

Quantum Natural Gradient vs. Adam Optimizer in Variational Quantum Classifiers: Crossover Analysis and Information Acquisition Efficiency

Desi Amirullah¹, Lipantri Mashur Gultom^{*2}

^{1,2}Politeknik Negeri Bengkalis, Bengkalis, Riau, Indonesia

¹andes@polbeng.ac.id, ²lipantri@polbeng.ac.id

Received: June 12, 2026 | Revised: July 09, 2026 | Accepted: July 22, 2026

Abstract

Variational Quantum Circuits (VQCs) represent a central paradigm in near-term quantum machine learning, yet the comparative optimisation dynamics of quantum-aware and classical optimisers remain insufficiently characterised in realistic multi-class settings. We present a systematic empirical study comparing the Quantum Natural Gradient (QNG) optimizer against Adam within a VQC trained for ten-class digit recognition, employing eight qubits, three variational layers with RY-RZ-RX gate sequences, circular CX entanglement, and data re-uploading—yielding 72 trainable quantum parameters augmented by a classical linear readout head. A diagonal Quantum Fisher Information Matrix (QFIM) estimated via the parameter-shift fidelity metric underpins QNG, with a lazy update scheme (every five gradient steps) to equalise computational cost with Adam. Over 1,000 training epochs, Adam achieves 95% test accuracy while QNG achieves 92%, with QNG demonstrating markedly superior early convergence. Crossover analysis across four encoding-overlap bins confirms that QNG outperforms Adam exclusively in the low-overlap regime ($c_i < 0.54$), consistent with the theoretical predictions of Kimura and Mitarai [10]. Information acquisition efficiency measurements reveal qualitatively opposite scaling behaviours: Adam's Fisher-empirical g_i scales as $c_i^{-2.19}$, whereas QNG's QFIM-fidelity g_i is near scale-invariant ($c_i^{0.02}$). These findings provide actionable optimizer selection criteria for practical VQC deployments and offer the first empirical validation of the Kimura-Mitarai efficiency framework beyond the quantum phase estimation setting.

Keywords: Variational Quantum Circuit, Quantum Natural Gradient, Adam optimizer, Quantum Fisher Information Matrix, data re-uploading, crossover analysis, quantum machine Learning

1. INTRODUCTION

The realisation of near-term quantum advantage is closely tied to the effectiveness of variational hybrid algorithms, among which Variational Quantum Circuits (VQCs) are among the most widely studied [1], with noisy intermediate-scale quantum (NISQ) architectures remaining the dominant near-term testbed [2]. A VQC encodes classical data into a parametric quantum state, applies tunable unitary operations, and extracts class predictions via quantum measurement. The choice of optimiser critically governs convergence speed, susceptibility to barren plateaux [3], [4], and the final accuracy achievable for a given circuit depth and qubit budget. Classical first-order optimisers, principally Adam [5] dominate practical VQC training owing to their robustness and low memory requirements, and remain competitive against a broad range of alternatives in recent optimiser benchmarks [6]. However, the geometry of quantum state space, characterised by the Quantum Fisher Information Matrix (QFIM) F (the Fubini-Study metric tensor), is Riemannian rather than Euclidean. The Quantum Natural Gradient (QNG) [7] exploits

this geometry by preconditioning gradient updates with F , theoretically yielding parameter-invariant, second-order-efficient updates, a benefit recently confirmed both under improved initialisation schemes [8] and under realistic hardware noise [9]. Despite this theoretical appeal, systematic empirical comparisons against Adam in realistic multi-class classification tasks remain scarce.

A key missing dimension is the dependence of optimiser advantage on the quality of data encoding. Kimura and Mitarai [10] recently established a theoretical framework predicting a crossover in the relative advantage of geometry-aware optimisers as a function of the encoding overlap c_i a measure of how well encoded data states align with class centroids in Hilbert space and introduced the information acquisition efficiency g_i to quantify per-epoch learning efficiency. Their analysis, conducted in the quantum phase estimation (QPE) setting, has not previously been tested in a VQC classification context. This work makes the following contributions:

- a) a rigorous 1,000-epoch empirical comparison of Adam and QNG on ten-class digit classification with data re-uploading;
- b) the first empirical validation of the Kimura-Mitarai crossover prediction in a VQC setting;
- c) a corrected Fisher-empirical g_i estimator for Adam, resolving a degenerate constant-QFIM proxy in prior implementations; and
- d) practical pre-training guidelines for optimiser selection based on encoding overlap statistics, applicable to NISQ deployments without additional quantum resources.

2. METHODS

2.1 Variational Quantum Circuits

A VQC parametrises a unitary $U(x, w)$ acting on an N_Q -qubit register, where $x \in \mathbb{R}^{N_Q}$ are encoded input features and $w \in \mathbb{R}^p$ are trainable parameters. The output quantum state is

$$|\psi(x, w)\rangle = U(x, w) |0\rangle^{\otimes N_Q}$$

Angle encoding maps each PCA feature x_j to a single-qubit rotation via $R_Y(x_j)$. The data re-uploading technique [11] repeats this encoding at each variational layer l , enabling the circuit to approximate arbitrary functions via multi-frequency Fourier representations, a property confirmed in recent image-classification VQCs [12], [13] and analysed further in [14].

2.2 Quantum Fisher Information Matrix

The QFIM $F \in \mathbb{R}^{p \times p}$ is the metric tensor of the manifold of quantum states parametrised by w , defined via the Fubini-Study metric:

$$F_{jk} = 4 \operatorname{Re} [\langle \partial_j \psi | \partial_k \psi \rangle - \langle \partial_j \psi | \psi \rangle \langle \psi | \partial_k \psi \rangle]$$

where $|\partial_j \psi\rangle = \frac{\partial}{\partial w_j} |\psi(x, w)\rangle$. The diagonal element F_{jj} used in our diagonal QFIM approximation admits an exact parameter-shift fidelity estimator [7], [19]:

$$F_{jj} = 1 - |\langle \psi(w_j + \frac{\pi}{2}) | \psi(w_j - \frac{\pi}{2}) \rangle|^2$$

2.3 Quantum Natural Gradient

Standard gradient descent updates $w \leftarrow w - \eta \nabla_w L$. QNG [7] replaces the Euclidean gradient with the natural gradient preconditioned by the inverse QFIM:

$$w \leftarrow w - \eta F^\dagger g$$

where $g = \nabla_w L$ and $F^\dagger$ denotes the Moore-Penrose pseudoinverse. Under the diagonal approximation, each parameter update simplifies to

$$w_j \leftarrow w_j - \eta \frac{g_j}{F_{jj} + \lambda}$$

where $\lambda = 10^{-2}$ is a regularisation constant and updates are clipped to $[-5, 5]$ to prevent instability.

2.4 Parameter-Shift Rule

For gates of the form $G(w_j) = \exp(-i \frac{w_j}{2} P)$ with P a Pauli operator, the parameter-shift rule [15], [16] yields the exact gradient without finite-difference approximation:

$$\frac{\partial f_i}{\partial w_j} = \frac{f_i(w_j + \frac{\pi}{2}) - f_i(w_j - \frac{\pi}{2})}{2}$$

where $f_i(w) = \langle \psi(w) | Z_i | \psi(w) \rangle$ is the Pauli-Z expectation for qubit i . This yields the Jacobian $J \in \mathbb{R}^{N_Q \times p}$ with $J_{ij} = \partial f_i / \partial w_j$. Recent work on numerical gradient and Hessian estimation for VQAs provides further error-robust alternatives to this exact rule [17].

2.5 Information Acquisition Efficiency

Following Kimura and Mitarai [10], the information acquisition efficiency per epoch is

$$g_i = \frac{\bar{F}}{(N \cdot T)^2}$$

where $\bar{F} = \frac{1}{p} \sum_j F_{jj}$ is the mean diagonal QFIM, N is the batch size, and $T = N_Q \times L$ is the total depth parameter. For Adam, $\bar{F}$ is replaced by the Fisher-empirical estimator accumulated from the Jacobian computed during gradient steps:

$$\bar{F}_{\text{emp}} = \frac{1}{B} \sum_{i=1}^B \frac{1}{p} \sum_{j=1}^p \|J_j(x_i)\|^2$$

where B is the batch size and $J_j(x_i)$ is the j -th column of the Jacobian evaluated at sample x_i .

2.6 Encoding Overlap

The encoding overlap quantifies the alignment between the data-encoded state and the target class centroid in Hilbert space:

$$c_i = |\langle \psi_{\text{data}}(x_i) | \psi_{\text{centroid}}(c_k) \rangle|^2, \quad k = y_i$$

where $|\psi_{\text{centroid}}(c_k)\rangle$ is derived from the L_2 -normalised mean of encoded training samples in class k . This metric adapts the QPE-based definition of Kimura and Mitarai [10] to angle-encoded VQC states.

2.7 Dataset and Preprocessing

The scikit-learn digits dataset [18] comprises 1,797 samples of 8×8 grayscale handwritten digit images (classes 0–9), normalised to $[0, 1]$. Standard scaling (zero mean, unit variance) is applied before PCA reduction to $d = N_Q = 8$ components (52.9% explained variance). PCA features are mapped to angles via

$$x'_j = \frac{\pi}{2} \left(2 \frac{x_{j - \min_j x}}{\max_x - \min_j x + \varepsilon} - 1 \right), \quad x'_j \in \left[-\frac{\pi}{2}, \frac{\pi}{2} \right]$$

The dataset is split into $N_{\text{train}} = 300$ and $N_{\text{test}} = 100$ samples (stratified by class).

2.8 Circuit Architecture

The VQC employs $N_Q = 8$ qubits and $L = 3$ variational layers. Each layer l applies a hardware-efficient block

$$U_l(w) = CX_{\text{circ}} \otimes_{j=0}^{N_Q-1} R_x \left(w_{3j+2}^{(l)} \right) R_z \left(w_{3j+1}^{(l)} \right) R_y \left(w_{3j}^{(l)} \right)$$

where CX_{circ} denotes circular CNOT entanglement ($j \rightarrow j+1$ and $N_Q-1 \rightarrow 0$). Data re-uploading [11] inserts $R_Y(x'_j/2)$ on each qubit after every entanglement block except the last. The full circuit unitary is:

$$U(x, w) = \prod_{l=1}^L [E_l(x) U_l(w)] \cdot E_0(x)$$

where $E_0(x) = \otimes_j R_Y(x'_j)$ is the initial encoding layer and $E_l(x) = \otimes_j R_Y(x'_j/2)$ are the re-uploading layers. Total trainable quantum parameters: $N_W = N_Q \times L \times 3 = 72$.

2.9 Readout and Loss Function

The quantum feature vector $f(x, w) \in [-1, 1]^{N_Q}$ comprises Pauli-Z expectation values

$$f_i(x, w) = \langle \psi(x, w) | Z_i | \psi(x, w) \rangle$$

A classical linear head $W \in \mathbb{R}^{K \times N_Q}$ (with $K = 10$ classes) and bias $b \in \mathbb{R}^K$ produce logits $z = Wf + b$, followed by softmax.

$$p_k(x, w, W, b) = \frac{\exp(z_k)}{\sum_{k'=1}^K \exp(z_{k'})}$$

Training minimises the cross-entropy loss over a mini-batch B of size N .

$$L(w, W, b) = -\frac{1}{N} \sum_{i=1}^N \sum_{k=1}^K y_{ik} \log p_k(x_i)$$

2.10 Gradient Computation via Chain Rule

Quantum parameter gradients are obtained by the chain rule through the Jacobian $J \in \mathbb{R}^{N_Q \times p}$:

$$\frac{\partial L}{\partial w} = \sum_i \sum_k \left(\frac{\partial L}{\partial f_k} \right) \cdot J_{ki}$$

$$\frac{\partial L}{\partial w_j} = \sum_{i=1}^N \sum_{k=1}^K \frac{\partial L}{\partial f_k} \cdot J_{kj}$$

where $J_{kj} = \partial f_k / \partial w_j$ is computed exactly via the parameter-shift rule [15]. Classical head gradients follow from the softmax cross-entropy identity: $\partial L / \partial z_k = p_k - y_k$.

2.11 Adam Optimiser

Adam [5] maintains bias-corrected first and second moment estimates

$$\begin{aligned} m_t &= \beta_1 m_{t-1} + (1 - \beta_1) g_t \\ v_t &= \beta_2 v_{t-1} + (1 - \beta_2) g_t^2 \end{aligned}$$

with the parameter update:

$$w \leftarrow w - \eta \cdot \frac{\hat{m}_t}{\sqrt{\hat{v}_t} + \varepsilon}$$

Hyperparameters: $\beta_1 = 0.9$, $\beta_2 = 0.999$, $\eta = 0.05$, $\varepsilon = 10^{-8}$. Both optimisers apply cosine annealing:

$$\eta_t = \eta \cdot \left(0.5 + 0.5 \cos \left(\frac{\pi t}{T_{\max}} \right) \right)$$

2.12 Lazy QFIM Update Scheme

Full per-step QFIM computation costs $O(p)$ additional statevector evaluations per step, making QNG substantially more expensive than Adam. We update the QFIM cache every $Q_{\text{FREQ}} = 5$ steps via an exponential moving average (EMA):

$$\bar{F}_t = \alpha \bar{F}_{\text{new}} + (1 - \alpha) \bar{F}_{t-1}, \quad \alpha = 0.3$$

QFIM is computed from a single sample per update step, and the cached value is reused in intervening steps. This reduces QNG's overhead to near-parity with Adam ($< 10\%$ wall-clock difference over 1,000 epochs), a cost-saving strategy consistent with recent noise-robust QNG variants [8], [9].

2.13 Computational Environment

All experiments were executed on a single workstation running Windows 11 (64-bit), with an Intel Core i7 processor and 16 GB of RAM. No GPU acceleration or distributed/HPC resources were used; all statevector simulations were performed on CPU. The quantum circuits and training pipeline were implemented in Python 3, using Qiskit for circuit construction and NumPy/scikit-learn [18] for numerical operations and dataset handling (Section 2.7), with a custom NumPy statevector simulator used for the timing results reported in Section 3.1 ($\sim 5\times$ speedup over the reference Qiskit Statevector backend at 8 qubits, run under the same hardware configuration). Reporting the hardware and software configuration follows standard practice for reproducibility in empirical quantum machine learning studies [1], [6].

3. RESULTS AND DISCUSSION

3.1 Full Training Dynamics (Phase B)

Table 1 summarizes training outcomes over 1,000 epochs. Adam achieves a final test

accuracy of 95.0% and QNG achieves 92.0%. Both optimizers exhibit rapid early improvement: Adam reaches 85% at epoch ≈ 100 and QNG at epoch ≈ 150 , reflecting QNG's superior initial convergence due to curvature-adaptive step scaling. However, Adam continues to improve beyond epoch 400, ultimately surpassing QNG's plateau.

Table 1. Training Summary ($N_{\text{epochs}} = 1,000$; $N_{\text{train}} = 300$; $N_{\text{test}} = 100$).

Metric	Adam	QNG
Final test accuracy (%)	95.0	92.0
Epoch at 85% accuracy	≈ 100	≈ 150
Epoch at 90% accuracy	≈ 400	≈ 325
Minimum cross-entropy loss	0.048	0.054
Loss at epoch 1,000	0.096	0.093
Wall-clock time (s)	27,777	25,299

Adam's loss trajectory oscillates in epochs 75–500 (range 0.13–0.46), consistent with stochastic gradient noise and the limited 80-sample coverage per epoch. After epoch 500, loss stabilizes below 0.15. QNG converges more smoothly (loss: $0.895 \rightarrow 0.060$ from epochs 25–900) due to natural gradient curvature regulation, but plateaus at 92% accuracy from epoch 475 onward. This plateau is attributed to the diagonal QFIM approximation neglecting off-diagonal parameter correlations that become dominant near convergence [19], [20].

3.2 Crossover Analysis (Phase A)

Table 2 presents per-bin accuracy results (50 epochs per bin, up to 60 samples). A clear crossover is observed: QNG outperforms Adam exclusively in Bin 1 ($c_i \in [0.389, 0.540]$), achieving 18% versus Adam's 9%. From Bin 2 onward, Adam dominates by an increasing margin reaching 67% versus QNG's 46% in Bin 3.

Table 2. Crossover Results (50 epochs per bin; up to 60 samples per bin; QNG g_i column omitted for brevity, near-constant $\approx 9.3 \times 10^{-5}$ across all bins).

Bin	Range c_i	n	Adam (%)	QNG (%)	Winner	Adam g_i	QNG g_i
1	[0.389, 0.540]	12	9.0	18.0	QNG	1.48×10^{-4}	9.29×10^{-5}
2	[0.540, 0.690]	28	23.0	20.0	Adam	1.17×10^{-4}	9.33×10^{-5}
3	[0.690, 0.841]	58	67.0	46.0	Adam	7.80×10^{-5}	9.43×10^{-5}
4	[0.841, 0.992]	201	59.0	47.0	Adam	3.04×10^{-5}	9.36×10^{-5}

This pattern is qualitatively consistent with the Kimura-Mitarai prediction [10]: at low overlap, the QFIM-preconditioned update provides genuine advantage over Adam's momentum-based heuristic. The empirical crossover threshold $c^* \approx 0.54$ –0.62 lies well below the dataset mean overlap $\mu c = 0.849$, explaining Adam's overall Phase B superiority.

3.3 Information Acquisition Efficiency

The two estimators exhibit qualitatively opposite scaling. Adam's Fisher-empirical g_i fits a power law

$$g_i(\text{Adam}) \approx A \cdot c_i^{-2.19}$$

As c_i increases, class centroids become more geometrically distinct in the encoded Hilbert space, gradient norms $\|J_j(x)\|^2$ decrease, and each step acquires less Fisher information per unit update cost. In contrast, QNG's QFIM-fidelity g_i fits

$$g_i(\text{QNG}) \approx B \cdot c_i^{0.02}$$

The near-zero exponent reflects that $F_{jj} = 1 - |\langle \psi^+ | \psi^- \rangle|^2$ measures circuit-intrinsic geometric distinguishability of shifted states, which is largely independent of the data distribution. These empirical scalings diverge from the Kimura-Mitarai QPE predictions [10] ($g_i \sim c^l$ for QFT-QPE; $g_i \sim c^2$ for Hadamard-test QPE), which we attribute to the fundamental difference in task structure between QPE phase estimation and VQC multi-class classification.

3.4 Practical Optimiser Selection

Based on the empirical crossover threshold $c^* \approx 0.54-0.62$, we propose the following decision rule: if the pre-training mean encoding overlap $\hat{\mu}_c = \frac{1}{N} \sum_i c_i$ (computed cheaply via the overlap definition in Section 2.6 before training) satisfies $\hat{\mu}_c < c^*$, prefer QNG for its superior early convergence; otherwise, prefer Adam for its higher asymptotic accuracy, consistent with recent optimiser benchmarking results on variational circuits [6]. The quantity $\hat{\mu}_c$ is easily computed in $O(N \cdot N_Q)$ operations and requires no circuit evaluations beyond the initial forward pass.

3.5 Lazy QFIM Scheme

The lazy QFIM update ($Q_{\text{FREQ}} = 5$, $\alpha = 0.3$) reduces QNG's wall-clock time from a naïve $\approx p$ -fold overhead versus Adam to less than 10% (25,299s vs 27,777s). The EMA provides temporal smoothing consistent with the slow variation of QFIM near convergence, in line with recent noise-aware QNG scheduling strategies [8], [9]. We recommend $Q_{\text{FREQ}} \in \{5, 10\}$ as a practical default for circuits of similar depth.

3.6 Limitations and Future Work

The diagonal QFIM approximation discards off-diagonal correlations, likely contributing to QNG's accuracy plateau at epoch 475. Full QFIM or Kronecker-factored approximations [19] may narrow the accuracy gap, and QFIM-free stochastic natural-gradient schemes offer a lower-overhead alternative worth benchmarking against our lazy-update approach [20]. The dataset (300 training samples, 52.9% retained PCA variance) imposes an information-theoretic ceiling: classical SVM on identical features achieves 85%, suggesting the VQC's 95% reflects significant classical head contribution. Future work should examine: (i) block-diagonal and Kronecker-factored QFIM approximations [19]; (ii) the crossover behaviour under amplitude encoding on larger datasets (MNIST-784, Fashion-MNIST), building on recent rotation-invariant and re-uploading-based image classifiers [12], [13]; (iii) shot-noise effects when simulating hardware constraints, as examined for QNG under realistic noise models [9]; and (iv) the interaction between re-uploading depth L and the crossover threshold c^* .

4. CONCLUSION

We have presented a rigorous 1,000-epoch empirical comparison of Quantum Natural Gradient and Adam optimisers in a Variational Quantum Circuit for ten-class digit recognition. Adam achieves 95% test accuracy; QNG achieves 92% with markedly superior early convergence — reaching 89% at epoch 150 versus Adam's epoch 400. Crossover analysis across four encoding-overlap bins confirms the Kimura-Mitarai regime-dependence prediction [10]: QNG is advantageous for $c_i < c^* \approx 0.54\text{--}0.62$, while Adam dominates at higher overlap. The corrected Fisher-empirical g_i estimator reveals a strongly decaying power law ($c_i^{-2/9}$) for Adam, in contrast to QNG's near-constant QFIM profile ($c_i^{0.02}$), reflecting the data-dependent versus circuit-intrinsic nature of the respective information metrics. A lazy EMA-QFIM scheme equalises computational cost to within 10%. These findings provide actionable guidelines for optimiser selection in near-term quantum machine learning and constitute the first empirical validation of the Kimura-Mitarai efficiency framework in a variational classification context.

ACKNOWLEDGEMENTS

This work was supported by Politeknik Negeri Bengkalis

CONFLICT OF INTEREST

The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

REFERENCES

- [1] Y. Ruiqi, Z. Xinyun, and R. Shuang, "A review of quantum machine learning algorithms based on variational quantum circuit," *Journal of Computer Research and Development*, vol. 62, no. 4, pp. 820–838, 2025.
- [2] X.-D. Xie et al., "Advances in quantum computation in NISQ era," *Entropy*, vol. 27, no. 10, art. 1074, 2025. doi: 10.3390/e27101074
- [3] M. Larocca, S. Thanasilp, S. Wang, K. Sharma, J. Biamonte, P. J. Coles, L. Cincio, J. R. McClean, Z. Holmes, and M. Cerezo, "Barren plateaus in variational quantum computing," *Nature Reviews Physics*, vol. 6, pp. 361–381, 2024. doi: 10.1038/s42254-024-00701-9
- [4] J. Cunningham and J. Zhuang, "Investigating and mitigating barren plateaus in variational quantum circuits: a survey," *Quantum Information Processing*, vol. 24, art. 48, 2025. doi: 10.1007/s11128-025-04691-x
- [5] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. 3rd Int. Conf. Learning Representations (ICLR)*, San Diego, CA, USA, 2015. [Online]. Available: <https://arxiv.org/abs/1412.6980>
- [6] X. Bonet-Monroig, H. Wang, D. Vermetten, B. Senjean, C. Moussa, T. Bäck, V. Dunjko, and T. E. O'Brien, "Performance comparison of optimization methods on variational quantum algorithms," *Physical Review A*, vol. 107, no. 3, p. 032407, 2023. doi: 10.1103/PhysRevA.107.032407
- [7] J. Stokes, J. Izaac, N. Killoran, and G. Carleo, "Quantum natural gradient," *Quantum*, vol. 4, p. 269, 2020. doi: 10.22331/q-2020-05-25-269
- [8] M. Minervini, D. Patel, and M. M. Wilde, "Quantum natural gradient with thermal-state initialization," *Physical Review A*, vol. 112, no. 2, p. 022424, 2025. doi: 10.1103/PhysRevA.112.022424

- [9] F. Dell'Anna, R. Gómez-Lurbe, A. Pérez, and E. Ercolessi, "Quantum natural gradient optimizer on noisy platforms: Quantum approximate optimization algorithm as a case study," *Physical Review A*, vol. 110, no. 4, p. 042408, 2024. doi: 10.1103/PhysRevA.110.042408
- [10] R. Kimura and K. Mitarai, "Information acquisition efficiency in quantum phase estimation," arXiv:2603.12706, Mar. 2026. [Online]. Available: <https://arxiv.org/abs/2603.12706>
- [11] A. Pérez-Salinas, A. Cervera-Lierta, E. Gil-Fuster, and J. I. Latorre, "Data re-uploading for a universal quantum classifier," *Quantum*, vol. 4, p. 226, 2020. doi: 10.22331/q-2020-02-06-226
- [12] P. San Sebastián Sein, M. Cañizo, and R. Orús, "Image classification with rotation-invariant variational quantum circuits," *Physical Review Research*, vol. 7, no. 1, p. 013082, 2025. doi: 10.1103/PhysRevResearch.7.013082
- [13] A. Mukhanbet and B. Daribayev, "A hybrid quantum-classical architecture with data re-uploading and genetic algorithm optimization for enhanced image classification," *Computation*, vol. 13, no. 8, art. 191, 2025. doi: 10.3390/computation13080191
- [14] N. A. Nemkov, E. O. Kiktenko, and A. K. Fedorov, "Fourier expansion in variational quantum algorithms," *Physical Review A*, vol. 108, no. 3, p. 032406, 2023. doi: 10.1103/PhysRevA.108.032406
- [15] K. Mitarai, M. Negoro, M. Kitagawa, and K. Fujii, "Quantum circuit learning," *Physical Review A*, vol. 98, no. 3, p. 032309, 2018. doi: 10.1103/PhysRevA.98.032309
- [16] M. Schuld, V. Bergholm, C. Gogolin, J. Izaac, and N. Killoran, "Evaluating analytic gradients on quantum hardware," *Physical Review A*, vol. 99, no. 3, p. 032331, 2019. doi: 10.1103/PhysRevA.99.032331
- [17] Y. S. Teo, "Optimized numerical gradient and Hessian estimation for variational quantum algorithms," *Physical Review A*, vol. 107, no. 4, p. 042421, 2023. doi: 10.1103/PhysRevA.107.042421
- [18] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.
- [19] B. Koczor and S. C. Benjamin, "Quantum natural gradient generalized to noisy and non-unitary circuits," *Physical Review A*, vol. 106, no. 6, p. 062416, 2022. doi: 10.1103/PhysRevA.106.062416
- [20] I. Kolotouros and P. Wallden, "Random natural gradient," *Quantum*, vol. 8, p. 1503, 2024. doi: 10.22331/q-2024-10-22-1503