Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Comparative Analysis of Classical and Quantum Runge-Kutta Methods in Classification Tasks

Janakkumar Patel

janakkp220@gmail.com

Campbellsville University https://orcid.org/0009-0002-3527-9331

Research Article

Keywords: Quantum machine learning, Runge-Kutta methods, Vision Transformer, ordinary differential equations, variational quantum circuits, generative adversarial networks, parameter efficiency, NISQ computing

Posted Date: April 2nd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9286122/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: The authors declare no competing interests.

Abstract

Deep learning models have driven remarkable progress across image classification, natural language processing, and generative modelling. Yet their reliance on millions of trainable parameters brings substantial computational and environmental costs. This paper investigates whether Ordinary Differential Equation inspired Transformer architectures, built on first through fourth order Runge-Kutta numerical solvers, can reduce parameter counts while preserving classification quality. I implement both classical and quantum variants of these ODE-Transformer blocks inside Vision Transformer encoders and evaluate them on MNIST digit classification, CIFAR-10 object classification, and IMDb sentiment analysis. Classical fourth order Runge-Kutta blocks achieve 95 percent accuracy on MNIST and 42 percent on CIFAR-10, outperforming lower order methods by a small margin while training time grows with method complexity. An optimized RK4 variant that removes intermediate stages converges faster and uses fewer parameters without sacrificing performance. Quantum implementations match 91 percent accuracy on MNIST with roughly 40 percent fewer trainable parameters, though they require significantly more wall-clock training time due to circuit simulation overhead. On IMDb, the quantum enhanced model matches the classical 85 percent accuracy while halving the parameter count. Finally, a Quantum Runge-Kutta Transformer GAN called QRKT-GAN achieves an Inception Score of 74.89 and a Frechet Inception Distance of 6.78 on CIFAR-10, substantially outperforming the classical TransGAN baseline. These results suggest that quantum ODE-Transformer hybrids offer a viable path toward more parameter efficient deep learning, with practical advantages likely to grow as fault tolerant quantum hardware matures.

I. Introduction

Artificial intelligence models built on deep neural networks have transformed computer vision, natural language understanding, and generative modelling over the past decade. State of the art architectures such as Vision Transformers and large language models routinely contain hundreds of millions to billions of trainable parameters, demanding enormous computational resources and drawing significant electrical power during training and inference. This resource burden limits accessibility and raises environmental concerns that the research community increasingly acknowledges.

Researchers have pursued many strategies to reduce parameter counts while maintaining acceptable performance. Classical approaches include structured sparsity, knowledge distillation, product of sum matrix decompositions, and kernel sharing between convolutional layers. Transfer learning offers another path by reusing pretrained weights, though it requires suitable source models and domain compatibility. Despite these techniques, the fundamental tension between model capacity and resource consumption persists.

Quantum computing offers a fundamentally different avenue. Variational quantum circuits encode information in qubits that exploit superposition and entanglement, enabling compact representations that classical linear layers cannot achieve. Recent work on quantum Vision Transformers has shown that

replacing classical projection layers with parameterized quantum circuits can cut the parameter budget substantially without degrading accuracy on benchmark tasks.

Separately, the neural ODE perspective interprets Transformer residual blocks as discretized solutions to ordinary differential equations. Higher order Runge-Kutta solvers applied in this framework yield lower truncation error and thus more parameter efficient blocks. Combining the neural ODE viewpoint with variational quantum circuits creates an opportunity to compress Transformer architectures from two complementary directions simultaneously.

In this paper I investigate that combined approach. I implement first through fourth order Runge-Kutta ODE blocks in both classical and quantum settings, embed them in Vision Transformer encoders, and benchmark them on MNIST, CIFAR-10, and IMDb. I also propose QRKT-GAN, a quantum Runge-Kutta Transformer GAN, and compare it against the classical TransGAN on CIFAR-10 image generation. My goal is to quantify the parameter savings, accuracy trade-offs, and current hardware limitations of quantum ODE-Transformers across a range of tasks.

II. Related Work

Neural ODEs were introduced by Chen et al., who showed that residual networks can be viewed as Euler discretizations of continuous dynamical systems. Li et al. extended this idea to Transformers by interpreting each residual block as a higher order Runge-Kutta step, demonstrating that RK2 and RK4 blocks reduce truncation error and improve parameter efficiency compared to the standard residual formulation.

On the quantum side, Comajoan Cara et al. proposed quantum Vision Transformers that replace classical multi-head attention projections and MLP layers with variational quantum circuits, achieving competitive accuracy on particle physics classification tasks with significantly fewer parameters. Benedetti et al. provided a general framework for parameterized quantum circuits as machine learning models, and Cerezo et al. surveyed the broader landscape of variational quantum algorithms.

Fan et al. recently introduced a quantum inspired neural network using Runge-Kutta methods, though their approach operates entirely on classical hardware and does not involve actual quantum circuits. My work differs by implementing genuine variational quantum circuits within the ODE-Transformer framework and evaluating the resulting hybrid models on standard classification and generation benchmarks.

III. Background

A. ODE Interpretation of Transformer Layers

A standard Transformer layer applies multi-head attention followed by a feed-forward network, each wrapped in a residual connection. Denoting the combined attention and feed-forward function as F, the

residual update y(t+1) = y(t) + F(y(t)) is equivalent to a single Euler step. Higher order Runge-Kutta methods improve upon Euler by computing intermediate evaluations of F. The second order method averages two evaluations, the third order uses three weighted evaluations, and the classical fourth order method combines four evaluations with weights 1/6, 1/3, 1/3, 1/6. Because all evaluations share the same function F and input context, the Transformer reuses its attention and MLP weights across stages, increasing effective depth without proportional parameter growth.

B. Variational Quantum Circuits

A variational quantum circuit prepares a parameterized quantum state by applying a sequence of rotation and entangling gates to an initial state of qubits. Feature values from classical data are encoded through rotation angles, and trainable parameters are optimized alongside classical network weights using gradient based methods. For a hidden dimension of size n, a fully connected classical layer requires n squared plus n parameters, whereas a variational quantum circuit with two rotation layers requires only 2n trainable parameters. This quadratic to linear reduction is the primary motivation for quantum integration.

IV. Proposed Method

A. ODE-QViT Architecture

I construct hybrid Quantum Vision Transformers by replacing the classical linear projections in multi-head attention and the MLP sublayers with variational quantum circuits. Each quantum circuit encodes patch embeddings through RY rotations, applies two layers of trainable RX, RY, and RZ rotations interleaved with CNOT entangling gates and Hadamard gates, and measures expectation values to produce the output. The additional RY, RZ, and Hadamard gates compared to the baseline quantum Transformer design improve feature extraction by enabling richer traversal of the qubit state space.

These quantum attention heads and quantum MLPs are then embedded within Runge-Kutta ODE blocks of orders one through four. I also implement an optimized RK4 variant that sets all weighting coefficients to unity and removes the second and third intermediate stages, retaining only the first and fourth evaluations. This modification is not theoretically grounded in standard Runge-Kutta theory but reduces vanishing gradient effects in deeper models and empirically accelerates convergence.

B. QRKT-GAN

For image generation I propose QRKT-GAN, a generative adversarial network in which both the generator and discriminator are quantum Vision Transformers. The generator incorporates an optimized RK4 ODE block and progressively upscales from an 8 by 8 latent grid to 32 by 32 output images. The discriminator uses quantum Transformer encoders without ODE blocks due to memory constraints during training. Training uses the Wasserstein loss with gradient penalty, the AdamW optimizer, and data augmentation on CIFAR-10.

V. Experiments

A. Setup

All models use AdamW optimization with gradient clipping at norm 1 and a cosine learning rate schedule with linear warmup over 5000 steps from 0 to 1e-3. A batch size of 65 is used for classification tasks. Classical components are implemented in JAX and Flax. Quantum circuits are simulated using TensorCircuit on classical hardware. Hyperparameters are selected via random search on the classical baseline and applied unchanged to quantum variants.

B. MNIST Results

On MNIST with patch size 14, hidden size 6, three ODE blocks, and two attention heads, the classical RK4 model achieves 95 percent accuracy and 95 percent F1 with 5971 parameters. The quantum QRK4 reaches 91 percent accuracy with only 3520 parameters and 357 qubits, a 41 percent parameter reduction. The optimized quantum QRK4 variant converges four epochs earlier while maintaining the same accuracy. Training time roughly doubles for quantum models due to circuit simulation overhead.

C. CIFAR-10 Results

On CIFAR-10 with patch size 16, hidden size 12, one ODE block, and six attention heads, the classical optimized RK4 achieves 42 percent accuracy with 33634 parameters. The quantum QRK4 enhanced model reaches 34 percent accuracy with 20590 parameters and 390 qubits. The accuracy gap is larger here, reflecting the increased difficulty of color image classification with constrained quantum circuit depth on current simulators.

D. IMDb Results

On IMDb sentiment analysis with a maximum sequence length of 512, the classical optimized RK4 achieves 85 percent accuracy with 499316 parameters. The quantum enhanced model matches this 85 percent accuracy with only 243896 parameters and 141 qubits, representing a 51 percent parameter reduction with no loss in classification performance. This result demonstrates that quantum ODE-Transformers are particularly effective for sequence tasks where the input dimensionality is high.

E. QRKT-GAN Generation

On CIFAR-10 image generation after 40 training epochs, QRKT-GAN achieves an Inception Score of 74.89 and a Frechet Inception Distance of 6.78. The classical TransGAN baseline achieves an IS of 52.31 and FID of 46.97 under identical training conditions. QRKT-GAN thus produces substantially higher quality synthetic images, though it requires significantly more training time due to quantum circuit simulation on an NVIDIA A100 GPU cluster.

VI. Discussion

Three findings stand out across the experiments. First, quantum ODE-Transformer blocks consistently reduce parameter counts by 40 to 51 percent compared to classical counterparts, confirming the theoretical quadratic to linear scaling advantage of variational quantum circuits. Second, accuracy is preserved or only modestly reduced on MNIST and IMDb, though a meaningful gap appears on CIFAR-10 where the task complexity exceeds what current shallow quantum circuits can efficiently capture. Third, training time increases substantially for quantum models, as each forward pass requires full circuit simulation. This overhead will diminish as quantum hardware improves and native execution replaces simulation.

The strong QRKT-GAN results are encouraging but should be interpreted cautiously. The 40 epoch training limit, imposed by computational constraints, favors architectures that converge quickly. Longer training runs may narrow or reverse the gap between quantum and classical GANs. Nevertheless, the combination of ODE-based blocks with quantum circuits appears to provide a genuine advantage for generative modelling in the current resource constrained regime.

VII. Conclusion and Future Work

This paper demonstrates that integrating Runge-Kutta ODE solvers with variational quantum circuits in Transformer architectures yields meaningful parameter reductions across classification and generation tasks. Quantum implementations achieve competitive accuracy with 40 to 51 percent fewer parameters, and the proposed QRKT-GAN outperforms its classical baseline on image quality metrics. Current limitations center on simulation overhead and shallow circuit depth, both of which are expected to improve with advances in fault tolerant quantum hardware.

Future work should explore deeper variational circuits enabled by error corrected qubits, apply the framework to larger scale datasets and architectures, and benchmark on actual quantum processors as they become available. Testing across diverse quantum hardware platforms including trapped ion, superconducting, and photonic systems will provide a more complete understanding of the practical advantages and constraints of quantum ODE-Transformers.

References

1. Zhong YD, Zhang T, Chakraborty A, Dey B (2022) 'A neural ODE interpretation of transformer layers,' arXiv:2212.06011
2. Li B et al (2022) 'ODE transformer: An ordinary differential equation-inspired model for sequence generation,' arXiv:2203.09176
3. Comajoan M, Cara et al (2024) 'Quantum vision transformers for quark-gluon classification,' Axioms, vol. 13, no. 5, p. 323
4. Benedetti M, Lloyd E, Sack S, Fiorentini M (2019) 'Parameterized quantum circuits as machine learning models,' Quantum Science and Technology, vol. 4, no. 4

5. Cerezo M et al (2021) 'Variational quantum algorithms,' Nature Reviews Physics, vol. 3, no. 9, pp. 625–644
6. Fan Z, Zhang J, Zhang P, Lin Q, Gao H (2024) 'Quantum-inspired neural network with Runge-Kutta method,' Proc. AAAI, vol. 38, pp. 17977–17984
7. Vaswani A et al (2017) 'Attention is all you need,' Advances in NeurIPS, vol. 30
8. Dosovitskiy A et al (2020) 'An image is worth 16x16 words: Transformers for image recognition at scale,' arXiv:2010.11929
9. Jiang Y, Chang S, Wang Z (2021) 'TransGAN: Two pure transformers can make one strong GAN,' Advances in NeurIPS, vol. 34, pp. 14745–14758
10. Zhang SX et al (2023) 'TensorCircuit: A quantum software framework for the NISQ era,' Quantum, vol. 7, p. 912
11. Abbas A et al (2021) The power of quantum neural networks. Nat Comput Sci 1(6):403–409
12. Chen RTQ, Rubanova Y, Bettencourt J (2018) and D. Duvenaud, 'Neural ordinary differential equations,' Advances in NeurIPS, vol. 31
13. J. C. Butcher, 'A history of Runge-Kutta methods,' Applied Numerical Mathematics, vol. 20, no. 3, pp. 247–260, (1996)
14. Goodfellow IJ et al (2014) 'Generative adversarial nets,' Advances in NeurIPS, pp. 2672–2680
15. Nielsen MA, Chuang IL (2001) Quantum Computation and Quantum Information. Cambridge University Press
16. Wang Y, Liu J (2024) 'A comprehensive review of quantum machine learning: From NISQ to fault tolerance,' arXiv:2401.11351