_1, \dots, P_{16})$
      $y \leftarrow \text{FC}(P_{\text{final}}, O_{\text{classes}})$
    Step 6: Optimize: $L \leftarrow \text{CrossEntropy}(y, Y_i)$
      Update W and $\vec{\theta}$ via gradient descent with η
Output: Trained W, $\vec{\theta}$, evaluation metrics