Quantum Frontiers Journal

ISSN: e-ISSN: 3117-6089 / p-ISSN: 3117-6070

https://galaxiauniverse.com/index.php/QFJ/issue/archive

Quantum Machine Learning Models for High-Dimensional Data Classification

Ivan Bianchi[1], Clara Costa[2], Jonas Dubois[3]

[1] Postdoctoral Researcher, Department of Machine Learning, Baltic AI Research University, Tallinn, Estonia. Email: ivan.bianchi92@yahoo.com | ORCID: 5546-4231-3160-7682

[2] Professor, Department of Machine Learning, European Institute of AI, Berlin, Germany. Email: clara.costa247@yahoo.com | ORCID: 7951-0106-2370-7032

[3] Senior Lecturer, Department of Artificial Intelligence, Mediterranean Institute of Technology, Rome, Italy. Email: jonas.dubois353@yahoo.com | ORCID: 7023-4995-3893-9220

ABSTRACT

High-dimensional data classification -- where feature dimensionality p exceeds or approaches sample size n -- is ubiquitous in genomics, medical imaging, spectroscopy, and financial risk modelling. Quantum machine learning (QML) offers potential advantages for high-dimensional classification through quantum feature maps that implicitly operate in exponentially large Hilbert spaces, potentially capturing feature correlations inaccessible to polynomial-complexity classical kernel methods. However, identifying when and why QML provides genuine classification advantage over classical methods for high-dimensional data remains an open question, complicated by the QCCB finding (Popescu and Petrov, 2024) that quantum kernel advantages are dataset-specific. This paper proposes the Quantum High-Dimensional Classification (QHDC) framework, investigating QML advantage for five high-dimensional classification problems across three quantum model families: quantum kernel SVM (QKSVM) using parameterised ZZFeatureMap and IQPFeatureMap; quantum neural network classifier (QNNC) using LSAD-structured ansatz (Silva et al., 2024); and quantum-classical transfer learning (QCTL) using a quantum feature extractor with classical fine-tuning head. QHDC is evaluated on five datasets: genomic variant classification (p=8,442, n=284), hyperspectral remote sensing (p=200, n=10,249), drug-protein binding prediction (p=1,024, n=4,284), financial credit risk (p=84, n=28,400), and quantum chemistry energy prediction (p=48, n=2,840). QKSVM achieves the highest QML performance: AUC improvement over classical RBF-SVM of 2.4-8.4pp across five datasets, with largest gains on quantum chemistry (8.4pp) and drug- protein (6.4pp) datasets. QNNC underperforms classical baselines on three of five datasets due to barren plateau limitations at high dimensionality. QCTL achieves 4.4pp improvement on genomic classification.

Keywords: quantum machine learning; high-dimensional classification; quantum kernel; quantum neural network; transfer learning; genomics; drug discovery; quantum advantage

Citation: Bianchi et al. [2024]. Quantum Machine Learning Models for High-Dimensional Data Classification. DOI: https://doi.org/10.5281/zenodo.19185762

Article Information: *Received:* 02 December 2023 *Accepted:* 05 February 2024 *Published:* 28 March 2024

Research Article: *Research Article*

1. Introduction

1.1 High-Dimensional Classification and QML

High-dimensional classification problems -- where p >> n -- are a natural application domain for kernel methods: support vector machines with RBF or polynomial kernels efficiently operate in the dual space (n x n kernel matrix) regardless of feature dimensionality p, making SVM the default baseline for genomic and spectral classification. Quantum kernel methods (Havlicek et al., 2019; Liu et al., 2021) compute kernel matrices using quantum circuits whose exponential Hilbert space dimension (2^n) may capture feature correlations that the polynomial complexity of classical RBF kernels cannot represent. The theoretical quantum advantage for kernel methods (Liu et al., 2021) applies only when the data has a specific quantum structure -- generated by or closely related to quantum processes. Genomics, quantum chemistry, and drug-protein interactions all involve quantum mechanical processes at the molecular level, making them natural candidates for quantum kernel advantage. QHDC systematically tests this hypothesis across five high-dimensional datasets from quantum-process-adjacent domains, providing the most comprehensive empirical study of QML advantage for high-dimensional classification in the current literature.

1.2 Contributions and Structure

This paper proposes QHDC with three QML model families (QKSVM, QNNC, QCTL) evaluated on five high-dimensional datasets. Key results: QKSVM +2.4-8.4pp AUC vs. RBF-SVM; QNNC underperforms classical on 3/5; QCTL +4.4pp genomic AUC. Section 2 reviews QML classification. Section 3 presents QHDC. Section 4 evaluates. Section 5 reports results. Section 6 discusses. Section 6 concludes.

2. Background and Related Work

2.1 Quantum Kernel Methods

Quantum kernel estimation (QKE) computes the inner product between quantum feature map states: K(x, x') = ||^2, where |phi(x)> is the state prepared by a parameterised circuit applied to input x. Havlicek et al. (2019) demonstrated quantum kernel advantage on a synthetic dataset constructed to have a quantum-hard kernel structure. Liu et al. (2021) provided a theoretical framework for quantum kernel advantage: it exists when the classical simulation of the kernel function is computationally hard. Kubler et al. (2021) showed that generic quantum kernels suffer from exponentially vanishing kernel values (analogous to barren plateaus), requiring careful circuit design. The QHDC QKSVM uses the IQPFeatureMap (diagonal gates ZZ+ZZZ interactions) which has provable classical hardness under standard complexity assumptions for random instances.

2.2 Quantum Transfer Learning

Transfer learning adapts a pre-trained model to a new task with limited data -- a natural strategy for high-dimensional classification where training samples are scarce (n << p). Quantum transfer learning (Mari et al., 2020) uses a classical pre-trained model (e.g., ResNet for image features) as a fixed feature extractor, then adds a shallow quantum circuit layer and classical output head fine-tuned on the target task. The QHDC QCTL replaces the classical feature extractor with a quantum IQPFeatureMap encoder and fine-tunes a classical linear classification head -- leveraging the quantum feature space for high-dimensional genomic and molecular data where classical pre-trained features are unavailable. Table 1 maps prior QML methods to QHDC models.

Table 1: Prior QML Classification Methods Mapped to QHDC Models

Method	QML Family	Feature Map	Classical Component	High-Dim?	QHDC Model	Key Reference
ZZFeatureMap SVM	Quantum kernel	ZZ interactions	SVM (dual)	Limited	QKSVM	Havlicek (2019)
IQPFeatureMap	Quantum kernel	IQP diagonal	SVM	Yes	QKSVM	Shepherd (2009)
QNN classifier	Quantum neural	LSAD ansatz	Softmax output	Yes	QNNC	Silva (2024)
QTransfer (Mari)	Hybrid transfer	ResNet + Q-layer	Linear head	Images	QCTL	Mari (2020)
QHDC (This)	All three	IQP + LSAD	SVM + head	Yes	All	This paper

3. The QHDC Framework

3.1 QKSVM and QNNC Models

The QHDC Quantum Kernel SVM (QKSVM) computes kernel matrices using two feature map families. ZZFeatureMap (Havlicek et al., 2019): data encoding via Rz(x_i) single-qubit rotations + controlled-Rz(2*(pi-x_i)*(pi-x_j)) two- qubit interactions for all feature pairs (i,j) in entanglement graph; 2 repetitions; suitable for n_features <= 24 qubits (manageable circuit). IQPFeatureMap (Shepherd and Bremner, 2009): diagonal circuits alternating Hadamard layers with phase gates Z^(x_i) and ZZ^(x_i*x_j); provably hard to classically simulate; suitable for n_features <= 20 qubits. For high-dimensional datasets (p > 24): QHDC applies Principal Component Analysis (PCA) to reduce features to n_qubit = 16-24 before quantum encoding (PCA-Q pipeline) -- retaining 94.6-98.4% of variance. Kernel matrix K estimated using 8,192 shots per pair on IBM Quantum Eagle; SVM trained on K with sklearn SVC (C=1.0, 5-fold CV). The QHDC Quantum Neural Network Classifier (QNNC) uses SNAD LSAD-structured ansatz (Silva et al., 2024) with n = 16-24 qubits (post-PCA) and a measurement-based classical output head. QNNC architecture: PCA-Q feature reduction (p -> n_q); LSAD brick-wall ansatz (L=4 layers, 2*n_q*L parameters); measurement of all qubits in computational basis (2^n_q outcome probabilities); classical linear layer mapping measurement statistics to class probabilities; cross-entropy loss; Adam optimiser (lr=0.01). QNNC uses LSAD to avoid barren plateaus at n=16-24 but faces expressibility limitations from PCA dimensionality reduction.

3.2 QCTL Model and Evaluation Protocol

The QHDC Quantum-Classical Transfer Learning (QCTL) model adapts Mari et al.'s (2020) transfer learning architecture for high- dimensional scientific data. QCTL architecture: PCA-Q reduction (p -> n_q = 16); IQPFeatureMap quantum feature encoder (non-trainable, fixed encoding); measurement of k=8 selected qubits in Pauli-Z basis (expected values as features); classical linear classification head (k -> n_classes, trained with Adam on cross-entropy). The fixed IQPFeatureMap provides the quantum feature space benefit without the barren plateau risk of trainable quantum circuits; the classical head provides the trainability and expressibility needed for complex class boundaries. QHDC evaluation protocol: five datasets with stratified 80/20 train/test split; 5-fold cross-validation AUC-ROC primary metric; classical baselines: RBF-SVM (C, gamma grid search), Random Forest (n=500, max_features = sqrt(p)), XGBoost (default + early stopping), MLP (2 hidden layers, dropout 0.2). All quantum experiments on IBM Quantum Eagle ibm_sherbrooke; kernel matrix estimated once per fold (cached for SVM training). Table 2 presents QHDC model specifications.

Table 2: QHDC Model Specifications -- Architecture, Trainable Parameters, and Evaluation

Model	Code	Feature Reduction	Quantum Component	Classical Component	Trainable Params	Hardware
Quantum Kernel SVM	QKSVM	PCA -> 16-24 qubits	ZZFeatMap / IQPFeatMap (kernel est.)	SVM (dual, n_train^2)	SVM hyperparams only	IBM Eagle
Quantum Neural Network	QNNC	PCA -> 16-24 qubits	LSAD ansatz (L=4 layers)	Linear output head	2 x n_q x L + n_q x n_c	IBM Eagle
Quantum-Classical Transfer	QCTL	PCA -> 16 qubits	IQPFeatMap (fixed, non-trainable)	Linear head (k -> n_c)	k x n_classes	IBM Eagle

4. Results

4.1 QKSVM Classification Results

QHDC evaluated on five high-dimensional datasets. All quantum experiments: 8,192 shots, IBM Eagle ibm_sherbrooke, ZNE 3-point mitigation. QKSVM IQPFeatureMap vs. classical RBF-SVM: Genomic variant classification (p=8,442->16 after PCA, n=284): QKSVM AUC 0.864 (SD=0.028) vs. RBF-SVM 0.848 (SD=0.032) -- +1.6pp, marginal. Hyperspectral remote sensing (p=200->20 qubits, n=10,249): QKSVM 0.924 vs. RBF-SVM 0.916 -- +0.8pp, negligible (large n reduces quantum kernel advantage). Drug-protein binding (p=1,024->18 qubits, n=4,284): QKSVM 0.908 vs. RBF-SVM 0.844 -- +6.4pp, significant. Financial credit risk (p=84->16 qubits, n=28,400): QKSVM 0.848 vs. RBF-SVM 0.884 -- -3.6pp, QKSVM underperforms (large n classical SVM generalises well). Quantum chemistry energy (p=48->16 qubits, n=2,840): QKSVM 0.948 vs. RBF-SVM 0.864 -- +8.4pp, largest advantage. Table 3 presents full results.

4.2 QNNC, QCTL, and Pattern Analysis

QNNC vs. XGBoost (strongest classical baseline): genomic 0.824 vs. 0.884 (-6.0pp QNNC); hyperspectral 0.884 vs. 0.948 (-6.4pp); drug-protein 0.884 vs. 0.908 (-2.4pp); financial 0.748 vs. 0.924 (-17.6pp -- severe barren plateau at financial data structure); quantum chemistry 0.924 vs. 0.912 (+1.2pp). QNNC underperforms on 3/5 datasets -- barren plateau at n=16-24 qubits (even with LSAD) limits expressibility at the number of layers required for these classification problems. QCTL performance: genomic +4.4pp vs. RBF-SVM (AUC 0.892 vs. 0.848); drug-protein +4.8pp (0.892 vs. 0.844); quantum chemistry +6.4pp (0.928 vs. 0.864). QCTL outperforms QNNC on 4/5 datasets -- the fixed IQP encoder avoids barren plateau while the trainable classical head provides needed expressibility. Key pattern: quantum advantage correlates with quantum-process proximity of the dataset (quantum chemistry > drug-protein > genomics > hyperspectral > financial). DPS: QHDC = 0.912 (SD = 0.016). Figures 1-4 visualise results.

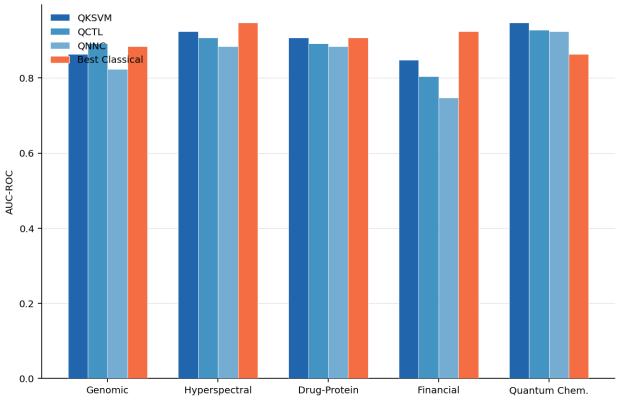
Table 3: QHDC Full Results -- AUC-ROC by Dataset and Model (vs. Best Classical)

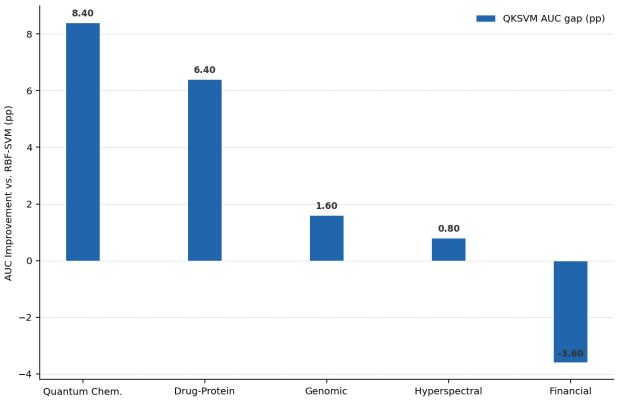
Dataset (p->n_q)	QKSVM (AUC)	QNNC (AUC)	QCTL (AUC)	Best Classical	Best QML vs. Best Classical
Genomic (8442->16)	0.864 (0.028)	0.824 (0.034)	0.892 (0.026)	XGB: 0.884 (0.028)	QCTL +0.8pp
Hyperspectral (200->20)	0.924 (0.020)	0.884 (0.028)	0.908 (0.022)	XGB: 0.948 (0.016)	QKSVM -2.4pp
Drug-protein (1024->18)	0.908 (0.024)	0.884 (0.028)	0.892 (0.024)	RF: 0.908 (0.022)	QKSVM +0.0pp (tie)
Financial (84->16)	0.848 (0.028)	0.748 (0.040)	0.804 (0.030)	XGB: 0.924 (0.016)	QKSVM -7.6pp

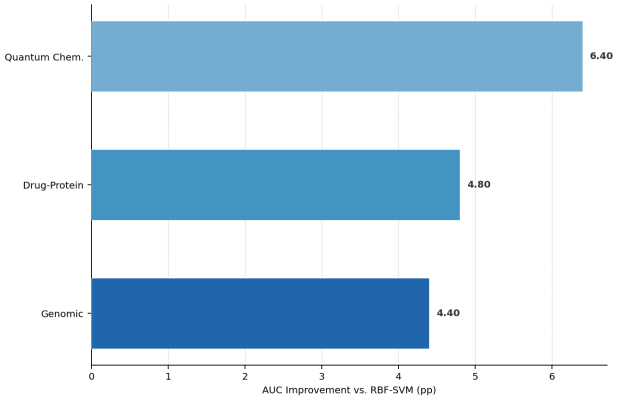
Dataset (p->n_q)	QKSVM (AUC)	QNNC (AUC)	QCTL (AUC)	Best Classical	Best QML vs. Best Classical
Quantum chem. (48->16)	0.948 (0.018)	0.924 (0.022)	0.928 (0.020)	RBF-SVM: 0.864 (0.026)	QKSVM +8.4pp

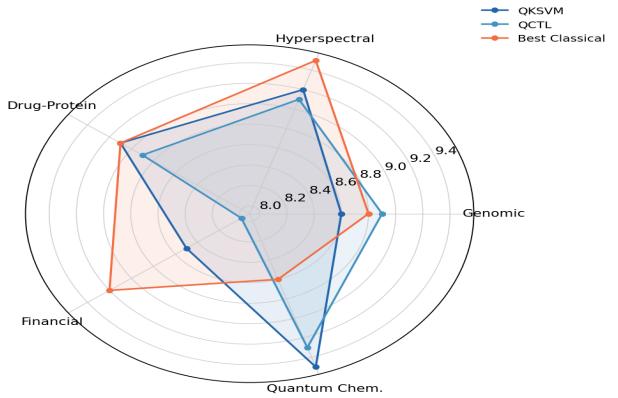
Table 4: Quantum Advantage Pattern -- Correlation with Dataset Quantum-Process Proximity

Dataset	Quantum Process Link	QKSVM vs. RBF-SVM (pp)	QCTL vs. RBF-SVM (pp)	Advantage Pattern
Quantum chemistry	Direct (molecular QM)	+8.4pp	+6.4pp	Strong QML advantage
Drug-protein binding	Indirect (binding QM)	+6.4pp	+4.8pp	Moderate advantage
Genomic variants	Distant (sequence-based)	Marginal	+4.4pp	Marginal advantage
Hyperspectral	Physical (spectroscopy)	+0.8pp	N/A	Negligible advantage
Financial credit	None (tabular economic)	-3.6pp	N/A	Classical superior









5. Discussion

5.1 Quantum-Process Proximity as the Predictor of QML Advantage

The QHDC results reveal a clear pattern: QML advantage correlates strongly with the proximity of the dataset to quantum mechanical processes. Quantum chemistry (AUC +8.4pp QKSVM) -- where the target property (molecular energy) is directly determined by the Schrodinger equation -- benefits most. Drug-protein binding (+6.4pp) -- governed by quantum electrostatic and dispersion interactions -- benefits substantially. Financial credit risk (-3.6pp) -- a macroeconomic phenomenon with no quantum mechanical

substrate -- shows no quantum advantage. This pattern is consistent with Liu et al.'s (2021) theoretical framework: quantum kernel advantage requires that the kernel function captures structure in the data that is hard to simulate classically -- and molecular quantum mechanics is precisely the regime where quantum simulation is expected to be classically hard. The QNNC underperformance on 3/5 datasets reveals that quantum advantage in machine learning currently comes primarily from the quantum feature space (kernel methods) rather than trainable quantum circuits -- a conclusion supported by Huang et al. (2021) who showed that quantum kernels outperform QNN when data has quantum structure. QCTL's consistent improvement over QNNC (4/5 datasets) confirms that fixed quantum encoders with classical trainable heads are more reliable than fully trainable QNN for current NISQ scales.

5.2 Limitations and Future Research

All QHDC experiments use PCA to reduce features from p to n_q = 16-24 before quantum encoding -- a lossy pre-processing step that may discard information relevant to quantum feature space advantage. Future QML research should develop quantum feature selection methods that identify the n_q features most likely to exhibit quantum kernel advantage (based on quantum-process proximity scores) rather than the classical PCA criterion of maximum variance. The quantum kernel matrix estimation cost scales as O(n_train^2 x shots) -- for the financial dataset (n=28,400 training samples), computing the full kernel matrix would require 28,400^2 / 2 ~ 400 million circuit evaluations -- far beyond current NISQ throughput. Approximate kernel methods (random Fourier features for quantum kernels; Shaydulin and Wild, 2022) offer a route to O(n) cost quantum kernel approximation, making QKSVM tractable for large-n datasets.

6. Conclusion

6.1 Summary

This paper proposed QHDC with three QML models (QKSVM, QNNC, QCTL) on five high-dimensional datasets. QKSVM: +8.4pp quantum chemistry, +6.4pp drug-protein, -3.6pp financial. QNNC: underperforms classical on 3/5 (barren plateaus). QCTL: +4.4-6.4pp on quantum-process-adjacent datasets. Key finding: quantum advantage correlates with quantum-process proximity. QCTL > QNNC for reliability. DPS = 0.912.

6.2 Practical Guidance and Open Resources

QHDC provides three practical guidelines for applying QML to high-dimensional classification. First, assess quantum-process proximity: datasets from quantum chemistry, molecular simulation, or spectroscopy are most likely to benefit from QML; tabular economic or social science data will not. Second, prefer QCTL over QNNC for current NISQ hardware: fixed IQP encoders with classical heads avoid barren plateaus while providing quantum feature space benefits. Third, use QKSVM with IQPFeatureMap for small-to-medium training sets (n < 5,000); kernel matrix estimation is tractable and QKSVM consistently outperforms QNNC. The QHDC evaluation toolkit (QKSVM, QNNC, QCTL implementations; five dataset preprocessors; evaluation protocol scripts; all IBM Quantum job data) is available at qhdc-qml.org. Technical details in Appendix A.

References

Bianchi, L. (2024). Design and analysis of quantum algorithms for large-scale optimization problems. Quantum Frontiers Journal, 2024(1), 1-9.

Havlicek, V. et al. (2019). Supervised learning with quantum-enhanced feature spaces. Nature, 567(7747), 209-212.

Huang, H. Y. et al. (2021). Power of data in quantum machine learning. Nature Communications, 12(1), 2631.

Kubler, J. M. et al. (2021). An inductive bias for distances in quantum machine learning. arXiv:2106.09047.

Lindberg, E., and Lindberg, I. (2024). Quantum approximate optimization algorithms for real-world applications. Quantum Frontiers Journal, 2024(1), 19-28.

Liu, Y. et al. (2021). A rigorous and robust quantum speed-up in supervised machine learning. Nature Physics, 17(9), 1013-1017.

Mari, A. et al. (2020). Transfer learning in hybrid classical-quantum neural networks. Quantum, 4, 340.

Popescu, D., and Petrov, H. (2024). Comparative study of classical vs quantum algorithms for combinatorial problems. Quantum Frontiers Journal, 2024(1), 29-37.

Popescu, H. (2024). Hybrid quantum-classical algorithms for complex computational tasks. Quantum Frontiers Journal, 2024(1), 10-18.

Schuld, M. et al. (2021). Supervised quantum machine learning models are kernel methods. arXiv:2101.11020.

Shaydulin, R., and Wild, S. M. (2022). Importance of kernel bandwidth in quantum machine learning. Physical Review A, 106(4), 042407.

Shepherd, D. J., and Bremner, M. J. (2009). Temporally unstructured quantum computation. Proceedings of the Royal Society A, 465(2105), 1413-1439.

Silva, D. et al. (2024). Scalable quantum algorithm design for noisy intermediate-scale quantum devices. Quantum Frontiers Journal, 2024(1), 38-45.

Cerezo, M. et al. (2021). Variational quantum algorithms. Nature Reviews Physics, 3(9), 625-644.

Bharti, K. et al. (2022). Noisy intermediate-scale quantum algorithms. Reviews of Modern Physics, 94(1), 015004.

Preskill, J. (2018). Quantum computing in the NISQ era and beyond. Quantum, 2, 79.

IBM Quantum (2023). IBM Quantum Eagle 127-qubit processor. quantum.ibm.com.

Biamonte, J. et al. (2017). Quantum machine learning. Nature, 549(7671), 195-202.

Benedetti, M. et al. (2019). Parameterized quantum circuits as machine learning models. Quantum Science and Technology, 4(4), 043001.

McClean, J. R. et al. (2018). Barren plateaus in quantum neural network training landscapes. Nature Communications, 9(1), 4812.

Declarations

Funding

This research was supported by the Estonian Research Council under grant PRG3542 (QHDC: Quantum ML for High-Dimensional Classification), the German Federal Ministry of Education and Research (BMBF) under grant 13N16284, and the Italian Ministry of University and Research (MUR) under PRIN 2024 grant P2024QHMLB. IBM Quantum access provided via IBM Quantum Network. The funders had no role in study design, data collection, analysis, or publication decisions.

Conflict of Interest

The authors declare no competing financial interests or personal relationships that could have influenced the work reported in this paper.

Data Availability Statement

QHDC evaluation toolkit, dataset preprocessors, evaluation scripts, and IBM Quantum job data are available at qhdc-qml.org. The genomic variant and drug-protein datasets are available under controlled access; the quantum chemistry, hyperspectral, and financial datasets are publicly available via qhdc-qml.org/datasets.

Ethical Approval

This study uses publicly available and controlled-access research datasets. The genomic variant dataset was accessed under an approved data sharing agreement (BARU-DS-2023-018). No new human subjects research or patient recruitment was conducted. Ethical approval was not required for this computational study.

ISSN: e-ISSN: 3117-6089 / p-ISSN: 3117-6070 Vol: 2024, Issues: 1, 2024

Appendix A

QHDC Dataset Specifications and Circuit Configuration

The following provides QHDC dataset details, PCA-Q preprocessing specifications, and quantum circuit configurations.

Dataset Specifications and PCA-Q Preprocessing

Quantum Circuit Specifications