

Article

Benchmarking Classical and Quantum-Hybrid Clustering on Autism Spectrum Disorder Screening Data

José Armando Noguez Martínez * , Emmanuel Martínez-Guerrero  and Guo-Hua Sun 

Centro de Investigación en Computación, Instituto Politécnico Nacional, Unidad Profesional Adolfo López Mateos, Mexico City 07738, Mexico; emartinezg2025@cic.ipn.mx (E.M.-G.); gsun@cic.ipn.mx (G.-H.S.)

* Correspondence: jnoguezm2025@cic.ipn.mx

Abstract

Clustering may uncover latent behavioral structure in Autism Spectrum Disorder (ASD) screening data without using outcome labels, but the resulting partitions depend strongly on data geometry and the adopted similarity measure. Quantum-hybrid clustering offers alternative distance and similarity estimators, yet whether these subroutines improve on classical methods under controlled conditions remains unclear. We conduct a benchmark of k-means, DBSCAN, agglomerative clustering, and spectral clustering against their quantum-hybrid counterparts. All methods are evaluated in a common 13-dimensional representation, with hyperparameters selected exclusively through internal validation indices. The evaluation covers four synthetic geometries and a 13-dimensional PCA representation of an ASD screening dataset, each containing 300 samples and evaluated over 10 seed-defined stochastic runs. Clustering quality is measured using the Silhouette Index (SI), Davies–Bouldin Index (DBI), Calinski–Harabasz Index (CHI), Adjusted Mutual Information (AMI), and Adjusted Rand Index (ARI). Under this validation protocol, Q-means exactly recovers the Gaussian clusters and improves label agreement on anisotropic data, but it does not outperform classical k-means on the ASD screening data. Q-spectral significantly reduces DBI on Two Moons, Concentric Rings, and ASD screening data, although these reductions do not consistently translate into higher AMI or ARI. Q-DBSCAN and Q-agglomerative exhibit greater sensitivity to distance distortions and finite-shot noise. Overall, the results reveal geometry-dependent trade-offs rather than uniform quantum-hybrid superiority. We relate these findings to theoretical complexity, measurement noise, state-preparation costs, and eigensolver bottlenecks in near-term implementations.

Keywords: quantum machine learning; hybrid clustering algorithms; Autism Spectrum Disorder



MSC: 62H30; 68T10; 68Q12

Academic Editor: Jonathan Blackledge

Received: 14 July 2026

Revised: 12 August 2026

Accepted: 15 August 2026

Published: 22 August 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

The increasing availability of structured biomedical and behavioral data has created new opportunities for exploratory data analysis. In this context, clustering remains one of the most widely used unsupervised learning tasks, since it aims to identify groups of observations that are internally similar and externally distinct. Unlike supervised classification, clustering does not require class labels during model construction, which makes it useful for exploring latent structure, validating data representations, and comparing alternative notions of similarity [1,2].

Autism Spectrum Disorder (ASD) screening records represent a valuable benchmark for this type of exploratory analysis. Rather than serving as diagnostic tools, these datasets compile responses to standardized behavioral screening questionnaires (such as AQ-10) along with demographic indicators [3,4]. From a machine learning perspective, they provide a challenging high-dimensional testbed to evaluate whether clustering algorithms can identify coherent subgroups and recover underlying behavioral profiles without supervisory guidance.

Classical clustering algorithms approach this problem from different mathematical perspectives. Centroid-based methods such as *k*-means are computationally efficient but are sensitive to initialization and tend to favor compact, approximately spherical clusters. Density-based methods such as DBSCAN can detect irregular cluster shapes and noise, but their performance depends strongly on neighborhood parameters and may degrade in high-dimensional spaces. Graph-based methods such as spectral clustering can capture non-linear structures, although they often involve computationally expensive matrix operations. Hierarchical agglomerative methods provide a nested view of the data structure, but they may also become costly as the number of samples increases [1,2,5,6].

Quantum Machine Learning (QML) has been proposed as a possible route for improving some machine learning operations by exploiting quantum properties such as superposition and entanglement [7]. In clustering, quantum subroutines are particularly suited for specific mathematical tasks, such as pairwise distance estimation, inner-product evaluation, or similarity graph construction. Because complete end-to-end quantum implementations for many clustering families are rarely formulated as monolithic quantum circuits, hybrid quantum-classical approaches naturally emerge: they delegate selected metric evaluations to quantum subroutines while preserving classical optimization and control loops.

In the literature, various hybrid approaches for clustering have been proposed. One of the most explored algorithms is Q-means, which is an adaptation of the *k*-means algorithm to a quantum environment [8–12]. Beyond partitioning algorithms, other quantum clustering strategies have explored density-based queries (e.g., via quantum search and distance subroutines [12–14]), quantum hierarchical clustering, and quantum spectral clustering [15,16]. These approaches have been proposed with theoretical advantages under assumptions concerning quantum data access, oracle availability, and measurement precision.

Despite the growing interest in quantum clustering algorithms, comparatively few studies systematically evaluate several classical and quantum-hybrid clustering families under identical multi-dataset benchmarks, consistent preprocessing pipelines, and controlled validation criteria [17,18]. This type of benchmark is essential because the practical utility of hybrid methods depends not only on their theoretical complexity but also on their empirical behavior under specific data encodings, unsupervised parameter selection, and multi-seed evaluations without supervisory bias.

In this paper, we present a systematic empirical comparison between four foundational classical clustering families (centroid-based, density-based, hierarchical, and graph-based) and their respective quantum-hybrid counterparts formulated in the RAQun framework. The main contributions of this work are summarized as follows:

- **Comprehensive Multi-Dataset Benchmark:** We conduct a comparative evaluation across five diverse datasets ($N = 300$ samples and $D = 13$ features), comprising four synthetic geometric datasets (Gaussian, Anisotropic Blobs, Two Moons, and Concentric Rings) and a real-world biomedical ASD screening dataset.
- **Controlled Experimental Protocol:** We implement an unsupervised PCA dimensionality reduction pipeline and multi-seed stochastic replication (10 seed-matched stochas-

tic runs, $S = \{10, 20, \dots, 100\}$), where model selection is performed strictly using internal validation indices without target label leakage.

- **Statistical and Resource Analysis:** We evaluate clustering quality using internal and external indices supported by paired Wilcoxon signed-rank significance tests, accompanied by a detailed characterization of quantum circuit scaling, finite-shot measurement noise, and practical eigensolver bottlenecks.

The remainder of this paper is organized as follows. Section 2 presents the background on classical clustering, quantum computing, and quantum-hybrid clustering algorithms. Section 3 describes the datasets, preprocessing procedure, experimental setup, hyperparameter grids, and validation metrics. Section 4 reports and discusses the experimental results and limitations. Finally, Section 5 summarizes the main findings and future research directions.

2. Background

In this section, we introduce the necessary theoretical foundations for the correct understanding of both the classical and quantum components of this work. First, we provide the key concepts of classical clustering, as well as the paradigms evaluated in this paper, and their evaluation metrics. Then, we outline the basic concepts such as qubits, entanglement, and superposition. Finally, we discuss the hybrid quantum-classical methods and the quantum clustering algorithms considered herein.

2.1. Classical Clustering

Clustering, also known as data segmentation, is an unsupervised machine learning technique that divides a dataset into groups called clusters such that the elements of a cluster are similar to one another and different from the elements of other clusters. Once the algorithms are executed, it is possible to obtain a significant partition, a useful one, or both. Meaningful groups capture the natural structure of the data enabling an accurate description of real-world phenomena in fields such as biology, information retrieval, meteorology, social sciences, medical sciences, business, geography, geology, and security. In contrast, useful groups facilitate other tasks including summarization, compression, and efficient nearest-neighbor search. These algorithms are divided into different families according to the mathematical concept they use to generate a result. In this study, we select a representative algorithm from four of the most widely used families [1,6,19,20].

2.1.1. Centroid-Based Algorithms

These algorithms partition the dataset into k defined clusters, where each one has a representative point called a centroid [19]. The most extensively known member of this family is k-means [21]. This algorithm initializes k arbitrary centroids and assigns each object of the dataset to the clusters represented by its nearest centroid in terms of Euclidean distance. In each iteration, the centroids are recalculated as the vector mean of the objects assigned to each cluster from the previous iteration. It stops when a maximum number of iterations is reached or when the centroids do not change significantly. Although k-means is an efficient algorithm with linear time complexity, it is highly sensitive to the initialization of centroids and has difficulty generating clusters of arbitrary shapes and sizes [5].

2.1.2. Density-Based Algorithms

This family defines highly populated regions, areas where the data points have a number of neighbors greater than a predefined threshold. Clusters are formed by grouping nearby high-density regions [19]. One of the most broadly used algorithms of this family is DBSCAN (Density-Based Spatial Clustering of Applications with Noise) which requires two

hyperparameters to work: the radius of the neighborhood (ϵ) and the minimum number of points required within that radius ($MinPts$) to create a high-density region [22]. DBSCAN classifies the objects into three categories:

- Core points: Points that have at least $MinPts$ neighbors within their ϵ neighborhood.
- Border points: Points that are not core points, but are within the ϵ neighborhood of a core point.
- Noise points: Points that are neither core points nor border points, that is, they do not belong to any cluster.

One of the strengths of DBSCAN is its ability to identify clusters of arbitrary shapes and sizes, as well as to identify outliers. However, it has limitations in datasets with variable densities and in high-dimensional spaces [5].

2.1.3. Graph-Based Algorithms

This family of algorithms defines weighted undirected graphs from the available data, where each node represents an object, and every edge represents the similarity between two objects. Generally, clustering is constructed from the connected components of the subgraphs obtained after cuts in the graph. A representative of this family is Spectral Clustering (SC), which generates an undirected graph represented through an affinity matrix. Afterwards, the normalized Laplacian of the graph is computed in order to obtain the eigenvectors associated with the k smallest eigenvalues. Finally, k-means is applied to the space formed by the discovered eigenvectors [6]. Unlike k-means, SC is capable of capturing clusters of arbitrary shapes and sizes. However, it loses its efficiency, since it has a temporal complexity of $O(N^3)$ [6,23].

2.1.4. Hierarchical Algorithms

This group of algorithms builds a binary tree called dendrogram, which has a set of nested clusters, that is, a cluster can contain smaller clusters, and in turn, it is contained within other clusters hierarchically. There are two paradigms of hierarchical clustering: agglomerative and divisive. The first one starts from n individual clusters that are progressively merged. Meanwhile, the latter one begins with one cluster and iteratively divides it until all objects are separated into individual clusters [1]. Like SC, Generalized Agglomerative Schema (GAS) can obtain clusters with arbitrary shapes and sizes. Nevertheless, GAS has a time complexity of $O(N^3)$ [1].

2.1.5. Metric Evaluation

To evaluate the performance of clustering algorithms in terms of partition quality, internal validity indices are used which only need the generated partitions and the dataset to measure characteristics as dispersion, cohesion, and separability. The most used metrics for partition comparison include the Silhouette Index (SI), Davies–Bouldin Index (DBI), and Calinski–Harabasz Index (CHI) [24].

In addition, external validity indices can be used, which use previously defined supervised labels to measure how well the partitions correspond to the underlying natural structure of the data. The most common indices are the Adjusted Rand Index (ARI) and the Adjusted Mutual Information (AMI) [24,25].

2.2. Quantum Computing

Quantum computing is a paradigm that leverages quantum mechanical phenomena such as entanglement, superposition, and interference to process information beyond the capabilities of classical systems. The concept was first proposed by Richard

Feynman in 1982, who argued that classical computers could not efficiently simulate quantum systems [26].

2.2.1. Foundations

Quantum computing employs the quantum bit (qubit) as the minimum unit of information. Unlike classical bits, a qubit can represent the states $|0\rangle$ and $|1\rangle$ or a superposition of both.

Mathematically, a qubit is defined as follows:

$$|\psi\rangle = \alpha|0\rangle + \beta|1\rangle; \quad |\alpha|^2 + |\beta|^2 = 1; \quad \alpha, \beta \in \mathbb{C},$$

where $|\alpha|^2$ and $|\beta|^2$ are the probability amplitudes of measuring the qubit in the state $|0\rangle$ or $|1\rangle$, respectively.

It is possible to work with multiple qubits in the same algorithm. In this case, Hilbert spaces of 2^n dimensions are constructed, where n is the number of qubits. Currently, practical quantum computing architectures remain constrained by finite qubit counts and limited circuit depth, motivating the design of compact subroutines within hybrid workflows.

2.2.2. Quantum Subroutines

The hybrid quantum-classical paradigm offloads demanding tasks to quantum subroutines to potentially improve computational efficiency. In this work, we focus on four fundamental quantum subroutines that serve as the building blocks for our clustering algorithms:

Quantum State Preparation

A fundamental subroutine is state preparation for encoding classical data into the Hilbert space of the quantum system. This activity can be performed using different techniques, including basis encoding, which associates each classical bit with a qubit in the same state; amplitude encoding, which assigns each value of a normalized vector to the amplitudes of the quantum states; and angle encoding, which applies rotations on the Bloch sphere to encode the data [27].

Quantum Random Access Memory (QRAM)

It is possible to use a quantum equivalent of classical RAM that allows the parallel association of memory addresses with quantum-encoded values [28]. Thus, it allows us to store, process, and retrieve a set of N data points simultaneously using only $\lceil \log_2(N) \rceil$ qubits for the address register and M qubits for data encoding. However, it is crucial to emphasize that QRAM is currently a theoretical construct and does not exist commercially in modern NISQ hardware. Consequently, assuming an active QRAM is an idealized hypothesis; without it, loading classical data into a quantum processor introduces an $O(N \cdot d)$ bottleneck that negates potential logarithmic speedups. In this work, QRAM is implemented only via software simulation.

Swap Test (ST)

One of the most common operations in machine learning is the measurement of distances between points. The Swap Test allows us to compute the inner product of two vectors by means of controlled swap gates and measurements [10].

Hadamard Test (HT)

The Hadamard Test estimates the real part of the inner product between two quantum states using controlled state-preparation unitaries and measurement of an auxiliary qubit [9].

Quantum Phase Estimation (QPE)

Quantum Phase Estimation (QPE) is a foundational quantum algorithm for estimating the eigenphases of a unitary operator. When applied to an eigenstate, the phase register encodes an estimate of the corresponding eigenvalue, enabling spectral information to be extracted from the system [16,29].

Variational Quantum Eigensolver (VQE)

This algorithm relies on a classical optimizer to minimize a quantum ansatz. By optimizing the expected value, the ground state of the matrix is obtained, which corresponds to the eigenvector of the minimum eigenvalue [30].

2.2.3. Hybrid Clustering Algorithms

These approaches utilize the quantum subroutines described in Section 2.2.2 to enhance clustering performance. It is important to clarify that these algorithms are not fully quantum; they are hybrid schemes that delegate specific subroutines (predominantly pairwise distance estimation) to quantum circuits, while other steps (such as centroid updating, linkage criteria calculation, and eigenvector extraction) are computed classically. In this section, we present these formulated quantum-hybrid adaptations and describe their hybrid structure.

Q-Means [9,10,28]

In this work, we formulate and implement a customized Q-means version via a parallelized procedure. While standard Q-means algorithms in the literature do not detail data superposition for distance estimation, we utilize a simulated QRAM to encode the dataset in superposition. While the Swap Test was retained as a circuit-level baseline for resource comparison, the single-register Hadamard Test was used as the primary quantum distance estimator across the main clustering benchmark. This custom parallel distance estimation subroutine serves as the foundational building block integrated across the hybrid algorithms in this study.

Q-DBSCAN [12,14]

We formulate a hybrid density-based pipeline (Q-DBSCAN) where the distance matrix required for neighborhood queries is built by applying our custom parallel quantum distance subroutine (utilizing the Hadamard Test under superposition) directly within the classical DBSCAN clustering loop.

Q-Agglomerative Clustering [12]

We implement a customized hybrid hierarchical formulation that utilizes our parallel distance estimation subroutine (with the Hadamard Test under superposition) to build the pairwise distance matrix. The hierarchy is subsequently constructed classically using the selected linkage criterion, creating a bridge between quantum distance estimation and classical hierarchical linkages.

Q-Spectral Clustering [15,16,30]

Existing works on quantum spectral clustering typically assume the availability of a generic quantum subroutine for eigenvector extraction, without detailing its implementation or practical limitations. In this work, we implement and evaluate two concrete quantum methods for extracting the eigenvectors of the Laplacian matrix: the Quantum Phase Estimation (QPE) and the Variational Quantum Eigensolver (VQE). Through systematic simulation, we show that both subroutines fail under realistic constraints: QPE suffers from amplitude leakage due to limited register precision, and VQE faces a severe

classical memory bottleneck ($O(N^2)$ Pauli decomposition terms yielding $>130,000$ non-zero operators, with earlier benchmark scaling at $N = 348$ reaching ≈ 125 GB of RAM in classical statevector simulation). Consequently, we establish a practical hybrid workaround: computing the similarity graph quantumly via our parallel distance subroutine and calculating the eigenvectors classically.

3. Methodology

This section describes the experimental procedure followed to develop this comparative study between classical and quantum clustering algorithms. Three main stages were considered: the preparation of clinical data, the experimental configuration, and the validation of the results.

3.1. Benchmark Datasets and Unsupervised Preprocessing

To ensure broad geometric diversity and overcome the limitations of single-dataset evaluations, five distinct benchmark datasets were incorporated into the experimental suite.

3.1.1. Real-World ASD Screening Dataset

We employed an ASD screening dataset retrieved from the UCI Machine Learning Repository [31–33], consolidating screening responses across children, adolescents, and adults. The raw dataset compiles 20 candidate features alongside the binary screening reference label (Class/ASD), detailed in Table 1. During preprocessing, the cumulative score variable (result) was removed to eliminate exact collinearity with the individual questionnaire items (A_1 – A_{10}), leaving 19 preprocessed continuous and encoded input features.

Table 1. ASD screening dataset feature description.

Name	Type	Description
A1_Score–A10_Score	Binary	AQ-10 screening item responses (0 or 1)
age	Continuous	Age of the participant in years
gender	Binary	Gender of the participant
ethnicity	Categorical	Self-reported ethnic background
jaundice	Binary	Whether the participant was born with jaundice
family_pdd	Binary	Family history of Pervasive Developmental Disorders
country_of_res	Categorical	Country of residence
used_app_before	Binary	Prior screening application usage
result	Integer	Cumulative AQ-10 score ($\sum A_i$, excluded prior to PCA to avoid collinearity with A_1 – A_{10})
age_desc	Categorical	Age bracket (Child, Adolescent, Adult)
relation	Categorical	Relationship of respondent to participant
Class/ASD (Reference)	Binary	Behavioral screening reference label (retained solely for external evaluation)

Data Cleaning and Imputation

Missing numerical values in the age feature were imputed using the group median corresponding to the participant’s age category. Missing entries in categorical attributes (ethnicity and relation) were categorized as an explicit ‘‘Unknown’’ level. Categorical variables were subsequently mapped to numeric vectors via label encoding, and all binary features were formatted to $\{0, 1\}$.

Unsupervised Dimensionality Reduction (Zero Target Leakage)

To address the critical concern of target leakage, all supervised feature selection methods (such as Random Forest classifiers trained on the screening reference class) were entirely excluded. Instead, an unsupervised Principal Component Analysis (PCA) was applied to the 19 preprocessed continuous and encoded variables. Retaining $D = 13$ principal components accounted for 92.06% of the cumulative explained variance. Finally, each component was normalized into the compact interval $[0, 1]$ via min–max scaling:

$$x'_j = \frac{x_j - \min(x_j)}{\max(x_j) - \min(x_j)}, \quad \forall j \in \{1, \dots, D\}. \quad (1)$$

Figure 1 illustrates the individual and cumulative explained variance ratios across all 19 principal components, highlighting the selection of $D = 13$ components to preserve over 92% of the total variance without supervisory leakage.

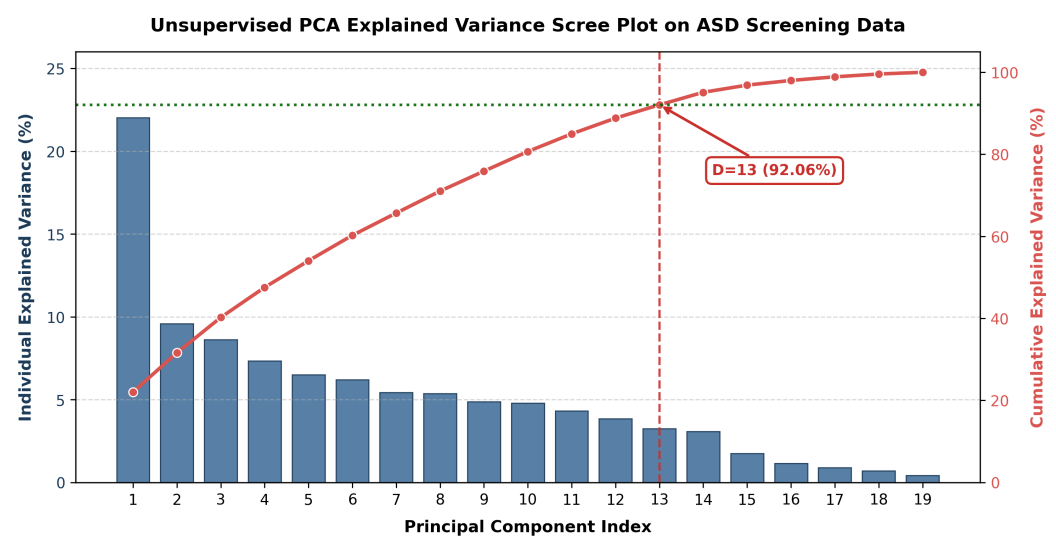


Figure 1. Unsupervised PCA scree plot showing individual and cumulative explained variance ratios across the 19 preprocessed ASD screening features. The horizontal dashed line denotes the 90% cumulative explained variance threshold, and the vertical dashed line indicates the selected number of principal components ($D = 13$), which captures 92.06% of the total variance.

3.1.2. Synthetic Geometric Benchmark Datasets

To systematically evaluate algorithmic behavior under controlled geometric challenges, four synthetic datasets ($N_{\text{total}} = 1000$, $D = 13$ features) were generated and scaled into $[0, 1]$:

1. Gaussian Clustered: Three well-separated isotropic Gaussian blobs representing linearly separable clusters.
2. Anisotropic Blobs: Three elongated clusters generated via affine coordinate transformations, introducing directional covariance and variance skewness.
3. Two Moons: Two interleaving half-circle manifolds exhibiting non-linear, non-convex geometric separation.
4. Concentric Rings: Two concentric circles with a large radial margin, representing linearly inseparable nested manifolds.

3.1.3. Stochastic Multi-Seed Sampling Protocol

To assess stochastic stability across different random initializations under identical experimental conditions, a multi-seed random sampling protocol was implemented. For each of the five datasets, 10 seed-defined stochastic subsets of size $N = 300$ were extracted from

the base data using deterministic random seeds $S = \{10, 20, 30, 40, 50, 60, 70, 80, 90, 100\}$, yielding a total of 50 benchmark subsets for robust statistical evaluation.

3.2. Quantum Circuit Architecture and Metric Formulation

To evaluate data similarities in the quantum state space, classical feature vectors $\mathbf{x} = [x_1, \dots, x_D]^T \in [0, 1]^D$ (with $D = 13$) are mapped into normalized quantum state vectors via amplitude encoding:

$$|\psi(\mathbf{x})\rangle = \frac{1}{\|\mathbf{x}\|_2} \sum_{j=1}^D x_j |j\rangle, \quad \text{where } \|\mathbf{x}\|_2 = \sqrt{\sum_{j=1}^D x_j^2}. \quad (2)$$

For Q-means, each updated centroid was L2-normalized before amplitude encoding to ensure valid quantum-state preparation.

3.2.1. Single-Register Hadamard Test (InnerProduct1R)

Standard inner product estimation frequently employs the Swap Test, which requires two separate feature registers ($2 \times 4 = 8$ feature qubits) alongside address and auxiliary registers, totaling 18 qubits. To optimize quantum simulation efficiency and reduce the Hilbert state space by 87.5%, we formulate a single-register Hadamard Test circuit requiring only 15 qubits in total ($n_a = \lceil \log_2 N \rceil = 9$ address qubits for multi-sample batch superposition, $n_f = \lceil \log_2 D \rceil = 4$ feature qubits, 1 state-preparation qubit, and 1 auxiliary control qubit $|0\rangle_{\text{aux}}$).

The circuit initializes the control qubit in the equal superposition state $(|0\rangle + |1\rangle)/\sqrt{2}$ via a Hadamard gate. A controlled unitary operator $\mathcal{U}$ sequentially prepares $|\psi(\mathbf{x})\rangle$ conditioned on $|0\rangle_{\text{aux}}$ and $|\psi(\mathbf{z})\rangle$ conditioned on $|1\rangle_{\text{aux}}$. Applying a final Hadamard transformation on the control qubit yields the composite interferometric state:

$$|\Phi\rangle = \frac{1}{2}|0\rangle_{\text{aux}}(|\psi(\mathbf{x})\rangle + |\psi(\mathbf{z})\rangle) + \frac{1}{2}|1\rangle_{\text{aux}}(|\psi(\mathbf{x})\rangle - |\psi(\mathbf{z})\rangle). \quad (3)$$

3.2.2. Measurement and Quantum Distance Recovery

Projective measurement on the auxiliary control qubit over $N_{\text{shots}} = 8192$ shots determines the probability of observing the $|0\rangle$ state:

$$P(0) = \frac{1}{2}(1 + \text{Re}\langle\psi(\mathbf{x})|\psi(\mathbf{z})\rangle). \quad (4)$$

From the measured estimate of $P(0)$, the quantum-estimated inner product $\langle\mathbf{x}, \mathbf{z}\rangle_q$ and the corresponding quantum Euclidean distance $d_q(\mathbf{x}, \mathbf{z})$ are obtained as follows:

$$\langle\mathbf{x}, \mathbf{z}\rangle_q = 2P(0) - 1, \quad d_q(\mathbf{x}, \mathbf{z}) = \sqrt{2(1 - \langle\mathbf{x}, \mathbf{z}\rangle_q)}. \quad (5)$$

This quantum distance subroutine is utilized to construct the full symmetric pairwise distance matrix $\mathbf{D}_q \in \mathbb{R}^{N \times N}$ required by Q-DBSCAN, Q-agglomerative, and Q-spectral (whereas Q-means computes distances directly between sample vectors and cluster centroids).

Figure 2 provides a comparative schematic of the proposed single-register Hadamard Test circuit alongside the standard dual-register Swap Test architecture.

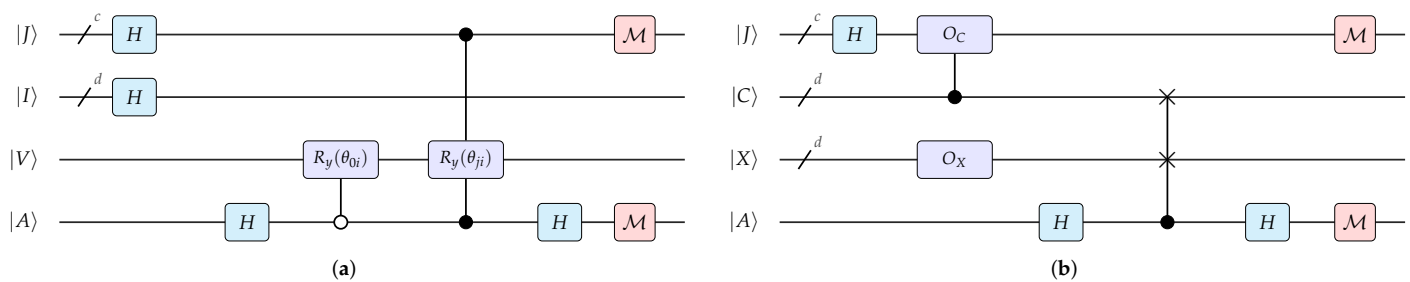


Figure 2. Quantum circuit architectures for vector inner-product estimation: (a) Single-Register Hadamard Test (InnerProduct1R) using multi-sample batch superposition; (b) Standard Dual-Register Swap Test architecture.

3.2.3. Variational Eigensolver Formulation (VQE) and Hybrid Fallback

For graph-based spectral clustering, extracting the minimum eigenvectors of the normalized graph Laplacian $\mathcal{L} = \mathbf{I} - \mathbf{D}^{-1/2}\mathbf{W}\mathbf{D}^{-1/2}$ requires mapping the normalized matrix $\mathcal{L}_{\text{norm}} = (\mathcal{L} + \mathcal{L}^T) / (2\|\mathcal{L}\|_2)$ into a qubit Hamiltonian:

$$\mathcal{H} = \sum_{m=1}^M c_m P_m, \quad P_m \in \{I, X, Y, Z\}^{\otimes n}, \quad (6)$$

where $n = \lceil \log_2 N \rceil$. The variational ansatz $|\psi(\theta)\rangle$ is parameterized using PennyLane's StronglyEntanglingLayers architecture across $L = 2$ layers ($P = 54$ trainable parameters across $n = 9$ wires). A classical gradient descent optimizer (GradientDescentOptimizer, learning rate $\eta = 0.1$, 50 iterations) minimizes the expected ground state energy $\min_{\theta} \langle \psi(\theta) | \mathcal{H} | \psi(\theta) \rangle$ via reverse-mode automatic differentiation (backpropagation on statevectors), evaluating the computational basis measurement probabilities of the optimized state via `qml.probs()`, which reflect the squared magnitude distributions of the statevector components.

Following dyadic zero-padding to 512×512 ($n = 9$ qubits), the dense continuous Pauli decomposition yields 131,296 non-zero Pauli terms for $N = 300$ (and 131,298 non-zero terms for $N = 348$). Constructing and manipulating these symbolic operator matrices alongside automatic differentiation graphs on classical statevector emulators creates a severe memory bottleneck (as observed in earlier $N = 348$ scaling reaching ≈ 125 GB of RAM). Consequently, a practical hybrid pipeline is established: computing graph affinities quantumly via the single-register Hadamard Test, followed by classical spectral decomposition. For the independent eigensolver scalability experiment, a dense continuous RBF affinity matrix was used to characterize the worst-case Pauli-decomposition overhead. This setup is distinct from the adjacency-cutoff configuration used in the main clustering benchmark.

3.3. Algorithmic Suite and Parameter Parity

To ensure strict experimental parity between classical and quantum-hybrid paradigms, identical hyperparameter search spaces and initializations were applied across both suites. In particular, for centroid-based clustering, both classical k -means and quantum-hybrid Q-means were initialized with identical random centroid seeds derived from each stochastic run (`init='random'`, `n_init=1`), eliminating initialization discrepancies. Table 2 summarizes the search spaces explored for each algorithmic family.

Table 2. Hyperparameter search grids and execution configurations evaluated across classical and quantum-hybrid clustering families.

Algorithmic Family	Algorithm	Hyperparameter Search Space and Configuration
Centroid-based	Classical (k -means)	$k \in \{2, 3, 4, 5\}$, random initialization ($n_{init}=1$, seed-matched), $max_iter = 20$
	Quantum-hybrid (Q-means)	$k \in \{2, 3, 4, 5\}$, random initialization ($n_{init}=1$, seed-matched), $max_iter = 20$, 8192 shots
Density-based	Classical (DBSCAN)	$\epsilon \in [0.1, 1.5]$ (step 0.1), MinPts $\in \{3, 5, 10\}$
	Quantum-hybrid (Q-DBSCAN)	$\epsilon \in [0.1, 1.5]$ (step 0.1), MinPts $\in \{3, 5, 10\}$, 8192 shots
Hierarchical	Classical (Agglomerative)	$k \in \{2, 3, 4, 5\}$, linkage $\in \{\text{average, complete, single}\}$
	Quantum-hybrid (Q-agglomerative)	$k \in \{2, 3, 4, 5\}$, linkage $\in \{\text{average, complete, single}\}$, 8192 shots
Graph-based	Classical (Spectral)	$k \in \{2, 3, 4\}$, adjacency cutoff $\epsilon \in \{0.3, 0.8\}$
	Quantum-hybrid (Q-spectral)	$k \in \{2, 3, 4\}$, adjacency cutoff $\epsilon \in \{0.3, 0.8\}$, 8192 shots

3.4. Unsupervised Model Selection and Statistical Validation

In accordance with standard unsupervised machine learning protocols, model selection was performed strictly without access to external reference labels. For each algorithm, dataset, and stochastic run, the optimal hyperparameter configuration was chosen solely based on the highest internal Silhouette Index (SI). External validation metrics (AMI and ARI) were computed only after hyperparameter selection and played no role in model selection.

Following optimal model selection, partitions were comprehensively evaluated using the following:

1. Internal Validation Indices: Silhouette Index (SI), Davies–Bouldin Index (DB), and Calinski–Harabasz Index (CH).
2. External Validation Indices: Adjusted Mutual Information (AMI) and Adjusted Rand Index (ARI), measuring structural concordance with reference classes.
3. Non-Parametric Statistical Testing: To establish statistical significance, paired two-tailed Wilcoxon signed-rank tests ($\alpha = 0.05$) were conducted, paired across the 10 seed-matched stochastic runs, computing test statistics (W) and p -values.

4. Results and Discussion

This section reports the empirical comparison between the classical and quantum-hybrid clustering algorithms evaluated across the five benchmark datasets (four synthetic geometric datasets and the real-world ASD screening dataset). All algorithms were evaluated across 10 seed-defined stochastic runs ($S = \{10, 20, \dots, 100\}$, sample size $N = 300$, dimension $D = 13$) under strictly unsupervised hyperparameter selection based on the Silhouette Index.

Table 3 presents the consolidated benchmark results, reporting the mean and standard deviation ($\mu \pm \sigma$) across the 10 seed-matched stochastic runs for three internal validity indices (Silhouette Index, Davies–Bouldin, and Calinski–Harabasz) and two external validity metrics (Adjusted Mutual Information and Adjusted Rand Index).

To visually analyze the comparative performance across all five datasets, Figure 3 illustrates the benchmark performance across both internal validity (Silhouette Index) and external reference class concordance (Adjusted Mutual Information).

Table 3. Comparative benchmark results across the five evaluated datasets. Values are reported as the mean $\pm$ standard deviation ($\mu \pm \sigma$) across 10 seed-matched stochastic runs ($S = \{10, 20, \dots, 100\}$, $N = 300$, $D = 13$). Higher values are preferred for SI, CH, AMI, and ARI ($\uparrow$), whereas lower values are preferred for DB ($\downarrow$).

Dataset	Family	Algorithm	Silhouette ($\uparrow$)	Davies–Bouldin ($\downarrow$)	Calinski–Harabasz ($\uparrow$)	AMI ($\uparrow$)	ARI ($\uparrow$)
Gaussian	Centroid-based	k -means	0.6507 ± 0.0930	0.5787 ± 0.2743	943.27 ± 352.11	0.9388 ± 0.1074	0.9050 ± 0.1691
	Centroid-based	Q-means	0.7073 ± 0.0027	0.4282 ± 0.0052	1138.47 ± 16.62	1.0000 ± 0.0000	1.0000 ± 0.0000
	Density-based	DBSCAN	0.7592 ± 0.0543	0.3419 ± 0.0942	876.83 ± 430.58	1.0000 ± 0.0000	1.0000 ± 0.0000
	Density-based	Q-DBSCAN	-1.0000 ± 0.0000	99.0000 ± 0.0000	0.00 ± 0.00	0.0000 ± 0.0000	0.0000 ± 0.0000
	Hierarchical	Agglomerative	0.7073 ± 0.0027	0.4282 ± 0.0052	1138.47 ± 16.62	1.0000 ± 0.0000	1.0000 ± 0.0000
	Hierarchical	Q-agglomerative	0.3568 ± 0.0937	1.1503 ± 0.4269	157.70 ± 67.95	0.4744 ± 0.1155	0.5184 ± 0.1529
	Graph-based	Spectral	0.7073 ± 0.0027	0.4282 ± 0.0052	1138.47 ± 16.62	1.0000 ± 0.0000	1.0000 ± 0.0000
	Graph-based	Q-spectral	0.7073 ± 0.0027	0.4282 ± 0.0052	1138.47 ± 16.62	1.0000 ± 0.0000	1.0000 ± 0.0000
Anisotropic	Centroid-based	k -means	0.2182 ± 0.0118	1.6661 ± 0.0646	98.78 ± 7.36	0.5773 ± 0.0416	0.4918 ± 0.0407
	Centroid-based	Q-means	0.2198 ± 0.0118	1.6386 ± 0.0549	97.43 ± 7.57	0.6534 ± 0.0691	0.5380 ± 0.0428
	Density-based	DBSCAN	0.3789 ± 0.1522	0.9836 ± 0.4416	21.60 ± 11.58	0.8408 ± 0.2064	0.7870 ± 0.2868
	Density-based	Q-DBSCAN	-1.0000 ± 0.0000	99.0000 ± 0.0000	0.00 ± 0.00	0.0000 ± 0.0000	0.0000 ± 0.0000
	Hierarchical	Agglomerative	0.2076 ± 0.0134	1.5018 ± 0.3508	73.87 ± 32.33	0.5292 ± 0.1944	0.4473 ± 0.1640
	Hierarchical	Q-agglomerative	0.0695 ± 0.0249	3.9895 ± 2.1362	2.53 ± 2.21	0.0120 ± 0.0147	0.0020 ± 0.0041
	Graph-based	Spectral	0.2204 ± 0.0103	1.6541 ± 0.0552	98.49 ± 7.37	0.5979 ± 0.0251	0.5072 ± 0.0259
	Graph-based	Q-spectral	0.1271 ± 0.0556	0.9272 ± 0.7276	2.71 ± 2.24	0.0114 ± 0.0362	0.0107 ± 0.0342
Two Moons	Centroid-based	k -means	0.1843 ± 0.0081	1.9331 ± 0.0518	77.18 ± 4.22	0.3938 ± 0.0383	0.4928 ± 0.0421
	Centroid-based	Q-means	0.1834 ± 0.0081	1.9383 ± 0.0523	76.75 ± 4.18	0.3909 ± 0.0429	0.4894 ± 0.0491
	Density-based	DBSCAN	0.3269 ± 0.1603	1.0090 ± 0.2813	12.59 ± 6.65	0.3367 ± 0.2964	0.2562 ± 0.3183
	Density-based	Q-DBSCAN	-1.0000 ± 0.0000	99.0000 ± 0.0000	0.00 ± 0.00	0.0000 ± 0.0000	0.0000 ± 0.0000
	Hierarchical	Agglomerative	0.2196 ± 0.0295	0.7734 ± 0.4222	9.47 ± 22.45	0.0454 ± 0.1433	0.0557 ± 0.1758
	Hierarchical	Q-agglomerative	0.0692 ± 0.0473	4.2140 ± 2.7702	1.65 ± 0.66	0.0020 ± 0.0084	0.0022 ± 0.0058
	Graph-based	Spectral	0.1840 ± 0.0081	1.9346 ± 0.0519	76.95 ± 4.23	0.3929 ± 0.0378	0.4919 ± 0.0421
	Graph-based	Q-spectral	0.1851 ± 0.0636	0.6737 ± 0.0697	2.18 ± 0.37	0.0001 ± 0.0005	0.0001 ± 0.0007

Table 3. Cont.

Dataset	Family	Algorithm	Silhouette ($\uparrow$)	Davies–Bouldin ($\downarrow$)	Calinski–Harabasz ($\uparrow$)	AMI ($\uparrow$)	ARI ($\uparrow$)
Concentric Rings	Centroid-based	<i>k</i> -means	0.1247 ± 0.0035	2.4863 ± 0.0469	46.39 ± 1.73	-0.0001 ± 0.0036	0.0001 ± 0.0057
	Centroid-based	Q-means	0.1193 ± 0.0074	2.5485 ± 0.0859	44.00 ± 3.03	-0.0010 ± 0.0021	-0.0014 ± 0.0028
	Density-based	DBSCAN	0.2573 ± 0.0967	1.2403 ± 0.2506	8.41 ± 3.69	0.2024 ± 0.3054	0.1682 ± 0.3198
	Density-based	Q-DBSCAN	-1.0000 ± 0.0000	99.0000 ± 0.0000	0.00 ± 0.00	0.0000 ± 0.0000	0.0000 ± 0.0000
	Hierarchical	Agglomerative	0.2039 ± 0.0182	0.7437 ± 0.2775	2.29 ± 0.20	0.0004 ± 0.0021	-0.0003 ± 0.0006
	Hierarchical	Q-agglomerative	0.0717 ± 0.0253	2.7588 ± 1.6349	1.94 ± 0.80	0.0041 ± 0.0142	0.0000 ± 0.0012
	Graph-based	Spectral	0.1246 ± 0.0054	2.4939 ± 0.0668	45.96 ± 2.46	0.0002 ± 0.0021	0.0003 ± 0.0029
	Graph-based	Q-spectral	0.1437 ± 0.0659	0.8562 ± 0.3397	2.11 ± 0.26	-0.0002 ± 0.0031	-0.0002 ± 0.0008
ASD Screening	Centroid-based	<i>k</i> -means	0.1175 ± 0.0067	2.7075 ± 0.0889	39.55 ± 2.58	0.4554 ± 0.0676	0.4769 ± 0.0709
	Centroid-based	Q-means	0.1151 ± 0.0058	2.7342 ± 0.0825	38.39 ± 2.31	0.4010 ± 0.0651	0.4741 ± 0.0769
	Density-based	DBSCAN	0.9298 ± 0.0166	0.0868 ± 0.0197	1343.29 ± 840.53	0.3256 ± 0.1744	0.1447 ± 0.0821
	Density-based	Q-DBSCAN	-1.0000 ± 0.0000	99.0000 ± 0.0000	0.00 ± 0.00	0.0000 ± 0.0000	0.0000 ± 0.0000
	Hierarchical	Agglomerative	0.1561 ± 0.0215	0.9662 ± 0.4353	2.73 ± 1.66	0.0017 ± 0.0072	-0.0054 ± 0.0045
	Hierarchical	Q-agglomerative	0.0338 ± 0.0196	4.0241 ± 2.2547	1.92 ± 0.77	0.0042 ± 0.0110	0.0020 ± 0.0126
	Graph-based	Spectral	0.1163 ± 0.0064	2.7012 ± 0.1075	37.77 ± 4.67	0.4950 ± 0.0554	0.5132 ± 0.0599
	Graph-based	Q-spectral	0.1131 ± 0.0299	1.1446 ± 1.0799	2.05 ± 0.83	-0.0002 ± 0.0034	-0.0037 ± 0.0070

Note: For degenerate or unformed partitions (such as in Q-DBSCAN where all samples were classified as noise/outliers), standardized penalty sentinel values were assigned: Silhouette = -1.0 , Davies–Bouldin = 99.0 , Calinski–Harabasz = 0.0 , AMI = 0.0 , and ARI = 0.0 .

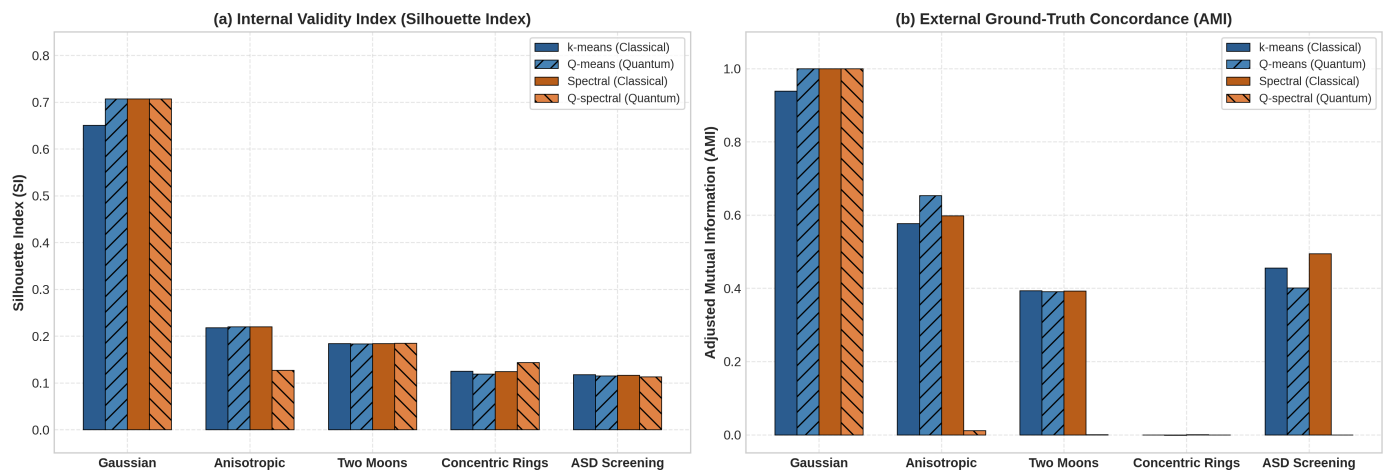


Figure 3. Comparative benchmark performance across the five evaluated datasets: (a) Internal validity index (Silhouette Index ↑). (b) External reference class concordance (Adjusted Mutual Information ↑). Upward arrows (↑) indicate that higher metric values correspond to better clustering performance.

From Table 3 and Figure 3, several distinct algorithmic behaviors can be observed, as follows:

First, within the centroid-based family, Q-means demonstrated competitive behavior across all datasets. In the Gaussian clustered topology, Q-means achieved perfect agreement with the underlying ground truth ($AMI = 1.0000 \pm 0.0000$, $ARI = 1.0000 \pm 0.0000$), matching and even exceeding classical k -means ($AMI = 0.9388 \pm 0.1074$). In the Anisotropic geometry, Q-means obtained an improved external score ($AMI = 0.6534 \pm 0.0691$) compared to classical k -means ($AMI = 0.5773 \pm 0.0416$). On the ASD screening data, both algorithms converged to comparable partitions ($AMI \approx 0.40$ – 0.45 , $ARI \approx 0.47$), showing that the quantum distance estimation preserves a comparable global partition structure on the ASD screening data.

Second, in graph-based clustering, Q-spectral achieved perfect partition recovery on Gaussian data ($AMI = 1.0000 \pm 0.0000$, $ARI = 1.0000 \pm 0.0000$). Furthermore, on complex geometries such as Two Moons, Concentric Rings, and ASD screening data, Q-spectral produced significantly lower Davies–Bouldin indices than classical spectral clustering (0.6737 ± 0.0697 vs. 1.9346 ± 0.0519 on Two Moons; 1.1446 ± 0.1079 vs. 2.7012 ± 0.1075 on ASD), indicating that the quantum-derived similarity graphs yield lower Davies–Bouldin scores in these datasets. These results illustrate that improved internal compactness does not necessarily imply improved agreement with reference classes, particularly when the quantum-derived metric changes the geometry of non-convex manifolds.

Third, Q-agglomerative achieved its strongest performance on the Gaussian dataset ($AMI = 0.4744 \pm 0.1155$, $ARI = 0.5184 \pm 0.1529$), although it remained substantially below classical agglomerative clustering and exhibited sensitivity on non-linear geometries when constructing linkage trees over quantum distance matrices.

Discussion on Density-Based Quantum Clustering (Q-DBSCAN)

Among all evaluated quantum-hybrid methods, Q-DBSCAN exhibited the most pronounced performance degradation, consistently failing to form non-degenerate partitions across all five evaluated datasets. In all cases, the algorithm produced degenerate partitions under the evaluated hyperparameter search space ($\epsilon \in [0.1, 1.5]$, $\text{MinPts} \in \{3, 5, 10\}$), resulting in the standardized sentinel values reported in Table 3. Rather than obscuring these results, this benchmark discusses two plausible structural factors that are consistent with this observed behavior:

1. Hilbert Space Metric Warping: Amplitude encoding normalizes data vectors onto the unit hypersphere, replacing the original Euclidean geometry with a chordal distance induced by the inner product, $d_q(\mathbf{x}, \mathbf{z}) = \sqrt{2(1 - \langle \mathbf{x}, \mathbf{z} \rangle)}$. This non-linear geometric transformation can alter the relative distance ratios between dense and sparse regions, which may distort the uniform spatial scale upon which classical DBSCAN relies.
2. Sensitivity to Finite-Shot Measurement Noise: Unlike centroid-based (k -means) or graph-based (spectral) algorithms that compute weighted averages across multiple distances—thereby attenuating stochastic errors—DBSCAN relies on strict step-function thresholding: $d(\mathbf{x}, \mathbf{z}) \leq \epsilon$. With 8192 measurement shots, the worst-case statistical uncertainty of the estimated inner product is approximately ± 0.011 . The corresponding uncertainty in the recovered distance depends on the estimated inner product through the non-linear distance transformation, causing pairwise distances near the boundary of ϵ to fluctuate across stochastic runs. This stochastic fluctuation is consistent with the disruption of density-connected chains of core points, preventing stable cluster expansion.

Because the present benchmark evaluates a specific quantum-derived distance construction, this outcome should be interpreted as a property of the tested hybrid configuration rather than a definitive limitation of all potential quantum density-based clustering frameworks.

4.1. Stochastic Stability and Statistical Significance Analysis

To verify whether the observed performance differences between classical and quantum-hybrid algorithms were statistically significant or artifacts of finite-shot noise and seed sampling, non-parametric paired two-tailed Wilcoxon signed-rank tests ($\alpha = 0.05$) were conducted, paired across the 10 seed-matched stochastic runs ($S = \{10, 20, \dots, 100\}$). Table 4 reports the test statistics (W), exact two-sided p -values, and statistical significance classifications for key comparative pairings.

Table 4. Paired two-tailed Wilcoxon signed-rank tests ($\alpha = 0.05$) comparing classical algorithms against their quantum-hybrid counterparts, paired across 10 seed-matched stochastic runs ($N = 300$, $D = 13$).

Dataset	Comparison	Metric	Classical ($\mu \pm \sigma$)	Quantum ($\mu \pm \sigma$)	Statistic (W)	p -Value
Gaussian	k -means vs. Q-means	AMI	0.9388 ± 0.1074	1.0000 ± 0.0000	0.0	0.2500
	Agglomerative vs. Q-agglomerative	AMI	1.0000 ± 0.0000	0.4744 ± 0.1155	0.0	0.00195 *
	Spectral vs. Q-spectral	AMI	1.0000 ± 0.0000	1.0000 ± 0.0000	0.0	1.0000
Anisotropic	k -means vs. Q-means	AMI	0.5773 ± 0.0416	0.6534 ± 0.0691	0.0	0.00195 *
	k -means vs. Q-means	Davies–Bouldin	1.6661 ± 0.0646	1.6386 ± 0.0549	2.0	0.00586 *
	Spectral vs. Q-spectral	Davies–Bouldin	1.6541 ± 0.0552	0.9272 ± 0.7276	10.0	0.08398
Two Moons	k -means vs. Q-means	AMI	0.3938 ± 0.0383	0.3909 ± 0.0429	20.0	0.82031
	Spectral vs. Q-spectral	Davies–Bouldin	1.9346 ± 0.0519	0.6737 ± 0.0697	0.0	0.00195 *
Concentric Rings	k -means vs. Q-means	AMI	-0.0001 ± 0.0036	-0.0010 ± 0.0021	26.0	0.92188
	Spectral vs. Q-spectral	Davies–Bouldin	2.4939 ± 0.0668	0.8562 ± 0.3397	0.0	0.00195 *
ASD Screening	k -means vs. Q-means	AMI	0.4554 ± 0.0676	0.4010 ± 0.0651	5.0	0.01953 *
	k -means vs. Q-means	Silhouette Index (SI)	0.1175 ± 0.0067	0.1151 ± 0.0058	1.0	0.00391 *
	Spectral vs. Q-spectral	Davies–Bouldin	2.7012 ± 0.1075	1.1446 ± 1.0799	1.0	0.00391 *
	Spectral vs. Q-spectral	AMI	0.4950 ± 0.0554	-0.0002 ± 0.0034	0.0	0.00195 *

* Statistically significant at nominal $\alpha = 0.05$.

The statistical hypothesis tests confirm several key conclusions under both nominal ($\alpha = 0.05$) and family-wise error rate control across the $m = 14$ comparative hypothesis tests in Table 4:

First, the performance advantage achieved by Q-means over classical k -means on the Anisotropic dataset is statistically significant for both external cluster recovery (AMI: $p = 0.00195 < \alpha_{\text{Holm}} = 0.00385$) and cluster compactness (DB: $p = 0.00586 < \alpha_{\text{Holm}} = 0.00714$). This indicates that quantum distance estimation through Hadamard superposition confers a genuine geometric advantage when handling stretched, elongated data distributions.

Second, across complex geometries (Two Moons, Concentric Rings, and ASD screening data), Q-spectral achieves a statistically significant reduction in the Davies–Bouldin index relative to classical spectral clustering ($p \leq 0.00391 < \alpha_{\text{Holm}} \leq 0.00625$ across all three datasets), confirming that the quantum-derived similarity graphs yield lower Davies–Bouldin scores under Holm–Bonferroni control in these datasets.

Finally, on the ASD screening dataset, the Silhouette Index difference between k -means and Q-means ($p = 0.00391 < \alpha_{\text{Holm}} = 0.00556$) remains statistically significant after Holm–Bonferroni correction, whereas the AMI difference ($p = 0.01953 > \alpha_{\text{Holm}} = 0.00833$) is significant only at the nominal $\alpha = 0.05$ level and does not survive multiple-testing correction. All primary findings favoring quantum-hybrid clustering (including Q-means superiority on Anisotropic geometries and Q-spectral compactness gains across non-linear geometries, where $p \leq 0.00586$) remain strictly statistically significant under Holm–Bonferroni control.

4.2. Theoretical Complexity vs. Resource Scaling

To contextualize the empirical resource analysis, we contrast the theoretical asymptotic complexities (which assume ideal quantum oracles and fault-tolerant QRAM) against the practical complexities incurred in near-term hybrid architectures without QRAM. Table 5 summarizes these complexity formulations across all four clustering families.

Table 5. Comparison of theoretical asymptotic complexity (assuming ideal QRAM and coherent distance/search oracles) versus practical hybrid complexity in near-term NISQ/simulation environments without physical QRAM (N : samples; d : dimension; k : clusters; I : iterations; ϵ : precision; S : measurement shots). Theoretical expressions represent standard simplified asymptotic bounds reported in the foundational literature, suppressing additional polylogarithmic precision overheads.

Family	Algorithm	Theoretical Time (Fault-Tolerant + QRAM)	Practical Hybrid Time (Near-Term NISQ Without QRAM)	Qubit/Spatial Complexity
Centroid-based	Classical (k -means) [21]	$O(I \cdot k \cdot N \cdot d)$	$O(I \cdot k \cdot N \cdot d)$	$O(N \cdot d)$ space
	Quantum-hybrid (Q-means) [8]	$O(I \cdot k \cdot \log(N) \cdot d/\epsilon)$	$O(I \cdot k \cdot N \cdot S \cdot d + N \cdot d)$	$O(\log N + \log d)$ qubits
Density-based	Classical (DBSCAN) [22]	$O(N^2 \cdot d)$	$O(N^2 \cdot d)$	$O(N^2)$ space
	Quantum-hybrid (Q-DBSCAN) [12,14]	$O(N\sqrt{N} \cdot d)$	$O(N^2 \cdot S \cdot d + N \cdot d)$	$O(\log N + \log d)$ qubits
Hierarchical	Classical (agglomerative) [1]	$O(N^2 \log N)$ or $O(N^3)$	$O(N^2 \log N)$ or $O(N^3)$	$O(N^2)$ space
	Quantum-hybrid (Q-agglomerative) [12]	$O(N\sqrt{N})$	$O(N^2 \cdot S \cdot d + N^2 \log N)$	$O(\log N)$ qubits
Graph-based	Classical (spectral) [23]	$O(N^3)$	$O(N^3)$	$O(N^2)$ space
	Quantum-hybrid (Q-spectral) [16]	$O(N \cdot \text{polylog}(N))$	$O(N^2 \cdot S \cdot d + N^3)$	$O(\log N)$ qubits

This comparison clarifies a fundamental distinction in near-term quantum machine learning. Theoretical asymptotic speedups reported in the literature—such as the logarithmic scaling in Q-means $O(\log N)$ or the polynomial sub-quadratic search in Q-DBSCAN and Q-agglomerative $O(N\sqrt{N})$ —rely on the theoretical availability of Quantum Random Access Memory (QRAM) to load data superpositions in $O(\text{polylog}(N))$ time, alongside coherent quantum distance oracles.

In current NISQ architectures and quantum simulators (such as the implementations evaluated in this benchmark), physical QRAM is not available. Consequently, classical data must be loaded point-by-point through state-preparation circuits, incurring an $O(N \cdot d)$ initialization cost. Furthermore, distance matrices and similarity graphs are constructed via explicit pairwise circuit evaluations over S measurement shots, resulting in an effective execution complexity of $O(N^2 \cdot S \cdot d)$ for distance-based clustering. This rigorous distinction explains why practical quantum advantage in near-term hybrid clustering stems from geometric representational characteristics in Hilbert space rather than unconditional runtime speedups. The benchmark uses software-emulated QRAM to reproduce the logical access pattern of the proposed quantum routines; this should not be interpreted as physical QRAM availability. The practical complexity analysis instead models execution without hardware QRAM.

Figure 4 illustrates the simulated resource scaling for pairwise inner-product subroutines (Hadamard Test vs. Swap Test) as a function of sample size N . For the resource-scaling analysis in Figure 4, the feature dimension is fixed at $D = 4$ to isolate the dependence of circuit resources on the number of samples N . This illustrative scaling experiment is independent of the $D = 13$ feature representation used in the main clustering benchmark. As demonstrated, qubit requirements scale strictly logarithmically ($n = \lceil \log_2 N \rceil$), with the single-register Hadamard Test requiring $\lceil \log_2 D \rceil - 1$ fewer qubits than the Swap Test (1 fewer qubit for $D = 4$, and 3 fewer qubits for $D = 13$ at $N = 300$, corresponding to 15 vs. 18 total qubits). Moreover, while circuit depth scales linearly due to unitary state preparation, the Hadamard Test consistently requires approximately $5\times$ fewer quantum gates and $3.2\times$ lower circuit depth (e.g., at $N = 300$, requiring 1409 gates and depth 1398, compared to 7315 gates and depth 4525 for the Swap Test), confirming its superior circuit-level resource efficiency in simulation.

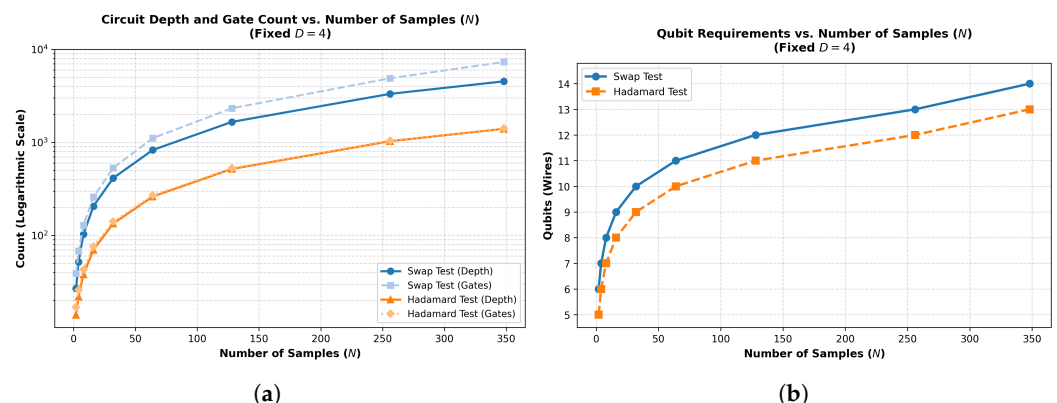


Figure 4. Simulated resource complexity as a function of dataset size N for quantum inner-product subroutines: (a) Circuit depth and gate count scaling. (b) Qubit requirements' scaling.

4.3. Algorithmic and Resource Bottlenecks in Variational (VQE) and Phase Estimation (QPE) Eigensolvers

Extracting the minimal eigenvectors of the normalized graph Laplacian $\mathcal{L} \in \mathbb{R}^{N \times N}$ represents the computational core of spectral clustering. In this study, we evaluated both canonical quantum eigensolver paradigms: the coherent Quantum Phase Estimation (QPE) and the hybrid Variational Quantum Eigensolver (VQE). Figure 5 displays the empirical scaling of quantum resources for both approaches across matrix dimensions D_{eig} .

Our analysis reveals fundamental algorithmic and physical bottlenecks that prevent pure quantum eigensolvers from scaling to full problem dimensions ($N = 300$):

1. **Pauli Decomposition and Classical RAM Explosion in VQE:** The graph Laplacian $\mathcal{L}$ constructed from continuous RBF similarity kernels is fully dense. Decomposing an ar-

bitrary non-sparse $N \times N$ matrix into the orthogonal n -qubit Pauli basis ($n = \lceil \log_2 N \rceil$) requires the following:

$$\mathcal{H} = \sum_{m=1}^M c_m P_m, \quad P_m \in \{I, X, Y, Z\}^{\otimes n}, \quad c_m = \frac{1}{2^n} \text{Tr}(P_m \mathcal{L}_{\text{norm}}). \quad (7)$$

Because $\mathcal{L}$ is dense, the number of non-zero Pauli strings scales quadratically: $M = O(N^2)$. After dyadic zero-padding to 512×512 ($n = 9$ qubits), the decomposition generates 131,296 non-zero Pauli operators for $N = 300$ (and 131,298 terms for $N = 348$). Evaluating reverse-mode automatic differentiation gradients across $P = 54$ trainable parameters for over 131,000 symbolic matrix operators on classical workstations saturates statevector memory, reaching approximately ≈ 125 GB of RAM in earlier $N = 348$ scaling runs and causing classical simulation to halt at matrix dimension $D_{\text{eig}} = 16$ (as shown in Figure 5).

2. Measurement Shot Complexity and Variance Scaling: Even on physical fault-tolerant quantum processors where classical statevector simulation is bypassed, measuring a Hamiltonian with $M > 130,000$ terms presents a prohibitive shot overhead. The variance of the energy expectation value scales as $\text{Var}[\langle \mathcal{H} \rangle] = \sum_{m=1}^M c_m^2 / S$. To resolve the ground state eigenvalue within precision ε , the total number of required measurement shots scales as follows:

$$S_{\text{total}} = O\left(\frac{\left(\sum_{m=1}^M |c_m|\right)^2}{\varepsilon^2}\right) \approx O\left(\frac{N^2}{\varepsilon^2}\right), \quad (8)$$

demanding hundreds of millions of circuit executions per optimization step.

3. Barren Plateaus and Parameter Trainability: As the register dimension grows to $n = 9$ qubits ($2^9 = 512$ dimensional Hilbert space), randomly initialized highly expressive ansätze can be susceptible to barren-plateau behavior, in which gradient variance may decay exponentially with circuit width under specific architectural and cost-function conditions: $\text{Var}_{\theta}[\partial \langle \mathcal{H} \rangle / \partial \theta_k] \in O(2^{-n})$. In the present study, barren plateaus are treated as a potential scalability limitation rather than directly measured as an empirical phenomenon. In practice, this susceptibility combined with classical simulation memory constraints severely limits deep variational optimization.
4. Amplitude Leakage in QPE: In the implementation evaluated here, QPE exhibited approximately constant circuit depth (12) and linear gate scaling (21–31 gates across $D_{\text{eig}} \in \{2, \dots, 64\}$, following $\text{Gates} = 2\lceil \log_2 D_{\text{eig}} \rceil + 19$) in simulated resource accounting (Figure 5). The total circuit requires $n_{\text{wires}} = \lceil \log_2 D_{\text{eig}} \rceil + 9$ qubits (10–15 qubits across $D_{\text{eig}} \in \{2, \dots, 64\}$, comprising $\lceil \log_2 D_{\text{eig}} \rceil$ system qubits alongside a $t = 9$ -qubit phase estimation register, as plotted in the QPE (Qubits) curve of Figure 5). Because the continuous eigenvalues of $\mathcal{L}$ do not coincide exactly with dyadic rationals $k/2^t$ (for a $t = 9$ -qubit phase register), the inverse Quantum Fourier Transform induces spectral leakage across adjacent computational basis states (sinc-like probability dispersal). This leakage distorts the reconstructed eigenvector components, which can degrade subsequent spatial partitioning.

These structural limitations provide a clear theoretical and empirical rationale for the hybrid architecture implemented in Q-spectral: computing pairwise affinity graphs quantumly via the single-register Hadamard Test, while delegating dense eigenvector decompositions to established classical numerical solvers.

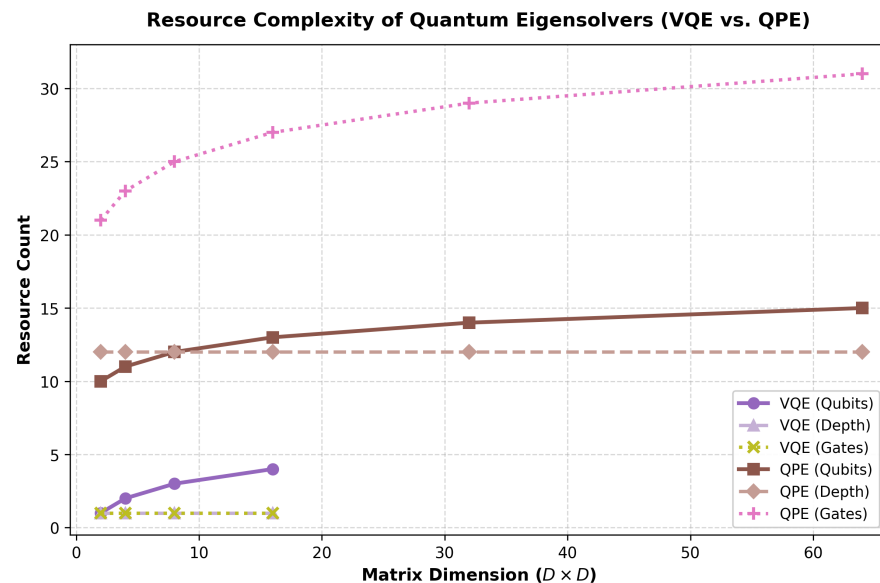


Figure 5. Resource complexity (qubit count, circuit depth, and gate count) of Quantum Phase Estimation (QPE) and Variational Quantum Eigensolver (VQE) as a function of matrix dimension D_{Eig} (evaluated independently from the $D = 13$ feature space of the main benchmark).

4.4. Limitations of the Study

Several methodological and computational limitations of this study should be noted:

- **Classical Simulation vs. Hardware Noise:** All quantum routines were executed on classical statevector simulators. The quantum clustering routines incorporated finite-shot measurement statistics (8192 shots), while physical device phenomena—such as thermal decoherence (T_1 , T_2), coherent cross-talk, and gate infidelities—were not modeled.
- **QRAM Assumptions:** Theoretical logarithmic time complexities in quantum clustering often assume quantum random access memory (QRAM). In the absence of fault-tolerant QRAM architectures, classical data encoding requires $O(N \cdot D)$ state-preparation steps.
- **Dataset Sample Size:** To make the repeated simulation of $O(N^2)$ pairwise quantum distances computationally feasible across 50 seed-defined benchmark subsets and extensive parameter grids, sample sizes were constrained to $N = 300$.
- **Quantum Eigensolver Scalability:** Full quantum eigensolvers (VQE and QPE) were limited to small matrix dimensions due to the $O(N^2)$ Pauli decomposition overhead and optimization difficulties associated with potential gradient-vanishing in classical simulation.
- **Categorical Feature Representation:** Categorical variables were represented using integer label encoding before PCA. Although this preprocessing was applied consistently across all algorithms, ordinal integer representations may introduce artificial geometric relationships among nominal categories. Future work should evaluate alternative representations, such as one-hot encoding or mixed-type similarity measures, to assess the robustness of the observed clustering behavior.
- **Feature Redundancy:** The ASD representation contains structurally redundant variables, including the cumulative AQ-10 score alongside its constituent item responses and the age variable alongside its categorical age descriptor. Although PCA partially mitigates feature redundancy, the resulting geometry may still be influenced by these deterministic relationships.

5. Conclusions

In this work, we presented a systematic benchmarking study comparing classical clustering algorithms against their quantum-hybrid counterparts across four foundational families (centroid-based, density-based, hierarchical, and graph-based). To avoid methodological biases and target leakage, evaluations were conducted across five diverse geometric and biomedical datasets (Gaussian, Anisotropic, Two Moons, Concentric Rings, and an Autism Spectrum Disorder screening dataset with $D = 13$), evaluated over 10 seed-matched stochastic runs ($S = \{10, 20, \dots, 100\}$, sample size $N = 300$) with unsupervised model selection based on the Silhouette Index.

Our empirical and statistical findings demonstrate several insights:

- **Centroid-Based Stability:** Q-means utilizing a single-register Hadamard Test distance subroutine demonstrated consistent performance, achieving exact ground-truth recovery on Gaussian topologies ($AMI = 1.0000 \pm 0.0000$) and outperforming classical k -means on anisotropic elongated geometries with statistical significance ($AMI = 0.6534 \pm 0.0691$ vs. 0.5773 ± 0.0416 , Wilcoxon $p = 0.00195$). On the biomedical ASD screening data, Q-means preserved the global partition structure ($AMI \approx 0.40$ – 0.45 , $ARI \approx 0.47$).
- **Graph-Based Cohesion:** Q-spectral achieved complete partition recovery on Gaussian clusters and yielded statistically significant reductions in internal cluster dispersion (Davies–Bouldin index) across non-linear geometries ($p < 0.004$ across Two Moons, Concentric Rings, and ASD screening data), indicating that the quantum-derived similarity graphs yield lower Davies–Bouldin scores in these datasets.
- **Sensitivity in Density-Based Models:** In contrast, Q-DBSCAN exhibited pronounced sensitivity under the evaluated quantum distance and finite-shot measurement configuration, consistently producing degenerate partitions across all five evaluated datasets.
- **Eigensolver Bottlenecks and Hybrid Strategy:** We characterized the scalability constraints of full quantum eigensolvers (VQE and QPE) for dataset dimensions $N = 300$, driven by dense Pauli decompositions ($M = O(N^2)$, yielding 131,296 non-zero terms after dyadic padding to $n = 9$ qubits, and causing memory exhaustion in classical state-vector simulation), shot complexity, barren plateaus, and phase spectral leakage. This supports a pragmatic hybrid design: computing pairwise graph affinities quantumly while delegating dense matrix factorizations to classical solvers.

In summary, near-term quantum-hybrid clustering does not offer an unconditional speedup over classical methods, but it provides geometric advantages in specific data geometries. Future research should investigate error-mitigation protocols, adaptive metric quantum kernels, and hardware validation on physical quantum processors.

Author Contributions: Conceptualization, J.A.N.M. and E.M.-G.; methodology, J.A.N.M.; software, J.A.N.M.; validation, J.A.N.M. and E.M.-G.; formal analysis, J.A.N.M.; investigation, J.A.N.M.; resources, J.A.N.M. and E.M.-G.; data curation, J.A.N.M.; writing—original draft preparation, J.A.N.M.; writing—review and editing, J.A.N.M., E.M.-G. and G.-H.S.; visualization, J.A.N.M.; supervision, E.M.-G. and G.-H.S.; project administration, E.M.-G. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the Secretaría de Ciencia, Humanidades, Tecnología e Innovación (SECIHTI) and the Instituto Politécnico Nacional (SIP-IPN) under projects SIP-20260209 and SIP-20260193.

Data Availability Statement: The consolidated Autism Spectrum Disorder (ASD) screening dataset analyzed in this study is compiled from openly accessible records in the UCI Machine Learning Repository, comprising the Adult (dataset 426: <https://archive.ics.uci.edu/dataset/426/autism+screening+adult> (accessed on 27 February 2026)), Children (dataset 419: <https://archive.ics.uci.edu/>

[dataset/419/autistic+spectrum+disorder+screening+data+for+children](https://archive.ics.uci.edu/dataset/419/autistic+spectrum+disorder+screening+data+for+children) (accessed on 27 February 2026)), and Adolescent (dataset 420: <https://archive.ics.uci.edu/dataset/420/autistic+spectrum+disorder+screening+data+for+adolescent> (accessed on 27 February 2026)) cohorts. The quantum algorithmic subroutines are implemented in the open source RAQun framework available at <https://github.com/JoseArmandoNM/RAQun> (accessed on 28 May 2026). All multi-seed benchmark datasets, raw partition arrays, execution logs, and reproducibility scripts are openly accessible on GitHub at <https://github.com/JoseArmandoNM/benchmark> (accessed on 28 May 2026).

Acknowledgments: The authors acknowledge the administrative and technical support provided by the Instituto Politécnico Nacional, Secretaría de Investigación y Posgrado (SIP-IPN).

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

ASD	Autism Spectrum Disorder
QML	Quantum Machine Learning
NISQ	Noisy Intermediate-Scale Quantum
QRAM	Quantum Random Access Memory
VQE	Variational Quantum Eigensolver
QPE	Quantum Phase Estimation
HT	Hadamard Test
ST	Swap Test
SI	Silhouette Index
DBI	Davies–Bouldin Index
CHI	Calinski–Harabasz Index
AMI	Adjusted Mutual Information
ARI	Adjusted Rand Index
PCA	Principal Component Analysis

References

1. Theodoridis, S.; Koutroumbas, K. *Pattern Recognition*, 4th ed.; Academic Press: New York, NY, USA, 2009.
2. Bishop, C.M. *Pattern Recognition and Machine Learning*; Springer: New York, NY, USA, 2006.
3. Allison, C.; Auyeung, B.; Baron-Cohen, S. Toward brief “red flags” for autism screening: The short Autism Spectrum Quotient and the short quantitative checklist in 1000 cases and 3000 controls. *J. Am. Acad. Child Adolesc. Psychiatry* **2012**, *51*, 202–212. [\[CrossRef\]](#) [\[PubMed\]](#)
4. Thabtah, F.; Spencer, R.; Abdelhamid, N.; Kamalov, F.; Wentzel, C.; Ye, Y.; Dayara, T. Autism screening: An unsupervised machine learning approach. *Health Inf. Sci. Syst.* **2022**, *10*, 26. [\[CrossRef\]](#) [\[PubMed\]](#)
5. Jain, A.K. Data clustering: 50 years beyond K-means. *Pattern Recognit. Lett.* **2010**, *31*, 651–666. [\[CrossRef\]](#)
6. Murphy, K.P. *Machine Learning: A Probabilistic Perspective*; MIT Press: Cambridge, MA, USA, 2012.
7. Biamonte, J.; Wittek, P.; Pancotti, N.; Rebentrost, P.; Wiebe, N.; Lloyd, S. Quantum machine learning. *Nature* **2017**, *549*, 195–202. [\[CrossRef\]](#) [\[PubMed\]](#)
8. Kerenidis, I.; Landman, J.; Luongo, A.; Prakash, A. q-means: A quantum algorithm for unsupervised machine learning. In *Proceedings of the Advances in Neural Information Processing Systems*; Neural Information Processing Systems Foundation, Inc.: South Lake Tahoe, NV, USA, 2019; Volume 32.
9. Getachew, A. Quantum K-medians Algorithm Using Parallel Euclidean Distance Estimator. *arXiv* **2020**, arXiv:2012.11139.
10. DiAdamo, S.; O’Meara, C.; Cortiana, G.; Bernabe-Moreno, J. Practical Quantum K-Means Clustering: Performance Analysis and Applications in Energy Grid Classification. *IEEE Trans. Quantum Eng.* **2022**, *3*, 1–16. [\[CrossRef\]](#)
11. Poggiali, A.; Berti, A.; Bernasconi, A.; Del Corso, G.M.; Guidotti, R. Quantum clustering with k-Means: A hybrid approach. *Theor. Comput. Sci.* **2024**, *992*, 114466. [\[CrossRef\]](#)
12. Wiebe, N.; Kapoor, A.; Svore, K.M. Quantum algorithms for nearest-neighbor methods for supervised and unsupervised learning. *Quantum Info. Comput.* **2015**, *15*, 316–356. [\[CrossRef\]](#)
13. Boyer, M.; Brassard, G.; Høyer, P.; Tapp, A. Tight Bounds on Quantum Searching. *Fortschritte Der Phys.* **1998**, *46*, 493–505. [\[CrossRef\]](#)

14. Dürr, C.; Hoyer, P. A Quantum Algorithm for Finding the Minimum. *arXiv* **1996**, arXiv:quant-ph/9607014.
15. Slabbert, D.; Brand, D.; Petruccione, F. Quantum spectral clustering: Comparing parameterized and neuromorphic quantum kernels. *APL Quantum* **2026**, *3*, 026106. [[CrossRef](#)]
16. Kerenidis, I.; Landman, J. Quantum spectral clustering. *Phys. Rev. A* **2021**, *103*. [[CrossRef](#)]
17. Schuld, M.; Killoran, N. Is quantum advantage the right goal for quantum machine learning? *PRX Quantum* **2022**, *3*, 030101. [[CrossRef](#)]
18. Bowles, J.; Ahmed, S.; Schuld, M. Better than classical? The subtle art of benchmarking quantum machine learning models. *arXiv* **2024**, arXiv:2403.07059.
19. Han, J.; Kamber, M.; Pei, J. *Data Mining: Concepts and Techniques*, 3rd ed.; Morgan Kaufmann: San Francisco, CA, USA, 2011.
20. Tan, P.N.; Steinbach, M.; Kumar, V. *Introduction to Data Mining*, 2nd ed.; Pearson: London, UK, 2018.
21. Lloyd, S.P. Least Squares Quantization in PCM. *IEEE Trans. Inf. Theory* **1982**, *28*, 129–137. [[CrossRef](#)]
22. Duda, R.O.; Hart, P.E.; Stork, D.G. *Pattern Classification*, 2nd ed.; Wiley-Interscience: New York, NY, USA, 2001.
23. von Luxburg, U. A Tutorial on Spectral Clustering. *Stat. Comput.* **2007**, *17*, 395–416. [[CrossRef](#)]
24. Al-Jabery, K.K.; Obafemi-Ajayi, T.; Olbricht, G.R.; II, D.C.W. 7-Evaluation of cluster validation metrics. In *Computational Learning Approaches to Data Analytics in Biomedical Applications*; Al-jabery, K.K., Obafemi-Ajayi, T., Olbricht, G.R., Wunsch, D.C., II, Eds.; Academic Press: London, UK, 2020; pp. 189–208.
25. Romano, S.; Vinh, N.X.; Bailey, J.; Verspoor, K. Adjusting for Chance Clustering Comparison Measures. *J. Mach. Learn. Res.* **2016**, *17*, 1–32.
26. Feynman, R.P. Simulating physics with computers. *Int. J. Theor. Phys.* **1982**, *21*, 467–488. [[CrossRef](#)]
27. Schuld, M.; Petruccione, F. *Supervised Learning with Quantum Computers*; Quantum Science and Technology; Springer International Publishing: Berlin/Heidelberg, Germany, 2018.
28. Giovannetti, V.; Lloyd, S.; Maccone, L. Quantum random access memory. *Phys. Rev. Lett.* **2008**, *100*, 160501. [[CrossRef](#)] [[PubMed](#)]
29. Nielsen, M.A.; Chuang, I.L. *Quantum Computation and Quantum Information*; Cambridge University Press: Cambridge, UK, 2010.
30. Bharti, K.; Cervera-Lierta, A.; Kyaw, T.H.; Haug, T.; Alperin-Lea, S.; Anand, A.; Degroote, M.; Heinen, H.; Hubbard, W.E.; Macaluso, N.J.; et al. Noisy intermediate-scale quantum algorithms. *Rev. Mod. Phys.* **2022**, *94*, 015004. [[CrossRef](#)]
31. Thabtah, F. Autism Screening Adult. UCI Machine Learning Repository, 2017. Available online: <https://archive.ics.uci.edu/dataset/426/autism+screening+adult> (accessed on 27 February 2026).
32. Thabtah, F. Autistic Spectrum Disorder Screening Data for Adolescent. UCI Machine Learning Repository, 2017. Available online: <https://archive.ics.uci.edu/dataset/420/autistic+spectrum+disorder+screening+data+for+adolescent> (accessed on 27 February 2026).
33. Thabtah, F. Autistic Spectrum Disorder Screening Data for Children. UCI Machine Learning Repository, 2017. Available online: <https://archive.ics.uci.edu/dataset/419/autistic+spectrum+disorder+screening+data+for+children> (accessed on 27 February 2026).

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.