Trinity College Dublin
Coláiste na Tríonóide, Baile Átha Cliath
The University of Dublin

School of Engineering

# Building a Hybrid Quantum-Classical Vision Transformer for Classification

Advik Bahadur

**Algorithms for Quantum Computing II**

December 3, 2025

MAI Electronic and Computer Engineering

# 1 Motivation

The intersection of quantum computing and deep learning is an exciting and rapidly developing area of research. Quantum algorithms introduce novel computational paradigms that could enhance machine learning models by leveraging quantum phenomena such as superposition and entanglement to explore high-dimensional feature spaces more efficiently. These properties allow quantum computers to identify patterns and correlations that are intractable for classical architectures. In theory, we may even achieve exponential speed-ups for higher-order exploration, although this remains unproven given current hardware limitations.

Even with the constraints of today's noisy intermediate-scale quantum (NISQ) devices—limited qubit counts for larger models, noisy outputs, and the need for longer coherence times for fault-tolerant computation—we can still realise meaningful benefits through hybrid machine learning models. These include improved model capabilities, better generalisation, and enhanced training efficiency.

In the age of Artificial Intelligence, one of the key drivers of recent breakthroughs has been the attention mechanism in [1]. These models power large language models, image and video generation systems, and a wide range of computer-vision tasks that shape the modern world. Given the substantial gains already achieved using purely classical algorithms and hardware, this project aims to develop a hybrid quantum–classical transformer [2] architecture to evaluate the potential advantages of incorporating quantum computation into vision classification tasks.

Vision Transformers (ViTs) represent a modern architecture where the self-attention mechanism—-particularly the Query-Key-Value (QKV) projections—is computationally expensive. As gathered by Zhang et al. [2] that hybrid quantum-classical transformers can be effective when quantum components are strategically placed to handle specific bottlenecks while classical layers manage the bulk of the computation.

# 2 Project Objectives

**Primary Objective:** Develop a working hybrid quantum-classical Vision transformer that can be used for classification for various datasets of increasing complexity, using selective quantum processing of attention mechanisms and compare its performance against a classical baseline model of equivalent capacity.

**Secondary Objectives:**

1. Implement clear architectural separation between quantum and classical components.
2. Establish reproducible evaluation metrics (accuracy, training time, inference latency).
3. Document quantum circuit design, noise characteristics, and resource utilisation.
4. Provide a foundation for future extensions to additional quantum operations.

# 3 Proposed Approach

## 3.1 Feasibility and Challenges

- **Circuit Depth vs. Noise:** Deep circuits accumulate errors exponentially on NISQ hardware [3]. **Potential Mitigation:** *Limit to 5–8 gate layers per QKV circuit.*
- **Barren Plateaus:** Gradients vanish in deep/wide quantum circuits, halting training [3]. **Potential Mitigation:** *Use shallow circuits; careful parameter initialization; monitor gradient*

*norms.*

- **Qubit Constraints:** IBM free tier limits practical circuits to $\leq$8 qubits. **Potential Mitigation:** *Design per-head circuits within this budget; simulate locally for rapid iteration.*
- **Quantum-Classical Gradient Flow:** Parameter shift rule adds overhead; noisy gradients destabilize classical optimizers [3]. **Potential Mitigation:** *Use gradient-free optimizers (COBYLA, SPSA) as fallback; increase shot counts to 1024+.*
- **Encoding Overhead:** Repeated quantum-classical conversions slow training [3]. *Apply quantum processing to one transformer block only.*

## 3.2 Realisitic Assessment

This is a proof-of-concept project, not a production system. The goal is not to beat classical models, but to demonstrate:

- Successful integration of quantum and classical components
- Measurable performance on a real task
- Develop and test a small-scale transformer architecture that can be scaled up with improving quantum hardware.

## 3.3 Minimum Viable Product

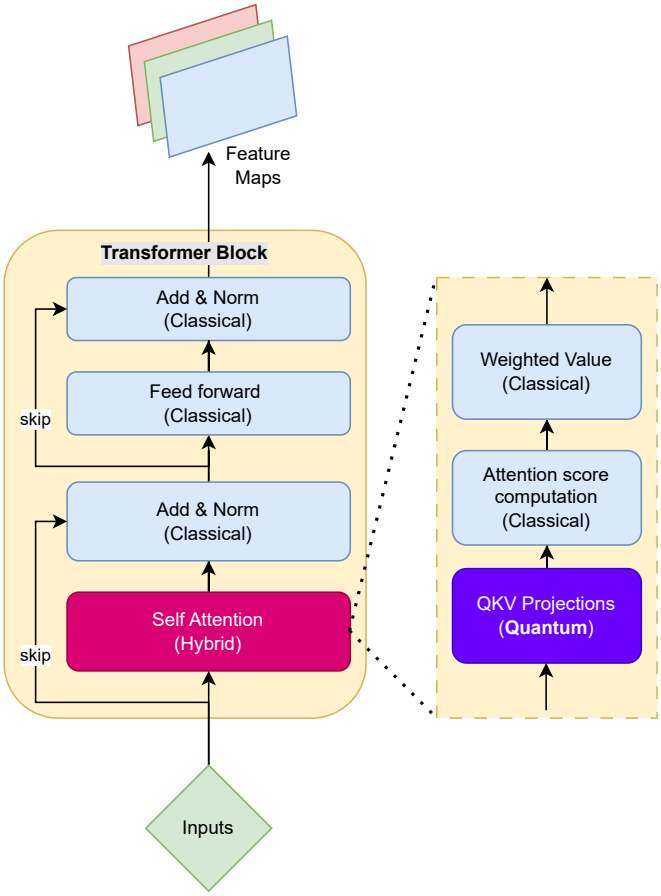


Figure 1: Proposed transformer model showcasing the ViT transformer and the hybrid split.

Our hybrid architecture (Figure 1) strategically limits quantum processing to the Query-Key-Value (QKV) projections within the multi-head attention mechanism. This design maximises potential quantum advantage while minimising resource consumption: QKV transformations are computationally expensive and critical to transformer performance, making them ideal candidates for quantum enhancement [3]. All other operations—patch embedding, positional encoding, feed-forward networks, and classification heads—remain classical, ensuring stable training and manageable overhead.

**Dataset Strategy:** We plan to adopt a phased validation approach. Initial development would start off with a minimal dataset such as the MNIST digits [4] (28×28, 10 classes) as a well-understood baseline for rapid prototyping and debugging. Once the hybrid pipeline is functional, we can look to scale to more complex datasets including the Architectural Heritage Elements (AHE) dataset [5] a 10,235-image collection of heritage architectural features (64×64 resolution) across 10 categories, including altars, columns, domes, gargoyles, and stained glass. We can limit the computational expense on the quantum machines by limiting the model to select 3–4 discriminative classes, providing a realistic benchmark for quantum-enhanced vision transformers.

### 3.4 Advanced Scope (Stretch Goals)

If MVP succeeds, possible extensions include:

1. Multi-Head Quantum Attention: Extend quantum QKV to multiple attention heads in parallel
2. Quantum Error Mitigation: Implement zero-noise extrapolation or readout error mitigation
3. Dynamic Circuit Depth: Adaptively adjust quantum circuit depth based on noise measurements per qubit
4. Additional Classes: Scale to 5–6 classes; increase model capacity
5. Hybrid Feed-Forward: Explore quantum processing in feed-forward layers (separate MVP ifattempted)

## 4 Conclusion

This project presents a pragmatic approach to quantum machine learning: realistic scope, clear comparison, and documented trade-offs. By implementing a hybrid quantum-classical VisionTransformer on a concrete task, we aim to advance understanding of when and how quantum processing can complement classical ML in near-term settings. The MVP focuses on reproducibility and measurable comparison; advanced extensions are deferred to post-MVP phases. Success is not defeating classical methods, but demonstrating functional integration and providing a platform for future exploration.

# Bibliography


[1] A. Vaswani et al., *Attention Is All You Need*, arXiv:1706.03762 [cs], Aug. 2023. DOI: 10.48550/arXiv.1706.03762. Accessed: Dec. 3, 2025. [Online]. Available: http://arxiv.org/abs/1706.03762.

[2] H. Zhang, Q. Zhao, M. Zhou, and L. Feng, *HQViT: Hybrid Quantum Vision Transformer for Image Classification*, arXiv:2504.02730 [cs] version: 1, Apr. 2025. DOI: 10.48550/arXiv.2504.02730. Accessed: Nov. 21, 2025. [Online]. Available: http://arxiv.org/abs/2504.02730.

[3] H. Zhang et al., *A Survey of Quantum Transformers: Architectures, Challenges and Outlooks*, arXiv:2504.03192 [quant-ph], Nov. 2025. DOI: 10.48550/arXiv.2504.03192. Accessed: Dec. 3, 2025. [Online]. Available: http://arxiv.org/abs/2504.03192.

[4] *MNIST Dataset*, en. Accessed: Dec. 3, 2025. [Online]. Available: https://www.kaggle.com/datasets/hojjatk/mnist-dataset.

[5] *Architectural Heritage Elements Image64 Dataset*, en. Accessed: Dec. 3, 2025. [Online]. Available: https://www.kaggle.com/datasets/ikobzev/architectural-heritage-elements-image64-dataset.