

## Highlights

### **FireQuan: A lightweight hybrid quantum-classical architecture with Patch Embedding for multi-domain image classification**

Quang Nhan Hoang, Trung Thanh Pham, Nhut Minh Nguyen, Linh Le, Choong Seon Hong, Duc Ngoc Minh Dang

- FireQuan: A lightweight hybrid quantum-classical model for multi-domain image classification.
- Patch-based encoding eliminates CNOT gates, reducing circuit depth by 99.6% vs. FRQI/NEQR.
- Fire512 Head reduces parameters by 98.9% and FLOPs by 98% compared to ResNet50.
- Competitive accuracy on 13 datasets while being 85 times smaller than ResNet50.
- Maintains a stable generalization gap below 10% on noisy and imbalanced data.

# FireQuan: A lightweight hybrid quantum-classical architecture with Patch Embedding for multi-domain image classification

Quang Nhan Hoang<sup>①a</sup>, Trung Thanh Pham<sup>②b</sup>, Nhut Minh Nguyen<sup>①a</sup>, Linh Le<sup>③c</sup>, Choong Seon Hong<sup>④d</sup>, Duc Ngoc Minh Dang<sup>①a,\*</sup>

<sup>a</sup>*AiTA Lab, Faculty of Information Technology, FPT University, D1 Street,  
Saigon Hi-tech Park, Tang Nhon Phu Ward, Ho Chi Minh City, 71320, Vietnam*

<sup>b</sup>*Department of Artificial Intelligence, Kyung Hee University, Yongin, Republic of Korea*

<sup>c</sup>*McGill University, affiliated with Mila, 845 Sherbrooke Street West, Montreal,  
QC H3A 0G4, Quebec, Canada*

<sup>d</sup>*Department of Computer Science and Engineering,  
Kyung Hee University, Yongin, 446-701, Republic of Korea*

---

## Abstract

Quantum Machine Learning (QML) offers notable computational advantages; however, the limited qubit count and system noise inherent to the Noisy Intermediate-Scale Quantum (NISQ) era pose significant obstacles to processing real-world images and large, specialized datasets. To address these challenges, we introduce FireQuan, a hybrid quantum-classical architecture for multi-domain image classification. The framework centers on two contributions: (1) the Fire512 Head, a compact convolutional feature extractor that reduces the number of parameters by up to 98.90% and FLOPs by over 98.00% compared to ResNet50, while preserving network depth for learning complex features; and (2) a patch-based encoding strategy that combines amplitude and angle encoding principles with data re-uploading to load classical features into qubits using only single-qubit rotation gates, thereby eliminating Controlled-NOT (CNOT) gates during the encoding phase. This encoding reduces physical circuit depth by over 99.60% and the total gate count by over 97.00% relative to Flexible Representation of Quantum Images (FRQI) and Novel Enhanced Quantum Representation (NEQR) for equivalent-sized feature vectors. Empirical evaluation across 13 datasets spanning 5 domains

---

\*Corresponding author: Duc Ngoc Minh Dang (ducdnm2@fe.edu.vn)

demonstrates that FireQuan performs competitively, achieving 95.74% on EuroSAT and 86.70% on PatchCamelyon (PCAM), while outperforming several Quantum Support Vector Machine (QSVM), Quantum Convolutional Neural Network (QCNN), and contemporary hybrid methods. FireQuan maintains a generalization gap below 10.00% even on datasets with high noise and class imbalance, highlighting its practical value for current quantum systems.

*Keywords:* Quantum Machine Learning, Hybrid Quantum-Classical Model, Multi-Domain Image Classification, Patch Embedding, Lightweight Architecture

---

## 1. Introduction

Image classification is increasingly important across industries such as medicine, transportation, and agriculture, where it reduces manual labor and improves accuracy. Deep learning models such as ResNet (He et al., 2016) and VGGNet (Simonyan and Zisserman, 2015) achieve strong results but require large datasets and extensive training, incurring high computational costs that limit their use in resource-constrained settings. As application domains incorporating heterogeneous data continue to grow, implementing multi-domain classification on edge devices becomes increasingly difficult (Manukian et al., 2019; Wang et al., 2020) due to model size, computational cost, and poor convergence in data-scarce environments.

Quantum machine learning (QML) has emerged as a promising approach to address these limitations (Bressan, 1978; Morandi, 2011; Alda, 1980). QML integrates classical theories with high-dimensional Hilbert spaces, enabling the modeling of complex systems through superposition and quantum entanglement. However, in the Noisy Intermediate-Scale Quantum (NISQ) era (Preskill, 2018), limited qubit availability, shallow circuit depth, and hardware noise constrain the implementation of fully quantum approaches. Consequently, hybrid quantum-classical models have become the preferred solution, where classical neural networks perform feature extraction while parameterized quantum circuits (PQCs) provide enhanced nonlinear representation with relatively few parameters (Benedetti et al., 2019; Xie et al., 2025; Park et al., 2025; Chen et al., 2025; He et al., 2024; Li et al., 2023; Baek et al., 2023; Liang et al., 2021; Ganjefar et al., 2015; da Silva et al., 2016; Zhang et al., 2022; Marcianò et al., 2022).

Despite these advances, most QML research relies on small-scale datasets (Senokosov et al., 2024; Long et al., 2025) such as MNIST (Deng, 2012), FashionMNIST (Xiao et al., 2017), or CIFAR10 (Giuste and Vizcarra, 2020), which do not capture the complexity of real-world multi-domain classification, including data noise, variable shapes, and distortions. Many existing hybrid architectures exhibit limited generalization on heterogeneous, noisy data (Wu et al., 2025). Overfitting occurs when classical backbones have too many parameters or when data encoding fails to preserve local features (Benedetti et al., 2019; Xie et al., 2025). Minimizing model parameters while achieving adequate accuracy, especially in medical or agricultural contexts, remains a challenge (Wu et al., 2025; Biamonte et al., 2017; Schuld and Killoran, 2022).

To address these challenges, our contributions are as follows:

- We introduce FireQuan, a lightweight hybrid quantum-classical architecture for multi-domain image classification that uses 85 times fewer parameters than ResNet50 (He et al., 2016).
- We design Fire512 Head, a compact convolutional backbone based on SqueezeNet (Iandola et al., 2016), which preserves network depth while reducing the number of parameters by up to 98.90% and FLOPs by 98.00% compared to ResNet50.
- We propose a patch-based encoding strategy that integrates amplitude and angle encoding principles with data re-uploading to encode features using only single-qubit gates, reducing physical circuit depth by over 99.60% and total gate count by over 97.00% compared with FRQI (Parigi et al., 2025) and NEQR (Zhang et al., 2013) for equivalent-sized inputs.

The paper is structured as follows: Section 2 reviews related work. Section 3 presents the methodology. Section 4 describes the experimental setup. Section 5 reports results. Section 6 presents ablation studies and further analysis. Section 7 concludes the paper.

## 2. Related work

QML development faces fundamental challenges in the NISQ era, constrained by qubit count, circuit depth, gate precision, and computational

cost. A core tension exists: increasing the efficiency of data encoding into the Hilbert space often requires higher circuit complexity, whereas reducing complexity may degrade encoding quality. This section reviews key research directions in data encoding methods and hybrid quantum-classical architectures, identifying gaps that motivated FireQuan.

### *2.1. Quantum encoding methods*

Encoding classical data into quantum states is a critical step in QML, directly influencing performance and hardware requirements (Huang et al., 2021; Zheng et al., 2025). Existing methods can be grouped into four categories based on their mapping principles and hardware costs (Pande, 2024; Khan et al., 2024; Sharma and N, 2024).

#### *2.1.1. Direct mapping encodings*

This group includes Basis Encoding (Benioff, 1980), Unary Encoding (Romero et al., 2017), Angle Encoding (Mitarai et al., 2018), and Phase Encoding (Coppersmith, 2002). Each data feature is mapped to a basis state or single-qubit gate configuration. These methods have simple construction and low circuit depth since no two-qubit gates are used, making them noise-resistant. However, qubit requirements scale linearly with input dimensionality, often exceeding the capacity of current quantum computers for high-dimensional data. They are therefore most suitable for small or low-dimensional datasets.

#### *2.1.2. Global compression encodings*

To address qubit limitations, Amplitude Encoding compresses  $N$  features into  $\log_2 N$  qubits by encoding data as quantum-state probability amplitudes. While theoretically optimal for qubit efficiency, this approach requires deep circuits with many two-qubit gates for state preparation, which is problematic for NISQ hardware. Furthermore, the near-uniform distribution of encoded data in the Hilbert space reduces measurement efficiency and contributes to barren plateaus during training (Wang et al., 2025).

#### *2.1.3. Image-specific quantum encodings*

Specialized methods such as FRQI (Parigi et al., 2025) and NEQR (Zhang et al., 2013; Haque et al., 2023) preserve both pixel values and spatial structure by encoding images as quantum states. However, these techniques require substantial numbers of multi-qubit control gates, particularly CNOT

gates, resulting in rapidly increasing circuit depth. As image size grows, the required gates and circuit depth exceed the coherence time of current devices, rendering these methods primarily suitable for idealized simulations (Oukaira, 2025; McDonough et al., 2022).

These trade-offs highlight the need for encoding methods that minimize reliance on two-qubit gates, maintain low circuit depth, and preserve local information, a gap that motivates our patch-based approach.

## 2.2. Hybrid quantum-classical models

Hardware constraints make it difficult to run fully quantum models on large datasets. Hybrid quantum-classical models combine classical neural networks for feature extraction with quantum circuits for nonlinear representation in Hilbert space (Liu et al., 2021; Hasan and Mahdy, 2025). CNNs are commonly used as backbones, leveraging spatial structure through weight sharing and local convolutions (O’Shea and Nash, 2015).

Although hybrid approaches reduce qubit requirements and increase training stability, most rely on large-scale or pretrained CNN backbones. This dependence incurs significant parameter costs and reduces the intended computational efficiency. Moreover, pretrained models are typically optimized for specific domains and may underperform on different datasets (Zhang et al., 2023; Zhou et al., 2023). Some research attempts to remove the classical component entirely by encoding inputs directly into quantum circuits (Han et al., 2025; Alam and Ghosh, 2022; Schuld et al., 2020; Jouzdani et al., 2024), but these methods are limited to small datasets due to qubit and circuit-depth constraints.

These observations motivate the development of hybrid architectures that balance efficient classical preprocessing with compact quantum processing, meeting NISQ-era hardware constraints.

## 2.3. Lightweight CNN architectures

Lightweight CNN architectures such as MobileNet (Howard et al., 2017) and ShuffleNet (Zhang et al., 2018) reduce FLOPs through depthwise separable convolutions and channel shuffling. However, their parameter counts remain too high for direct integration into hybrid QML systems with limited qubits. SqueezeNet (Iandola et al., 2016) achieves accuracy comparable to AlexNet with 50 times fewer parameters, using Fire modules that compress and expand features efficiently. Each Fire module consists of a Squeeze layer  $S$  with  $1 \times 1$  filters and an Expand layer merging  $E_{1 \times 1}$  and  $E_{3 \times 3}$  branches

with  $1 \times 1$  and  $3 \times 3$  filters, respectively. The structure is illustrated in Figure 1 and formalized in Eqs. 1–4.

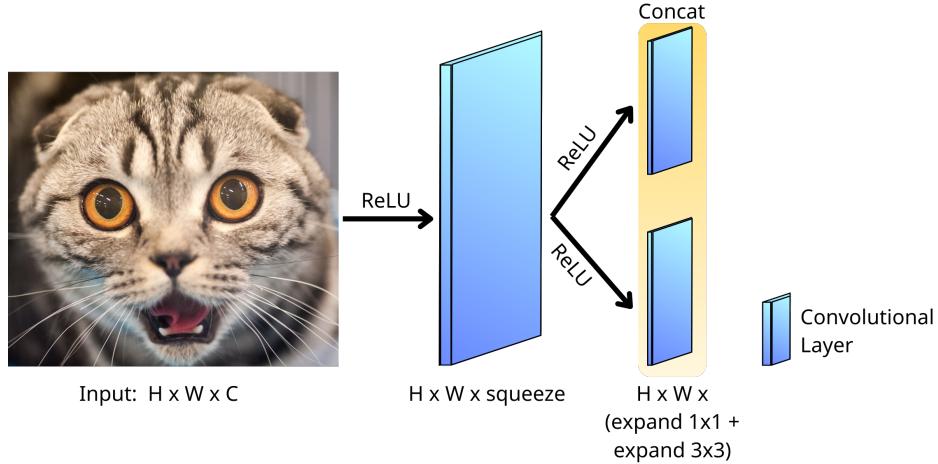


Figure 1: Structure of a Fire Module.

$$S = \text{ReLU}(\text{Conv}_{1 \times 1}(\mathbf{X}, \mathbf{W}_S)) \in \mathbb{R}^{H \times W \times C_S}, \quad (1)$$

$$E_{1 \times 1} = \text{ReLU}(\text{Conv}_{1 \times 1}(S, \mathbf{W}_{E_{1 \times 1}})), \quad (2)$$

$$E_{3 \times 3} = \text{ReLU}(\text{Conv}_{3 \times 3}(S, \mathbf{W}_{E_{3 \times 3}})), \quad (3)$$

$$E_{final} = \text{Concat}(E_{1 \times 1}, E_{3 \times 3}) \in \mathbb{R}^{H \times W \times (C_{E_{1 \times 1}} + C_{E_{3 \times 3}})}, \quad (4)$$

where all convolutional layers use stride 1 and  $3 \times 3$  filters are padded by 1 to preserve spatial dimensions. The ReLU activation (Agarap, 2018) is defined as  $\text{ReLU}(x) = \max(0, x)$ .

While SqueezeNet provides greater depth with fewer parameters than other classical benchmarks, directly integrating pretrained models into hybrid QML systems poses challenges. Pretrained models are optimized for specific domains (e.g., ImageNet), which limits adaptability and increases the risk of negative transfer on multi-domain datasets. Furthermore, fine-tuning the backbone can cause divergence in the randomly initialized quantum component (Ceschini et al., 2025; Long et al., 2025). These considerations motivate the design of Fire512 Head, a reorganized and more compact architecture that retains network depth while further reducing parameters for efficient quantum integration.

### 3. Methodology

#### 3.1. Overall Architecture

FireQuan consists of four sequential modules: (1) the Fire512 Head for classical feature extraction, (2) Patch Embedding for quantum data encoding, (3) a QNN for nonlinear feature processing in Hilbert space, and (4) a classical linear classifier, as shown in Figure 2. An input image of size  $224 \times 224 \times 3$  is processed by Fire512 Head, which extracts features and reduces the data to a 512-dimensional vector. Patch Embedding then partitions this vector and encodes it onto qubits using single-qubit rotation gates. The QNN produces nonlinear features through entanglement operations, and the resulting expectation values serve as input to a fully connected classifier for label prediction. Each module is described below.

#### 3.2. Fire512 Head

Fire512 Head is a compact feature extractor designed to produce a fixed-dimensional latent representation suitable for quantum encoding, while maintaining sufficient network depth for hierarchical feature learning. The architecture is illustrated in Figure 3.

**Input and initial convolution.** The input is a  $224 \times 224 \times 3$  image tensor. The first layer applies a  $3 \times 3$  convolution with 32 filters and stride 2, followed by ReLU activation, to extract low-level features with direct spatial downsampling. A subsequent max-pooling layer ( $3 \times 3$  kernel, stride 2) further reduces the spatial dimensions, yielding a  $56 \times 56 \times 32$  tensor. This initial downsampling reduces the computational burden on subsequent layers and provides translational invariance.

**Hierarchical Fire module blocks.** The feature maps then pass through three processing blocks, each consisting of two consecutive Fire modules followed by a max-pooling layer:

- **Block 1:** Squeeze planes = 8, expand planes = 32 per branch ( $1 \times 1$  and  $3 \times 3$ ). The output is a  $28 \times 28 \times 64$  tensor after pooling. This block captures basic morphological features such as edges and textures.
- **Block 2:** Squeeze planes = 16, expand planes = 64 per branch. The output is a  $14 \times 14 \times 128$  tensor after pooling. This block reconstructs intermediate spatial-geometric features, with the increased channel count compensating for reduced spatial dimensions.

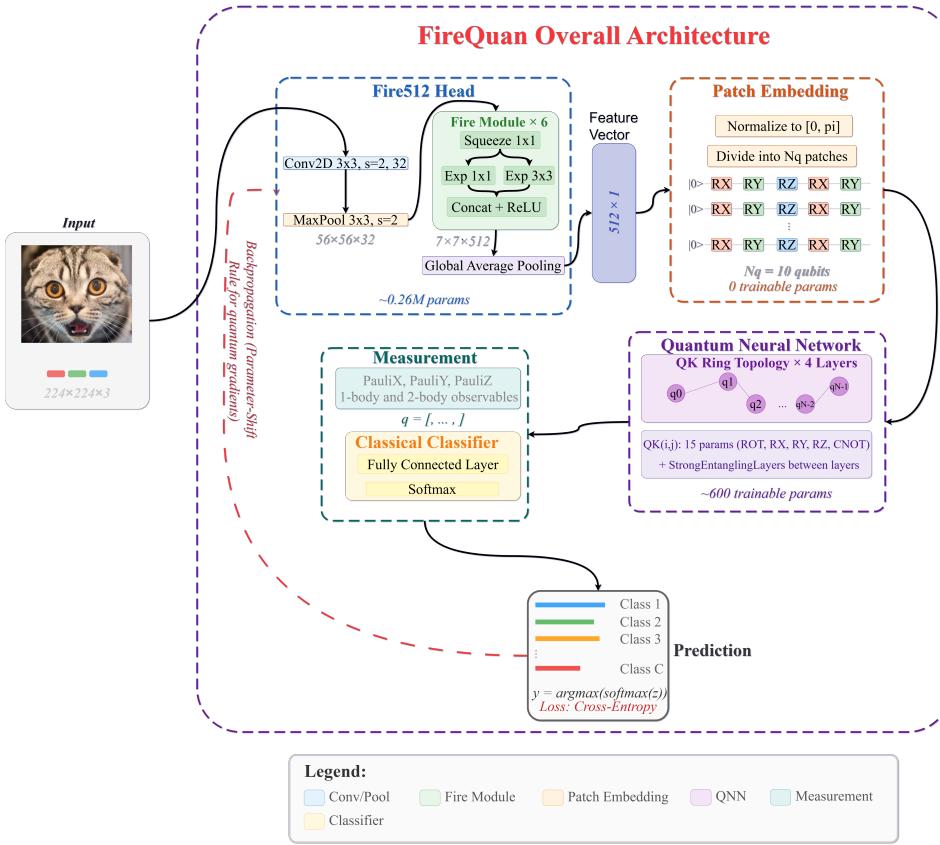


Figure 2: Pipeline of FireQuan. The Fire512 Head extracts features from the input image, Patch Embedding encodes them into a quantum circuit using single-qubit rotation gates, the QNN generates nonlinear features in Hilbert space, and a linear classifier produces the final prediction.

- **Block 3:** Squeeze planes = 32, expand planes = 128 per branch. The output is a  $7 \times 7 \times 512$  tensor after pooling. This block extracts the most abstract and semantically rich features.

**Global Average Pooling (GAP).** Finally, GAP (Lin et al., 2014) aggregates each of the 512 feature maps into a single scalar value, producing a 512-dimensional feature vector:

$$A = [a_1, a_2, \dots, a_{512}]^\top \in \mathbb{R}^{512}, \quad (5)$$

$$a_k = \frac{1}{H' \times W'} \sum_{i=1}^{H'} \sum_{j=1}^{W'} \mathbf{Y}_{i,j,k}, \quad k = 1, 2, \dots, 512, \quad (6)$$

where  $H'$ ,  $W'$  are the spatial dimensions of the final feature map  $\mathbf{Y}$ . Unlike flattening, GAP preserves spatial invariance and does not increase the parameter count, making it suitable as a deterministic interface for subsequent quantum encoding.

The fixed 512-dimensional output ensures consistent patch partitioning for the Patch Embedding module, regardless of the input content. With only 0.26M parameters, Fire512 Head is approximately 85 times smaller than ResNet50 while preserving 8 convolutional layers of depth for hierarchical feature extraction.

### 3.3. Patch Embedding

Amplitude Encoding is qubit-efficient but produces near-uniform distributions in Hilbert space that impede optimization (Wang et al., 2025; Mitsuda et al., 2024; Daimon and Matsushita, 2024). Meanwhile, Angle Encoding preserves gradient information but requires one qubit per feature, which is infeasible for high-dimensional inputs. Inspired by data re-uploading (Pérez-Salinas et al., 2020), which has shown that a single qubit can approximate arbitrary functions through sequential rotations, we propose Patch Embedding, a strategy that partitions the feature vector into patches and loads each patch onto a dedicated qubit using sequential single-qubit rotation gates. This approach combines the qubit efficiency of compression methods with the gradient-preserving properties of angle-based encodings, while eliminating all two-qubit gates during the encoding phase. The method is illustrated in Figure 4.

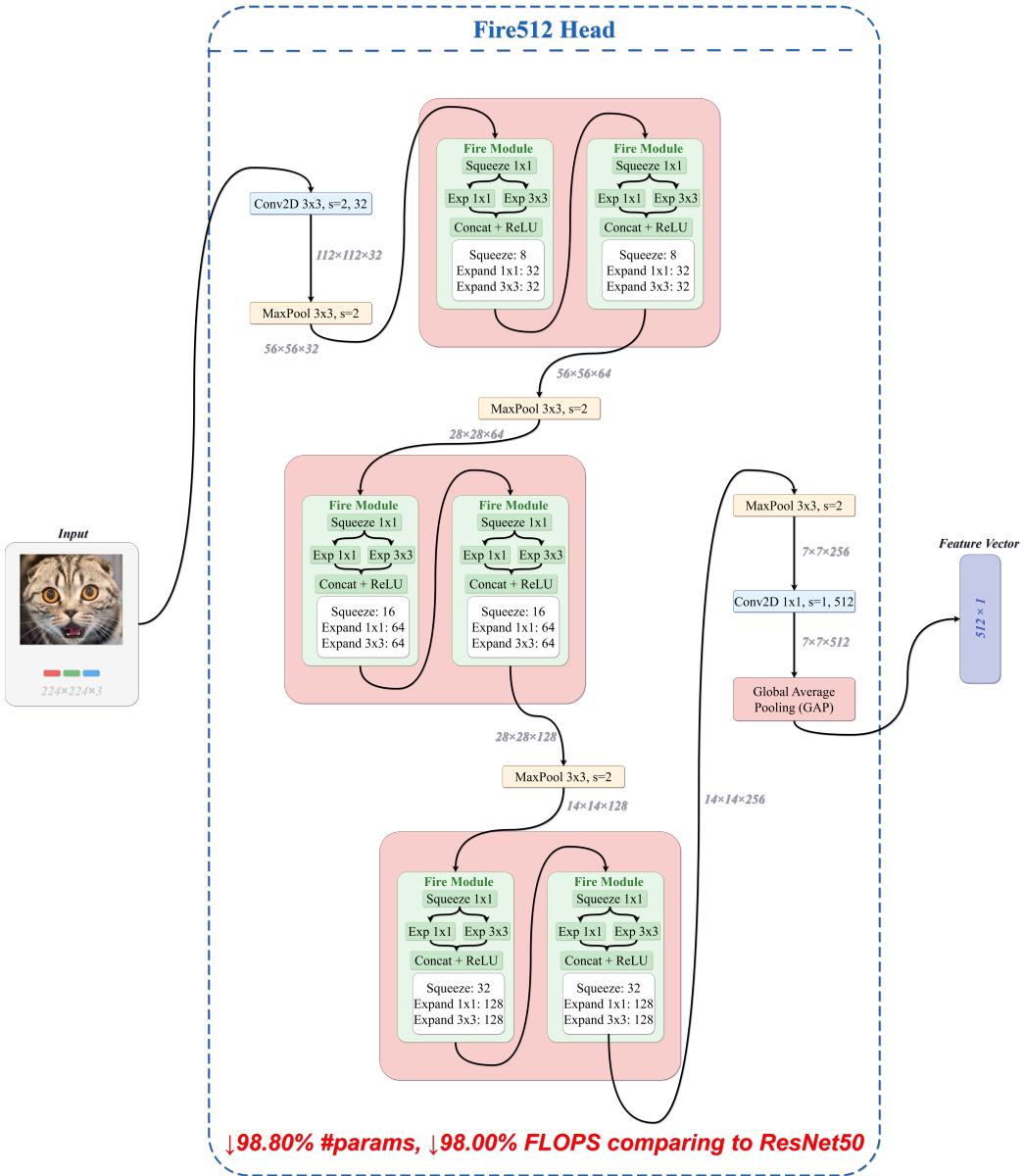


Figure 3: Structure of Fire512 Head, which extracts features from an input of shape  $224 \times 224 \times 3$  and produces a 512-dimensional feature vector.

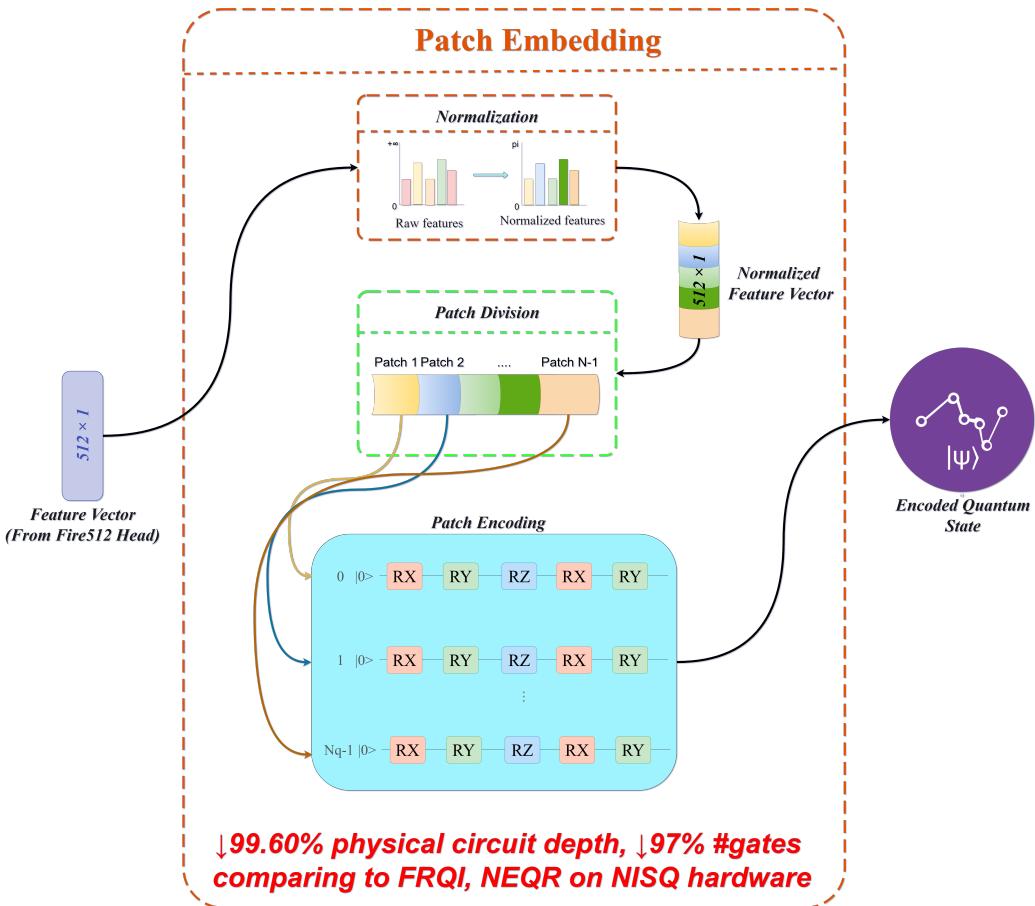


Figure 4: Patch Embedding: each qubit receives a patch of features encoded through sequential  $R_X$ ,  $R_Y$ ,  $R_Z$  rotation gates. No two-qubit gates are used.

**Normalization.** The 512-dimensional feature vector  $A$  from Fire512 Head is first scaled to  $[0, \pi]$ :

$$a'_k = \pi \cdot \frac{a_k - \min(A)}{\max(A) - \min(A) + \epsilon}, \quad (7)$$

where  $a'_k \in [0, \pi]$  and  $\epsilon = 10^{-7}$  prevents division by zero.

**Partitioning.** The normalized vector is divided into  $N_q$  patches (one per qubit). The  $i$ -th patch is:

$$A'_i = \left[ a'_{i \cdot \lfloor 512/N_q \rfloor}, a'_{i \cdot \lfloor 512/N_q \rfloor + 1}, \dots, a'_{(i+1) \cdot \lfloor 512/N_q \rfloor - 1} \right]. \quad (8)$$

**Encoding.** Each patch is encoded onto its corresponding qubit through a sequence of  $R_X$ ,  $R_Y$ ,  $R_Z$  rotation gates applied cyclically. The  $R_X$  and  $R_Y$  gates transform amplitudes, whereas  $R_Z$  modulates the quantum phase, creating nonlinear correlations among features within each patch. The encoding unitary for the  $i$ -th qubit is:

$$U_{\text{encode}}^i(A'_i) = \prod_{j=1}^m R_{\tau(j)}(a'_{i,j}), \quad (9)$$

where  $\tau(j) \in \{X, Y, Z\}$  cycles through the rotation axes and  $m = \lfloor 512/N_q \rfloor$  is the number of features per patch.

**Key advantage.** The complete elimination of two-qubit gates (e.g., CNOT) during encoding is the primary distinction of this method. On current hardware, two-qubit gates exhibit error rates 10–100 times higher than single-qubit gates (Arute et al., 2019). By using only single-qubit rotations, this approach minimizes noise accumulation and preserves coherence for longer periods before subsequent QNN computation. Furthermore, each qubit encodes  $\lfloor 512/N_q \rfloor$  features (approximately 51 for  $N_q = 10$ ), achieving substantial data compression without the deep state-preparation circuits required by Amplitude Encoding.

The final quantum state is a tensor product of individually encoded qubit states:

$$|\psi\rangle = \bigotimes_{i=0}^{N_q-1} (U_{\text{encode}}^i(A'_i)|0\rangle_i), \quad (10)$$

$$\rho = |\psi\rangle\langle\psi|. \quad (11)$$

Note that after Patch Embedding, the qubits are not yet entangled, correlations between different patches are established in the subsequent QNN stage. This separation of concerns allows the encoding to remain shallow and error-resistant, while the QNN handles inter-patch feature interactions.

### 3.4. Quantum Neural Network

After encoding, the QNN serves as a nonlinear feature processor in Hilbert space, as illustrated in Figure 5. Its central component is the Quantum Kernel ( $QK_{i,j}$ ), which operates on pairs of qubits to establish entanglement and extract nonlinear correlations from the independently encoded patches.

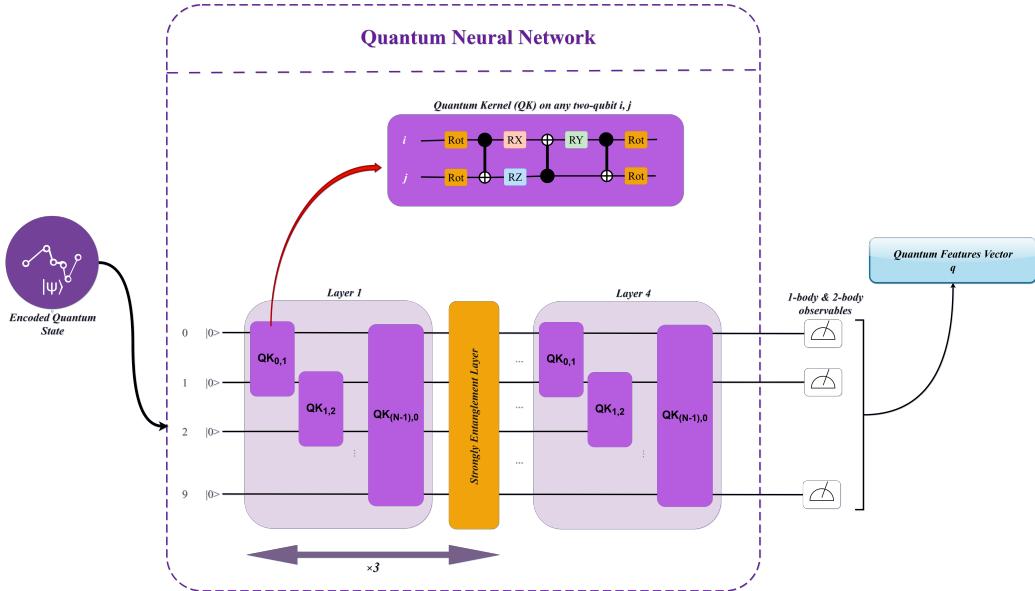


Figure 5: QNN architecture: Quantum Kernels ( $QK$ ) arranged in a ring topology with 1-body and 2-body observables for measurement. Four layers are stacked with intermediate strongly entangling layers.

**Quantum Kernel design.** Each  $QK_{i,j}$  contains 15 trainable parameters: four generalized rotation ( $ROT$ ) gates providing 12 angular parameters and three individual rotation gates ( $R_X$ ,  $R_Y$ ,  $R_Z$ ) refining features along each coordinate axis. Three CNOT gates establish entanglement between the qubit pair. This configuration balances expressibility and trainability, sufficient to prevent underfitting without overparameterization.

**Ring topology.** Rather than all-to-all connectivity, which requires  $O(N_q^2)$  CNOT gates, the QKs are arranged in a ring topology where each qubit in-

teracts with its nearest neighbor. This reduces both gate count and physical circuit depth while enabling information propagation through the full system across multiple layers.

**QNN construction.** FireQuan stacks four quantum layers with strongly entangling layers (Schuld et al., 2020) inserted between them to enable high-order feature interactions. The construction is detailed in Algorithm 1.

---

**Algorithm 1** QNN constructed from  $QK$  and Strongly Entangling Layers.

---

**Input:** State  $|\psi_{in}\rangle$ , parameters  $\mathbf{W}$ ,  $N_q$  qubits,  $N_{layer}$  quantum layers.

**Output:** Quantum circuit for a QNN.

```

1:  $|\psi\rangle \leftarrow |\psi_{in}\rangle$ 
2: for  $k = 1$  to  $N_{layer}$  do
3:   for  $i = 0$  to  $N_q - 1$  do
4:      $|\psi\rangle \leftarrow QK_{i, (i+1) \bmod N_q}(\mathbf{W}_{k,i}) |\psi\rangle$ 
5:   end for
6:   if  $k < N_{layer}$  then
7:      $|\psi\rangle \leftarrow \text{StrongEntanglingLayers}(|\psi\rangle)$ 
8:   end if
9: end for
```

---

All QNN parameters are optimized using the parameter-shift rule (Wierichs et al., 2022), which computes exact quantum gradients without classical approximation.

### 3.5. Measurement

After QNN processing, qubits are measured to extract classical features. We employ both one-body and two-body observables. A one-body observable yields:

$$\langle M_i \rangle = \langle \psi | M_i | \psi \rangle, \quad (12)$$

while a two-body observable measures correlations between qubit pairs:

$$\langle M_{ij} \rangle = \langle \psi | M_i \otimes M_j | \psi \rangle, \quad (13)$$

where  $M_i, M_j \in \{\text{PauliX}, \text{PauliY}, \text{PauliZ}\}$ . After performing  $k$  measurements, the resulting feature vector is:

$$\mathbf{q} = [\langle M_1 \rangle, \langle M_2 \rangle, \dots, \langle M_k \rangle]^\top \in \mathbb{R}^k. \quad (14)$$

The combination of one-body and two-body observables captures both individual qubit information and pairwise entanglement structure, producing features with high non-linearity.

### 3.6. Classical Classifier

A single fully connected layer maps quantum features to class logits:

$$\mathbf{z} = \mathbf{W}\mathbf{q} + \mathbf{b}, \quad (15)$$

where  $\mathbf{W} \in \mathbb{R}^{C \times k}$  and  $\mathbf{b} \in \mathbb{R}^C$  are trainable parameters. A simple linear classifier is used for two reasons: it maximizes the quantum component's contribution by avoiding complex post-processing, and quantum features already exhibit enhanced non-linearity from Hilbert space operations. The model is trained using Cross-Entropy Loss (Mao et al., 2023) with Softmax:

$$\hat{\mathbf{y}}_i = \frac{\exp(z_i)}{\sum_{j=1}^{N_{\text{classes}}} \exp(z_j)}, \quad i \in \{1, 2, \dots, N_{\text{classes}}\}, \quad (16)$$

$$\mathcal{L}(y, \hat{y}) = - \sum_{i=1}^{N_{\text{classes}}} y_i \log(\hat{y}_i), \quad (17)$$

where  $y$  and  $\hat{y}$  are the ground-truth and predicted labels, respectively.

## 4. Experimental Setup

### 4.1. Datasets

We evaluate FireQuan across 13 benchmark datasets from five domains, summarized in Table 1.

**Autonomous Driving and Traffic Sign Classification.** GTSRB (Stal-kamp et al., 2012) ( $\sim 50,000$  images, 43 classes) exhibits significant variation in illumination, perspective, and blur. BelgiumTS (Timofte et al., 2009) ( $\sim 9,000$  images, 62 classes) contains high noise levels and imbalanced class distributions under natural capture conditions.

**Agricultural and Plant Classification.** Fruit360 ( $\sim 90,000$  images, 131 classes), PlantVillage ( $\sim 50,000$  images, 38 classes), and DeepWeeds ( $\sim 17,000$  images, 9 classes) span controlled and natural environments with varying complexity.

**Medical and Biomedical Imaging.** PatchCamelyon (PCAM) (Veeling et al., 2018) (327,680 histopathology images, 2 classes), ISIC2019 (Hernández-Pérez et al., 2024; Codella et al., 2018) ( $> 25,000$  dermatology images, 8 classes), and HAM10000 (Tschantl, 2018) (10,015 images, 7 classes with severe class imbalance).

**Digit and Character Recognition.** EMNIST By Class (Cohen et al., 2017) (814,255 images, 62 classes) and SVHN (Netzer et al., 2011) (>600,000 images, 10 classes with background noise and overlapping digits).

**Earth Observation Image Classification.** EuroSAT (Helber et al., 2018) (27,000 images, 10 classes), RESISC45 (Cheng et al., 2017) (31,500 images, 45 classes), and UC Merced (Neumann et al., 2019) (2,100 images, 21 classes).

Table 1: Overview of 13 benchmark datasets across five domains.

Domain	Dataset	Total Samples	Classes
Autonomous Driving – Traffic Sign Classification	GTSRB	50,000	43
	BelgiumTS	9,000	62
Agricultural and Plant Classification	Fruit360	90,000	131
	PlantVillage	50,000	38
	DeepWeeds	17,000	9
Medical and Biomedical Imaging	PCAM	327,680	2
	ISIC2019	25,000	8
	HAM10000	10,015	7
Digit and Character Recognition	EMNIST (By Class)	814,255	62
	SVHN	600,000	10
Earth Observation Image Classification	EuroSAT	27,000	10
	RESISC45	31,500	45
	UC Merced	2,100	21

#### 4.2. Environmental Setup and Hyperparameters

FireQuan is implemented in Python using JAX, FLAX, PennyLane, Qiskit, and PyTorch. All experiments run on an NVIDIA A30 GPU (24GB vRAM) with a fixed random seed of 42. All models (both hybrid and classical) are trained from scratch without pretrained weights for 200 epochs with two warmup epochs. Early stopping halts training if no improvement occurs for 10 consecutive epochs. The learning rate is scheduled from  $2 \times 10^{-4}$  to  $1 \times 10^{-6}$ . All other hyperparameters (batch size, optimizer, loss function) are identical across models to ensure fair comparison. All experiments are conducted in simulation; results on physical quantum hardware are left for future work.

## 5. Experimental Results

### 5.1. Performance Comparison with Classical Baselines

We evaluate FireQuan across all five domains. Tables 2–6 report test accuracy compared with classical baselines, all trained from scratch under identical conditions.

**Digit and Character Recognition** (Table 2). FireQuan achieves 87.07% on EMNIST, within 1.3% of the best classical model, and 94.55% on SVHN, comparable to larger architectures.

Table 2: Performance comparison on Digit and Character Recognition datasets.

Models	EMNIST (By Class)	SVHN
SqueezeNet (Iandola et al., 2016)	87.17%	74.68%
ShuffleNet (Zhang et al., 2018)	<b>88.37%</b>	95.10%
EfficientNet-B0 (Tan and Le, 2020)	88.21%	<b>95.82%</b>
MobileNetV3 (Howard et al., 2017)	86.20%	95.39%
ResNet18 (He et al., 2016)	87.80%	95.32%
ResNet50 (He et al., 2016)	88.10%	95.09%
FireQuan (Ours)	87.07%	94.55%

**Autonomous Driving – Traffic Sign Classification** (Table 3). FireQuan achieves 95.21% on GTSRB, approximately 2–3% below the best classical models. On BelgiumTS, accuracy decreases to 85.98% due to high noise and class imbalance, though it remains more stable than SqueezeNet (80.85%).

Table 3: Performance comparison on Autonomous Driving – Traffic Sign Classification datasets.

Models	GTSRB	BelgiumTS
SqueezeNet (Iandola et al., 2016)	78.63%	80.85%
ShuffleNet (Zhang et al., 2018)	97.13%	96.55%
EfficientNet-B0 (Tan and Le, 2020)	94.24%	94.99%
MobileNetV3 (Howard et al., 2017)	90.78%	96.07%
ResNet18 (He et al., 2016)	<b>98.61%</b>	<b>97.96%</b>
ResNet50 (He et al., 2016)	97.28%	97.72%
FireQuan (Ours)	95.21%	85.98%

**Agricultural and Plant Classification** (Table 4). FireQuan achieves 96.63% on Fruit360 and 95.06% on PlantVillage. On DeepWeeds, which contains complex natural scenes, FireQuan achieves 73.47%, close to ResNet18

(74.59%) and higher than SqueezeNet (71.70%), ShuffleNet (69.88%), and MobileNetV3 (66.00%).

Table 4: Performance comparison on Agricultural and Plant Classification datasets.

Models	Fruit360	PlantVillage	DeepWeeds
SqueezeNet (Iandola et al., 2016)	99.54%	77.03%	71.70%
ShuffleNet (Zhang et al., 2018)	98.86%	98.82%	69.88%
EfficientNet-B0 (Tan and Le, 2020)	98.31%	<b>99.11%</b>	<b>84.46%</b>
MobileNetV3 (Howard et al., 2017)	97.45%	96.58%	66.00%
ResNet18 (He et al., 2016)	<b>99.55%</b>	98.11%	74.59%
ResNet50 (He et al., 2016)	98.66%	97.82%	79.63%
FireQuan (Ours)	96.63%	95.06%	73.47%

**Medical and Biomedical Imaging** (Table 5). FireQuan achieves 86.70% on PCAM, outperforming ResNet50 (85.21%) and EfficientNet-B0 (83.02%). On HAM10000, FireQuan achieves 76.39%, the highest among all models tested. These results suggest that the quantum circuit’s nonlinear representational capacity on the Bloch sphere is particularly beneficial for biomedical data with low contrast and indistinct lesion boundaries.

Table 5: Performance comparison on Medical and Biomedical Imaging datasets.

Models	PCAM	ISIC2019 Skin	HAM10000
SqueezeNet (Iandola et al., 2016)	81.51%	65.62%	73.69%
ShuffleNet (Zhang et al., 2018)	82.37%	64.87%	75.39%
EfficientNet-B0 (Tan and Le, 2020)	83.02%	67.08%	75.09%
MobileNetV3 (Howard et al., 2017)	84.48%	66.77%	72.59%
ResNet18 (He et al., 2016)	83.72%	67.30%	73.59%
ResNet50 (He et al., 2016)	85.21%	65.13%	74.99%
FireQuan (Ours)	<b>86.70%</b>	<b>67.91%</b>	<b>76.39%</b>

**Earth Observation Image Classification** (Table 6). FireQuan achieves the highest accuracy on EuroSAT (95.74%). On RESISC45 (67.16%) and UC Merced (62.24%), performance is lower, which we attribute to the limited representational capacity of the compact Fire512 Head on datasets where complex spatial hierarchies and large receptive fields are critical.

**Summary.** Across 13 datasets, FireQuan achieves accuracy comparable to or marginally lower than classical CNNs with 4–85 times more parameters.

Table 6: Performance comparison on Earth Observation Image Classification datasets.

Models	EuroSAT	RESISC45	UC Merced
SqueezeNet (Iandola et al., 2016)	93.76%	54.38%	52.34%
ShuffleNet (Zhang et al., 2018)	94.35%	74.67%	61.98%
EfficientNet-B0 (Tan and Le, 2020)	95.00%	84.39%	80.47%
MobileNetV3 (Howard et al., 2017)	88.39%	75.06%	48.18%
ResNet18 (He et al., 2016)	87.69%	77.14%	<b>86.20%</b>
ResNet50 (He et al., 2016)	95.30%	<b>77.77%</b>	83.07%
FireQuan (Ours)	<b>95.74%</b>	67.16%	62.24%

FireQuan outperforms all baselines on PCAM, HAM10000, and EuroSAT. On two datasets (BelgiumTS, RESISC45), performance gaps of 10–20% reflect the deliberate architectural trade-off of the compact Fire512 Head, which sacrifices representational capacity for parameter efficiency. These datasets favor high-capacity models with large receptive fields and deep feature hierarchies. Importantly, FireQuan exhibits stable training without extreme overfitting on these challenging datasets, suggesting robustness under noise and class imbalance.

### 5.2. Comparative Analysis with Quantum Architectures

We compare FireQuan with three quantum methods: QSVM (Havlíček et al., 2019), QCNN (Cong et al., 2019), and QAE (Liu et al., 2024; Asaoka and Kudo, 2025) on four datasets.

**With classical backbone** (Table 7). The three baseline methods use a pretrained ResNet18 (He et al., 2016) backbone (parameters frozen) with Amplitude Encoding. FireQuan uses its own Fire512 Head trained from scratch. FireQuan substantially outperforms all three methods across all four datasets. The gap is especially large on PlantVillage (95.06% vs. 67.95%) and SVHN (94.55% vs. 42.07%). We attribute this to two factors: the frozen pretrained backbone was optimized on ImageNet, whose distribution differs from these datasets; and Amplitude Encoding produces near-uniform representations on the Bloch sphere, limiting the quantum classifiers’ discriminative ability.

**Quantum circuit only** (Table 8). To isolate the quantum component’s contribution, we compare the QNN from FireQuan against QSVM, QCNN, and QAE without any classical backbone. All methods use Amplitude En-

Table 7: FireQuan vs. quantum methods with ResNet18 backbone (frozen pretrained weights + Amplitude Encoding).

Datasets	QSVM	QCNN	QAE	FireQuan (Ours)
EuroSAT	67.80%	70.78%	82.98%	<b>95.74%</b>
PCAM	83.37%	82.71%	83.39%	<b>86.70%</b>
PlantVillage	46.39%	49.33%	67.95%	<b>95.06%</b>
SVHN	35.17%	34.97%	42.07%	<b>94.55%</b>

coding with 10 qubits for fair comparison. Note that the overall accuracies are low across all methods because no classical feature extraction is performed. Even so, FireQuan’s QNN achieves the highest accuracy on all four datasets, outperforming alternatives by 2–11%, demonstrating the advantage of its ring-topology design and layered entanglement structure.

Table 8: Quantum circuit comparison without classical backbone (all methods use Amplitude Encoding, 10 qubits).

Datasets	QSVM	QCNN	QAE	FireQuan (Ours)
EuroSAT	27.97%	31.31%	32.02%	<b>39.88%</b>
PCAM	62.15%	61.61%	62.11%	<b>63.50%</b>
PlantVillage	29.13%	31.18%	31.61%	<b>42.96%</b>
SVHN	19.59%	20.24%	22.14%	<b>24.34%</b>

### 5.3. Benchmarking against Published Quantum and Hybrid Methods

Due to the limited number of published quantum methods with reported results on large-scale image datasets, we compare FireQuan with available works on three datasets where direct comparison is possible.

On PCAM (Table 9), FireQuan achieves 86.70%, surpassing a hybrid SVM classifier (Goyal et al., 2025) (85.55%, no pretrained backbone) and a hybrid ResNet18 architecture (Baral et al., 2025) (84.77%, with pretrained backbone).

On HAM10000 (Table 10), FireQuan achieves 76.39%, exceeding a hybrid QCNN with attention (Pandey and Mandal, 2025) (71.00%) and another hybrid architecture (V et al., 2025) (70.13%).

On EuroSAT (Table 11), FireQuan achieves 95.74%, above a quantum-inspired optimization method (Saravanan and N, 2025) (95.20%) and a hybrid model (Zaidenberg et al., 2021) (94.73%).

All FireQuan results are achieved without pretrained backbones. While the number of directly comparable works is limited, reflecting the current state of QML research where large-scale evaluation remains constrained, these results provide evidence that carefully designed lightweight hybrid architectures can achieve competitive performance.

Table 9: Comparison on the PCAM dataset with published quantum/hybrid methods.

Studies	Architecture	Pretrained Backbone	Test Accuracy
(Goyal et al., 2025)	Hybrid SVM Classifier	✗	85.55%
(Baral et al., 2025)	Hybrid ResNet18	✓	84.77%
FireQuan (Ours)	Hybrid Quantum-Classical	✗	<b>86.70%</b>

Table 10: Comparison on the HAM10000 dataset with published quantum/hybrid methods.

Studies	Architecture	Pretrained Backbone	Test Accuracy
(Pandey and Mandal, 2025)	Hybrid QCNN with attention	✗	71.00%
(V et al., 2025)	Hybrid Quantum-Classical	✗	70.13%
FireQuan (Ours)	Hybrid Quantum-Classical	✗	<b>76.39%</b>

Table 11: Comparison on the EuroSAT dataset with published quantum/hybrid methods.

Studies	Architecture	Pretrained Backbone	Test Accuracy
(Saravanan and N, 2025)	Quantum Hippopotamus Optimization	✗	95.20%
(Zaidenberg et al., 2021)	Hybrid Quantum-Classical	✗	94.73%
FireQuan (Ours)	Hybrid Quantum-Classical	✗	<b>95.74%</b>

## 6. Discussion

### 6.1. Ablation Study

To quantify each module’s contribution, we evaluate variants of FireQuan with individual components removed on four datasets (Table 12).

**Fire512 Head.** Removing Fire512 Head requires resizing images to  $18 \times 18 \times 3$  and using Amplitude Encoding directly. This causes severe information loss, with accuracies dropping to 24–64%. The result confirms that Fire512 Head is essential for preserving key features while reducing dimensionality to fit within limited qubit resources.

Table 12: Ablation study: test accuracy when individual modules are removed.

Configuration	EuroSAT	PCAM	PlantVillage	SVHN
w/o Fire512 Head	39.88%	63.50%	42.96%	24.34%
w/o QNN	91.50%	76.98%	94.29%	91.74%
w/o Patch Embedding	90.31%	77.59%	93.47%	91.91%
FireQuan (Full)	<b>95.74%</b>	<b>86.70%</b>	<b>95.06%</b>	<b>94.55%</b>

**QNN.** Without the QNN, Fire512 Head features are mapped directly to the classifier. Performance drops consistently (e.g., PCAM: 86.70% → 76.98%; EuroSAT: 95.74% → 91.50%), indicating that the QNN’s nonlinear transformations in Hilbert space provide discriminative power beyond what the classical features alone offer.

**Patch Embedding.** Replacing Patch Embedding with a single FC layer that reduces features to 10 dimensions for standard Angle Encoding degrades performance across all datasets (e.g., EuroSAT: 95.74% → 90.31%). Patch Embedding’s ability to distribute 512 features across qubits via sequential rotation gates encodes substantially more information than a 10-dimensional projection.

The three modules serve complementary roles: Fire512 Head captures hierarchical visual features, Patch Embedding distributes them efficiently across qubits, and the QNN introduces nonlinear correlations through entanglement.

### 6.2. Backbone Computational Efficiency

Table 13 compares Fire512 Head with classical backbones (classifiers removed) on computational efficiency metrics.

Fire512 Head uses 0.26M parameters, reductions of 63.9%, 91.2%, and 98.9% relative to SqueezeNet, MobileNetV3, and ResNet50, respectively. The FP32 model is 0.98 MB (91× smaller than ResNet50), and INT8 quantization reduces it to 0.25 MB. At 0.07 GFLOPs, it is approximately 73% lower than SqueezeNet and 98% lower than ResNet50. On GPU, Fire512 Head achieves 550.2 FPS with 1.82 ms inference time. Estimated energy consumption is 0.4543 J, approximately 30% lower than SqueezeNet and 86% lower than EfficientNet-B0. These characteristics confirm Fire512 Head’s suitability for resource-constrained deployment and integration with quantum processing modules.

Table 13: Backbone computational efficiency and quantization analysis.

Model	Device	#Params (M)	FLOPs (G)	Inference time (ms)	FPS	FP32 size (MB)	INT16 size (MB)	INT8 size (MB)	Est. energy (J)
Fire512 Head (Ours)	CPU	0.26	0.07	6.29	158.9	0.98	0.49	0.25	0.4543
Fire512 Head (Ours)	GPU	0.26	0.07	1.82	550.2	0.98	0.49	0.25	0.4543
SqueezeNet	CPU	0.72	0.26	8.52	117.4	2.76	1.38	0.69	0.6563
SqueezeNet	GPU	0.72	0.26	2.63	381.0	2.76	1.38	0.69	0.6563
ShuffleNet	CPU	1.25	0.15	21.90	45.7	4.78	2.39	1.20	2.4879
ShuffleNet	GPU	1.25	0.15	9.95	100.5	4.78	2.39	1.20	2.4879
EfficientNet-B0	CPU	4.01	0.41	38.30	26.1	15.29	7.64	3.82	3.2067
EfficientNet-B0	GPU	4.01	0.41	12.83	78.0	15.29	7.64	3.82	3.2067
MobileNetV3	CPU	2.97	0.23	25.73	38.9	11.34	5.67	2.83	2.2807
MobileNetV3	GPU	2.97	0.23	9.12	109.6	11.34	5.67	2.83	2.2807
ResNet18	CPU	11.18	1.82	18.50	54.0	42.64	21.32	10.66	0.9591
ResNet18	GPU	11.18	1.82	3.84	260.6	42.64	21.32	10.66	0.9591
ResNet50	CPU	23.51	4.13	44.40	22.5	89.68	44.84	22.42	2.1973
ResNet50	GPU	23.51	4.13	8.79	113.8	89.68	44.84	22.42	2.1973

### 6.3. Hardware-aware Cost Analysis of Patch Embedding

To assess feasibility on real quantum hardware, we compiled all encoding methods into physical circuits using IBM’s native gate set via Qiskit. The input is a fixed 512-dimensional feature vector. Table 14 reports qubits, physical circuit depth, total gates, and CNOT gate count.

Table 14: Hardware cost of encoding methods for a 512-dimensional input, compiled to IBM’s native gate set.

Methods	#Qubits	Physical Depth	#Physical Gates	#CNOT	Time (ms)	Memory (MB)
Basis Encoding (Benioff, 1980)	512	1	268	0	—	—
Amplitude Encoding (Grover, 1996)	9	2,370	2,401	502	263.02	0.10
Phase Encoding (Coppersmith, 2002)	512	4	2,048	0	—	—
FRQI (Le et al., 2011)	10	138,450	165,823	61,440	19,155.73	0.07
NEQR (Zhang et al., 2013)	11	47,615	65,288	20,688	8,682.33	0.13
Unary Encoding (Romero et al., 2017)	512	1	1	0	—	—
Angle Encoding (Mitarai et al., 2018)	512	4	2,048	0	—	—
Data Re-uploading (Pérez-Salinas et al., 2020)	1	2,560	2,560	0	183.87	0.01
Patch Embedding (Ours)	10	158	1,548	0	214.88	0.07

Direct mapping methods (Basis, Angle, Unary, Phase Encoding) avoid CNOT gates but require up to 512 qubits, exceeding the capacity of current processors. Amplitude Encoding uses 9 qubits but has a depth of 2,370 with 502 CNOTs. FRQI and NEQR require 61,440 and 20,688 CNOTs, respectively, with depths far exceeding device coherence times.

**Important caveat:** The comparison between Patch Embedding and image-specific methods (FRQI, NEQR) should be interpreted carefully. FRQI and NEQR encode raw pixel values with spatial structure, whereas Patch Embedding encodes a pre-extracted 512-dimensional feature vector. The circuit

cost reductions therefore reflect both the encoding design and the dimensionality reduction performed by Fire512 Head.

Patch Embedding uses 10 qubits with a physical depth of only 158 and 1,548 single-qubit gates, zero CNOTs. Compared with feasible compression methods, it achieves a depth 93% lower than Amplitude Encoding and eliminates all two-qubit gate errors. Its simulation time (214.88 ms) is comparable to Amplitude Encoding (263.02 ms) and two orders of magnitude faster than FRQI (19,155.73 ms). Data Re-uploading achieves the lowest memory (0.01 MB) but incurs a depth of 2,560 on a single qubit, limiting parallelism. Patch Embedding distributes computation across 10 qubits with  $16\times$  lower depth, enabling better utilization of multi-qubit hardware.

#### 6.4. Generalization and Overfitting

Given the noise and limited samples in several datasets, we evaluate the generalization gap  $\mathcal{G} = \mathcal{A}_{\text{train}} - \mathcal{A}_{\text{test}}$ , where  $\mathcal{A}_{\text{train}}$  and  $\mathcal{A}_{\text{test}}$  are the highest training and test accuracies. Figure 6 shows  $\mathcal{G}$  for all models across 13 datasets.

Classical models exhibit generalization gaps of 15% to 49% on challenging datasets such as DeepWeeds, HAM10000, PCAM, RESISC45, and UC Merced. By contrast, FireQuan maintains  $\mathcal{G} < 10\%$  across all datasets. This stability is likely attributable to the compact parameter count of Fire512 Head (which limits overfitting capacity) combined with the implicit regularization effect of the quantum circuit’s constrained Hilbert space.

#### 6.5. Limitations

Several limitations of this work should be noted. First, all experiments are conducted in simulation; validation on physical quantum hardware is necessary to confirm the practical benefits, particularly regarding noise resilience. Second, only accuracy is reported; for imbalanced datasets such as HAM10000, ISIC2019, and PCAM, metrics such as F1-score and AUC-ROC would provide a more complete evaluation. Third, all results use a single random seed (42); reporting mean and standard deviation across multiple seeds would strengthen the statistical rigor. Fourth, the scalability of Patch Embedding to larger qubit counts (e.g., 20, 50 qubits) has not been explored. Finally, the claim that Patch Embedding may help alleviate barren plateaus requires formal gradient variance analysis, which we leave for future work.

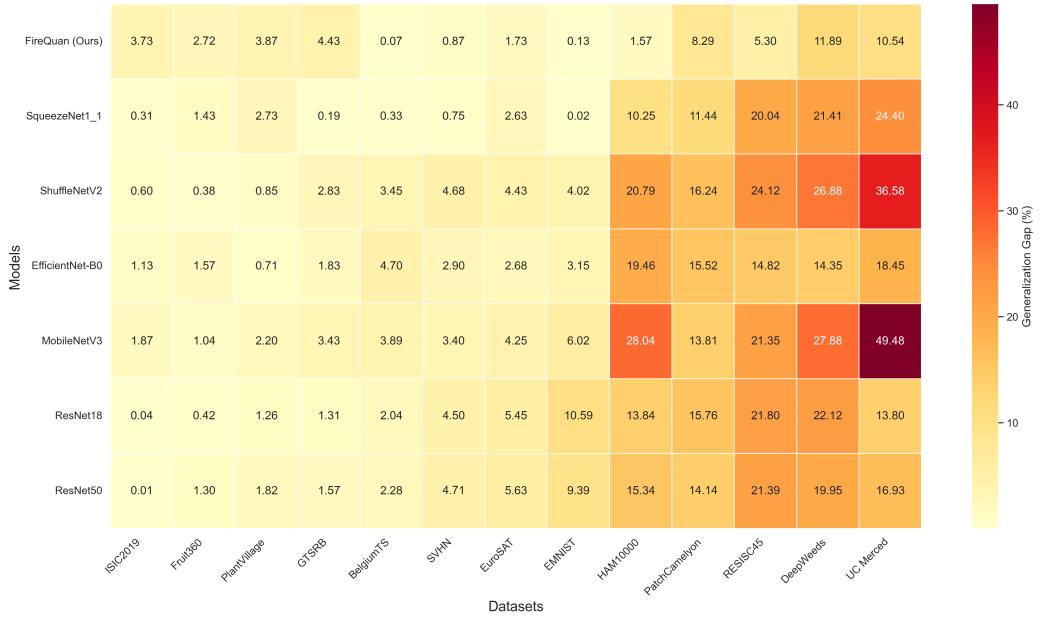


Figure 6: Generalization gap ( $\mathcal{G} = \mathcal{A}_{\text{train}} - \mathcal{A}_{\text{test}}$ ) for all models across 13 datasets (lower is better).

## 7. Conclusion and Future Work

This paper presents FireQuan, a hybrid quantum-classical architecture for multi-domain image classification in the NISQ era. The architecture integrates a compact Fire512 Head (0.26M parameters,  $85\times$  smaller than ResNet50) for feature extraction with a patch-based encoding strategy that loads features into qubits using only single-qubit rotation gates, eliminating CNOT gates during encoding. Experiments across 13 datasets in five domains demonstrate competitive accuracy with classical CNNs, with notable advantages on biomedical datasets (86.70% on PCAM, 76.39% on HAM10000) and remote sensing (95.74% on EuroSAT). The model maintains a generalization gap below 10% across all datasets, highlighting its robustness under noise and class imbalance.

Future work will pursue three directions. First, we plan to validate FireQuan on physical quantum hardware to assess real-world noise effects and circuit fidelity. Second, scaling the quantum component to more qubits and exploring the scalability of Patch Embedding will be investigated. Third, integrating FireQuan into federated learning frameworks for privacy-preserving

distributed training is a longer-term goal. Additionally, expanding the evaluation to include F1-score, AUC-ROC, and multi-seed experiments will strengthen the empirical foundation.

## Abbreviations

Abbreviation	Definition
BelgiumTS	Belgian Traffic Sign Dataset
CNN	Convolutional Neural Network
CNOT	Controlled-NOT Quantum Gate
EMNIST	Extended MNIST Dataset
EuroSAT	European Satellite Image Dataset
FC	Fully Connected (Layer)
FLOPs	Floating-Point Operations
FRQI	Flexible Representation of Quantum Images
GAP	Global Average Pooling
GPU	Graphics Processing Unit
GTSRB	German Traffic Sign Recognition Benchmark
HAM10000	Human Against Machine with 10000 training images
ISIC2019	International Skin Imaging Collaboration 2019 Dataset
NEQR	Novel Enhanced Quantum Representation
NISQ	Noisy Intermediate-Scale Quantum
PCAM	PatchCamelyon Dataset
PQCs	Parameterized Quantum Circuits
QAE	Quantum AutoEncoder
QCNN	Quantum Convolutional Neural Network
$QK$	Quantum Kernel
QML	Quantum Machine Learning
QNN	Quantum Neural Network
RESISC45	Remote Sensing Image Scene Classification (45 classes)
ResNet18	Residual Network – 18 layers
ResNet50	Residual Network – 50 layers
$Rot$	General rotation gate
$R_X$	The single qubit X rotation
$R_Y$	The single qubit Y rotation
$R_Z$	The single qubit Z rotation
SVHN	Street View House Numbers Dataset
UC Merced	UC Merced Land Use Dataset
vRAM	Video Random Access Memory
$N_{\text{classes}}$	Number of dataset classes

Continued on next page

Abbreviation – continued from previous page.

Abbreviation	Definition
$N_q$	Number of Qubits

### **Declaration of generative AI and AI-assisted technologies in the manuscript preparation process**

During the preparation of this study, the authors used Grammarly to improve the language and phrasing of the manuscript, as none of the authors are native English speakers. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

### **CRediT authorship contribution statement**

**Quang Nhan Hoang:** Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Writing – Original Draft, Visualization. **Trung Thanh Pham:** Conceptualization, Formal analysis, Validation, Writing – Review & Editing, Visualization. **Nhut Minh Nguyen:** Validation, Formal analysis, Writing – Review & Editing, Visualization. **Linh Le:** Validation, Formal analysis, Writing – Review & Editing, Supervision, Visualization. **Choong Seon Hong:** Validation, Formal analysis, Resources, Writing – Review & Editing, Supervision, Project administration. **Duc Ngoc Minh Dang:** Conceptualization, Validation, Formal analysis, Resources, Writing – Review & Editing, Supervision, Project administration.

### **Declaration of competing interest**

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Author agreement

All authors have seen and approved the final version of the manuscript being submitted. They warrant that the article is the authors' original work, has not received prior publication, and is not under consideration for publication elsewhere.

## References

- Agarap, A.F., 2018. Deep learning using rectified linear units (relu). ArXiv abs/1803.08375. URL: <https://api.semanticscholar.org/CorpusID:4090379>.
- Alam, M., Ghosh, S., 2022. DeepQMLP: A Scalable Quantum-Classical Hybrid Deep Neural Network Architecture for Classification , in: 2022 35th International Conference on VLSI Design and 2022 21st International Conference on Embedded Systems (VLSID), IEEE Computer Society, Los Alamitos, CA, USA. pp. 275–280. URL: <https://doi.ieeecomputerociety.org/10.1109/VLSID2022.2022.00060>, doi:10.1109/VLSID2022.2022.00060.
- Alda, V., 1980. Remark on two papers concerning axiomatics of quantum mechanics. *Aplikace matematiky* 25, 453–456. URL: <http://eudml.org/doc/15172>.
- Arute, F., Arya, K., Babbush, R., Bacon, D., Bardin, J.C., Barends, R., Biswas, R., Boixo, S., Brandao, F.G., Buell, D.A., et al., 2019. Quantum supremacy using a programmable superconducting processor. *Nature* 574, 505–510.
- Asaoka, H., Kudo, K., 2025. Quantum autoencoders for image classification. *Quantum Machine Intelligence* 7, 71.
- Baek, H., Yun, W.J., Park, S., Kim, J., 2023. Stereoscopic scalable quantum convolutional neural networks. *Neural Networks* 165, 860–867. URL: <https://www.sciencedirect.com/science/article/pii/S0893608023003386>, doi:<https://doi.org/10.1016/j.neunet.2023.06.027>.

- Baral, B., Bhagamiya, B., Majumder, R., Roy, D.D., Roy, T.D., 2025. Adversarial attacks on hybrid classical-quantum deep learning models for histopathological cancer detection. *APL Machine Learning* 3, 036106. URL: <https://doi.org/10.1063/5.0270673>, doi:10.1063/5.0270673, arXiv:<https://pubs.aip.org/aip/aml/article-pdf/doi/10.1063/5.0270673/20611418/036106.pdf>
- Benedetti, M., Lloyd, E., Sack, S., Fiorentini, M., 2019. Erratum: Parameterized quantum circuits as machine learning models (2019 quant. sci. tech. 4 043001). *Quantum Science and Technology* 5, 019601. URL: <https://doi.org/10.1088/2058-9565/ab5944>, doi:10.1088/2058-9565/ab5944.
- Benioff, P., 1980. The computer as a physical system: A microscopic quantum mechanical hamiltonian model of computers as represented by turing machines. *Journal of statistical physics* 22, 563–591.
- Biamonte, J., Wittek, P., Pancotti, N., Rebentrost, P., Wiebe, N., Lloyd, S., 2017. Quantum machine learning. *Nature* 549, 195–202. URL: <http://dx.doi.org/10.1038/nature23474>, doi:10.1038/nature23474.
- Bressan, A., 1978. On wave functions in quantum mechanics. i. *Rendiconti del Seminario Matematico della Università di Padova* 60, 77–98. URL: <http://eudml.org/doc/107706>.
- Ceschini, A., Carbone, A., Sebastianelli, A., Panella, M., Le Saux, B., 2025. On hybrid quanvolutional neural networks optimization. *Quantum Machine Intelligence* 7, 18.
- Chen, F., Zhao, Q., Feng, L., Chen, C., Lin, Y., Lin, J., 2025. Quantum mixed-state self-attention network. *Neural Networks* 185, 107123. URL: <https://www.sciencedirect.com/science/article/pii/S0893608025000024>, doi:<https://doi.org/10.1016/j.neunet.2025.107123>.
- Cheng, G., Han, J., Lu, X., 2017. Remote sensing image scene classification: Benchmark and state of the art. *Proceedings of the IEEE* 105, 1865–1883. doi:10.1109/JPROC.2017.2675998.
- Codella, N.C.F., Gutman, D., Celebi, M.E., Helba, B., Marchetti, M.A., Dusza, S.W., Kalloo, A., Liopyris, K., Mishra, N., Kittler, H., Halpern, A., 2018. Skin lesion analysis toward melanoma detection: A challenge

- at the 2017 international symposium on biomedical imaging (isbi), hosted by the international skin imaging collaboration (isic), in: 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018), pp. 168–172. doi:10.1109/ISBI.2018.8363547.
- Cohen, G., Afshar, S., Tapson, J., van Schaik, A., 2017. Emnist: Extending mnist to handwritten letters, in: 2017 International Joint Conference on Neural Networks (IJCNN), pp. 2921–2926. doi:10.1109/IJCNN.2017.7966217.
- Cong, I., Choi, S., Lukin, M.D., 2019. Quantum convolutional neural networks. *Nature Physics* 15, 1273–1278.
- Coppersmith, D., 2002. An approximate fourier transform useful in quantum factoring. URL: <https://arxiv.org/abs/quant-ph/0201067>, arXiv:quant-ph/0201067.
- da Silva, A.J., Ludermir, T.B., de Oliveira, W.R., 2016. Quantum perceptron over a field and neural network architecture selection in a quantum computer. *Neural Networks* 76, 55–64. URL: <https://www.sciencedirect.com/science/article/pii/S0893608016000034>, doi:<https://doi.org/10.1016/j.neunet.2016.01.002>.
- Daimon, S., Matsushita, Y.i., 2024. Quantum circuit generation for amplitude encoding using a transformer decoder. *Phys. Rev. Appl.* 22, L041001. URL: <https://link.aps.org/doi/10.1103/PhysRevApplied.22.L041001>, doi:10.1103/PhysRevApplied.22.L041001.
- Deng, L., 2012. The mnist database of handwritten digit images for machine learning research [best of the web]. *IEEE Signal Processing Magazine* 29, 141–142. doi:10.1109/MSP.2012.2211477.
- Ganjefar, S., Tofghi, M., Karami, H., 2015. Fuzzy wavelet plus a quantum neural network as a design base for power system stability enhancement. *Neural Networks* 71, 172–181. URL: <https://www.sciencedirect.com/science/article/pii/S0893608015001471>, doi:<https://doi.org/10.1016/j.neunet.2015.07.010>.
- Giuste, F.O., Vizcarra, J.C., 2020. Cifar-10 image classification using feature ensembles. URL: <https://arxiv.org/abs/2002.03846>, arXiv:2002.03846.

- Goyal, N., Uehara, G., Spanias, A., 2025. Quantum-enhanced cancer detection for histopathologic images, in: 2025 IEEE International Conference on Image Processing (ICIP), pp. 1342–1347. doi:[10.1109/ICIP55913.2025.11084595](https://doi.org/10.1109/ICIP55913.2025.11084595).
- Grover, L.K., 1996. A fast quantum mechanical algorithm for database search, in: Proceedings of the twenty-eighth annual ACM symposium on Theory of computing, pp. 212–219.
- Han, J., DiBrita, N.S., Cho, Y., Luo, H., Patel, T., 2025. Enqode: Fast amplitude embedding for quantum machine learning using classical data, in: 2025 62nd ACM/IEEE Design Automation Conference (DAC), pp. 1–7. doi:[10.1109/DAC63849.2025.11132921](https://doi.org/10.1109/DAC63849.2025.11132921).
- Haque, M.E., Paul, M., Ulhaq, A., Debnath, T., 2023. A novel state connection strategy for quantum computing to represent and compress digital images, in: ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 1–5. doi:[10.1109/ICASSP49357.2023.10094832](https://doi.org/10.1109/ICASSP49357.2023.10094832).
- Hasan, M.J., Mahdy, M.R.C., 2025. Bridging classical and quantum machine learning: Knowledge transfer from classical to quantum neural networks using knowledge distillation. URL: <https://arxiv.org/abs/2311.13810>, arXiv:2311.13810.
- Havlíček, V., Córcoles, A.D., Temme, K., Harrow, A.W., Kandala, A., Chow, J.M., Gambetta, J.M., 2019. Supervised learning with quantum-enhanced feature spaces. *Nature* 567, 209–212.
- He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition, in: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 770–778. doi:[10.1109/CVPR.2016.90](https://doi.org/10.1109/CVPR.2016.90).
- He, Z., Wei, J., Chen, C., Huang, Z., Situ, H., Li, L., 2024. Gradient-based optimization for quantum architecture search. *Neural Networks* 179, 106508. URL: <https://www.sciencedirect.com/science/article/pii/S0893608024004325>, doi:<https://doi.org/10.1016/j.neunet.2024.106508>.
- Helber, P., Bischke, B., Dengel, A., Borth, D., 2018. Introducing eurosat: A novel dataset and deep learning benchmark for land use and land cover

classification, in: IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium, pp. 204–207. doi:10.1109/IGARSS.2018.8519248.

Hernández-Pérez, C., Combalia, M., Podlipnik, S., Codella, N.C.F., Rotemberg, V., Halpern, A.C., Reiter, O., Carrera, C., Barreiro, A., Helba, B., Puig, S., Vilaplana, V., Malvehy, J., 2024. Bcn20000: Dermoscopic lesions in the wild. *Scientific Data* 11, 641. URL: <https://doi.org/10.1038/s41597-024-03387-w>, doi:10.1038/s41597-024-03387-w.

Howard, A.G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., Adam, H., 2017. Mobilenets: Efficient convolutional neural networks for mobile vision applications. URL: <https://arxiv.org/abs/1704.04861>, arXiv:1704.04861.

Huang, H.Y., Broughton, M., Mohseni, M., Babbush, R., Boixo, S., Neven, H., McClean, J.R., 2021. Power of data in quantum machine learning. *Nature Communications* 12. URL: <http://dx.doi.org/10.1038/s41467-021-22539-9>, doi:10.1038/s41467-021-22539-9.

Iandola, F.N., Han, S., Moskewicz, M.W., Ashraf, K., Dally, W.J., Keutzer, K., 2016. SqueezeNet: Alexnet-level accuracy with 50x fewer parameters and <math>\sim</math>0.5mb model size. URL: <https://arxiv.org/abs/1602.07360>, arXiv:1602.07360.

Jouzdani, P., Hashim, H.A., Mucciolo, E.R., 2024. Quantum algorithms for state preparation and data classification based on stabilizer codes. *Phys. Rev. A* 109, 022602. URL: <https://link.aps.org/doi/10.1103/PhysRevA.109.022602>, doi:10.1103/PhysRevA.109.022602.

Khan, M.A., Aman, M.N., Sikdar, B., 2024. Beyond bits: A review of quantum embedding techniques for efficient information processing. *IEEE Access* 12, 46118–46137. doi:10.1109/ACCESS.2024.3382150.

Le, P.Q., Dong, F., Hirota, K., 2011. A flexible representation of quantum images for polynomial preparation, image compression, and processing operations. *Quantum Information Processing* 10, 63–84.

Li, Y., Liu, R., Hao, X., Shang, R., Zhao, P., Jiao, L., 2023. Eqnas: Evolutionary quantum neural architecture search for image classification.

- Neural Networks 168, 471–483. URL: <https://www.sciencedirect.com/science/article/pii/S0893608023005348>, doi:<https://doi.org/10.1016/j.neunet.2023.09.040>.
- Liang, Y., Peng, W., Zheng, Z.J., Silvén, O., Zhao, G., 2021. A hybrid quantum-classical neural network with deep residual learning. Neural Networks 143, 133–147. URL: <https://www.sciencedirect.com/science/article/pii/S0893608021002240>, doi:<https://doi.org/10.1016/j.neunet.2021.05.028>.
- Lin, M., Chen, Q., Yan, S., 2014. Network in network. URL: <https://arxiv.org/abs/1312.4400>, arXiv:1312.4400.
- Liu, F., Bian, K., Meng, F., Zhang, W., Dahlsten, O., 2024. Information compression via hidden subgroup quantum autoencoders. npj Quantum Information 10, 74.
- Liu, J., Lim, K.H., Wood, K.L., Huang, W., Guo, C., Huang, H.L., 2021. Hybrid quantum-classical convolutional neural networks. Science China Physics, Mechanics & Astronomy 64, 290311.
- Long, C., Huang, M., Ye, X., Futamura, Y., Sakurai, T., 2025. Hybrid quantum-classical-quantum convolutional neural networks. Scientific Reports 15, 31780.
- Manukian, H., Traversa, F.L., Di Ventra, M., 2019. Accelerating deep learning with memcomputing. Neural Networks 110, 1–7. URL: <https://www.sciencedirect.com/science/article/pii/S0893608018302971>, doi:<https://doi.org/10.1016/j.neunet.2018.10.012>.
- Mao, A., Mohri, M., Zhong, Y., 2023. Cross-entropy loss functions: theoretical analysis and applications, in: Proceedings of the 40th International Conference on Machine Learning, JMLR.org.
- Marcianò, A., Chen, D., Fabrocini, F., Fields, C., Greco, E., Gresnigt, N., Jinklub, K., Lulli, M., Terzidis, K., Zappala, E., 2022. Quantum neural networks and topological quantum field theories. Neural Networks 153, 164–178. URL: <https://www.sciencedirect.com/science/article/pii/S0893608022002027>, doi:<https://doi.org/10.1016/j.neunet.2022.05.028>.

- McDonough, B., Mari, A., Shammah, N., Stemen, N.T., Wahl, M., Zeng, W.J., Orth, P.P., 2022. Automated quantum error mitigation based on probabilistic error reduction, in: 2022 IEEE/ACM Third International Workshop on Quantum Computing Software (QCS), pp. 83–93. doi:10.1109/QCS56647.2022.00015.
- Mitarai, K., Negoro, M., Kitagawa, M., Fujii, K., 2018. Quantum circuit learning. *Phys. Rev. A* 98, 032309. URL: <https://link.aps.org/doi/10.1103/PhysRevA.98.032309>, doi:10.1103/PhysRevA.98.032309.
- Mitsuda, N., Ichimura, T., Nakaji, K., Suzuki, Y., Tanaka, T., Raymond, R., Tezuka, H., Onodera, T., Yamamoto, N., 2024. Approximate complex amplitude encoding algorithm and its application to data classification problems. *Phys. Rev. A* 109, 052423. URL: <https://link.aps.org/doi/10.1103/PhysRevA.109.052423>, doi:10.1103/PhysRevA.109.052423.
- Morandi, Omar, B.L.F.G., 2011. Perturbation theory in terms of a generalized phase-space quantization procedure. *Bollettino dell’Unione Matematica Italiana* 4, 1–18. URL: <http://eudml.org/doc/290734>.
- Netzer, Y., Wang, T., Coates, A., Bissacco, A., Wu, B., Ng, A., 2011. Reading digits in natural images with unsupervised feature learning. URL: <https://api.semanticscholar.org/CorpusID:16852518>.
- Neumann, M., Pinto, A.S., Zhai, X., Housby, N., 2019. In-domain representation learning for remote sensing. URL: <https://arxiv.org/abs/1911.06721>, arXiv:1911.06721.
- O’Shea, K., Nash, R., 2015. An introduction to convolutional neural networks. URL: <https://arxiv.org/abs/1511.08458>, arXiv:1511.08458.
- Oukaira, A., 2025. Quantum hardware devices (qhds): Opportunities and challenges. *IEEE Access* 13, 98229–98241. doi:10.1109/ACCESS.2025.3576216.
- Pande, M.B., 2024. A comprehensive review of data encoding techniques for quantum machine learning problems, in: 2024 Second International Conference on Emerging Trends in Information Technology and Engineering (ICETITE), pp. 1–7. doi:10.1109/ic-ETITE58242.2024.10493306.

- Pandey, P., Mandal, S., 2025. A hybrid quantum–classical convolutional neural network with a quantum attention mechanism for skin cancer. *Scientific Reports* .
- Parigi, M., Khosrojerdi, M., Caruso, F., Banchi, L., 2025. Supervised quantum image processing. URL: <https://arxiv.org/abs/2507.22039>, arXiv:2507.22039.
- Park, S., Lee, H., Son, S.B., Jung, S., Kim, J., 2025. Quantum federated learning with pole-angle quantum local training and trainable measurement. *Neural Networks* 187, 107301. URL: <https://www.sciencedirect.com/science/article/pii/S0893608025001807>, doi:<https://doi.org/10.1016/j.neunet.2025.107301>.
- Pérez-Salinas, A., Cervera-Lierta, A., Gil-Fuster, E., Latorre, J.I., 2020. Data re-uploading for a universal quantum classifier. *Quantum* 4, 226.
- Preskill, J., 2018. Quantum Computing in the NISQ era and beyond. *Quantum* 2, 79. URL: <https://doi.org/10.22331/q-2018-08-06-79>, doi:10.22331/q-2018-08-06-79.
- Romero, J., Olson, J.P., Aspuru-Guzik, A., 2017. Quantum autoencoders for efficient compression of quantum data. *Quantum Science and Technology* 2, 045001.
- Saravanan, K., N, Y., 2025. Agricultural land use and land cover classification using self-attention deep learning techniques, in: 2025 International Conference on Next Generation Computing Systems (ICNGCS), pp. 1–7. doi:10.1109/ICNGCS64900.2025.11183072.
- Schuld, M., Bocharov, A., Svore, K.M., Wiebe, N., 2020. Circuit-centric quantum classifiers. *Phys. Rev. A* 101, 032308. URL: <https://link.aps.org/doi/10.1103/PhysRevA.101.032308>, doi:10.1103/PhysRevA.101.032308.
- Schuld, M., Killoran, N., 2022. Is quantum advantage the right goal for quantum machine learning? *PRX Quantum* 3, 030101. URL: <https://link.aps.org/doi/10.1103/PRXQuantum.3.030101>.

- Senokosov, A., Sedykh, A., Sagingalieva, A., Kyriacou, B., Melnikov, A., 2024. Quantum machine learning for image classification. *Machine Learning: Science and Technology* 5, 015040. URL: <https://doi.org/10.1088/2632-2153/ad2aef>, doi:10.1088/2632-2153/ad2aef.
- Sharma, S., N, R., 2024. Survey of encoding techniques for quantum machine learning. *Cybernetics and Physics* 13, 152–160. doi:10.35470/2226-4116-2024-13-2-152-160.
- Simonyan, K., Zisserman, A., 2015. Very deep convolutional networks for large-scale image recognition, in: International Conference on Learning Representations.
- Stallkamp, J., Schlipsing, M., Salmen, J., Igel, C., 2012. Man vs. computer: Benchmarking machine learning algorithms for traffic sign recognition. *Neural Networks* , –URL: <http://www.sciencedirect.com/science/article/pii/S0893608012000457>, doi:10.1016/j.neunet.2012.02.016.
- Tan, M., Le, Q.V., 2020. Efficientnet: Rethinking model scaling for convolutional neural networks. URL: <https://arxiv.org/abs/1905.11946>, arXiv:1905.11946.
- Timofte, R., Zimmermann, K., Van Gool, L., 2009. Multi-view traffic sign detection, recognition, and 3d localisation, in: 2009 Workshop on Applications of Computer Vision (WACV), pp. 1–8. doi:10.1109/WACV.2009.5403121.
- Tschandl, P., 2018. The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. URL: <https://doi.org/10.7910/DVN/DBW86T>, doi:10.7910/DVN/DBW86T.
- V, A.C., Joseph, E.R., Alasmari, S., 2025. A hybrid quantum-classical approach for multi-class skin disease classification using a 4-qubit model. *IEEE Access* 13, 114460–114470. doi:10.1109/ACCESS.2025.3581030.
- Veeling, B.S., Linmans, J., Winkens, J., Cohen, T., Welling, M., 2018. Rotation equivariant cnns for digital pathology, in: Frangi, A.F., Schnabel, J.A., Davatzikos, C., Alberola-López, C., Fichtinger, G. (Eds.), *Medical Image Computing and Computer Assisted Intervention – MICCAI 2018*, Springer International Publishing, Cham. pp. 210–218.

- Wang, X., Han, Y., Leung, V.C.M., Niyato, D., Yan, X., Chen, X., 2020. Convergence of edge computing and deep learning: A comprehensive survey. *IEEE Communications Surveys & Tutorials* 22, 869–904. doi:10.1109/COMST.2020.2970550.
- Wang, X., Wang, Y., Qi, B., Wu, R., 2025. Limitations of amplitude encoding on quantum classification. URL: <https://arxiv.org/abs/2503.01545>, arXiv:2503.01545.
- Wierichs, D., Izaac, J., Wang, C., Lin, C.Y.Y., 2022. General parameter-shift rules for quantum gradients. *Quantum* 6, 677.
- Wu, T., Bentellis, A., Sakhnenko, A., Lorenz, J.M., 2025. Generalization bounds in hybrid quantum-classical machine learning models. URL: <https://arxiv.org/abs/2504.08456>, arXiv:2504.08456.
- Xiao, H., Rasul, K., Vollgraf, R., 2017. Fashion-mnist: a novel image dataset for benchmarking machine learning algorithms. URL: <https://arxiv.org/abs/1708.07747>, arXiv:1708.07747.
- Xie, X., Gao, Y., Zhang, Y., 2025. An improved artificial protozoa optimizer for cnn architecture optimization. *Neural Networks* 187, 107368. URL: <https://www.sciencedirect.com/science/article/pii/S0893608025002473>, doi:<https://doi.org/10.1016/j.neunet.2025.107368>.
- Zaidenberg, D.A., Sebastianelli, A., Spiller, D., Le Saux, B., Ullo, S.L., 2021. Advantages and bottlenecks of quantum machine learning for remote sensing, in: 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS, pp. 5680–5683. doi:10.1109/IGARSS47720.2021.9553133.
- Zhang, R., Wang, J., Jiang, N., Li, H., Wang, Z., 2022. Quantum support vector machine based on regularized newton method. *Neural Networks* 151, 376–384. URL: <https://www.sciencedirect.com/science/article/pii/S0893608022001216>, doi:<https://doi.org/10.1016/j.neunet.2022.03.043>.
- Zhang, W., Deng, L., Zhang, L., Wu, D., 2023. A survey on negative transfer. *IEEE/CAA Journal of Automatica Sinica* 10, 305–329. doi:10.1109/JAS.2022.106004.

- Zhang, X., Zhou, X., Lin, M., Sun, J., 2018. Shufflenet: An extremely efficient convolutional neural network for mobile devices, in: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 6848–6856. doi:[10.1109/CVPR.2018.00716](https://doi.org/10.1109/CVPR.2018.00716).
- Zhang, Y., Lu, K., Gao, Y., Wang, M., 2013. Neqr: A novel enhanced quantum representation of digital images. *Quantum Information Processing* 12. doi:[10.1007/s11128-013-0567-z](https://doi.org/10.1007/s11128-013-0567-z).
- Zheng, Y.B., Zhao, X.L., Li, H.C., Li, C., Huang, T.Z., Zhao, Q., 2025. Tensor network decomposition for data recovery: Recent advancements and future prospects. *Neural Networks* 191, 107808. URL: <https://www.sciencedirect.com/science/article/pii/S0893608025006884>, doi:<https://doi.org/10.1016/j.neunet.2025.107808>.
- Zhou, K., Liu, Z., Qiao, Y., Xiang, T., Loy, C.C., 2023. Domain generalization: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 45, 4396–4415. doi:[10.1109/TPAMI.2022.3195549](https://doi.org/10.1109/TPAMI.2022.3195549).