

An Assessment of Quantum Principal Component Analysis

Mehjabin Khatoon^[0000-0002-4758-2744] and Niha Kamal
Basha^[0009-0005-5826-8071]

¹ Faculty of Engineering and Technology, Jamia Millia Islamia, Jamia Nagar, New Delhi 110025, India

mehjabinkhatoon@gmail.com

² Vellore Institute of Technology, Vellore Campus, Vellore, Tamil Nadu 632014, India
niha.k@vit.ac.in

Abstract. QPCA (Quantum Principal Component Analysis) is an algorithm that performs operations on a quantum dataset based on traditional PCA. Initially proposed for the simulation of the density matrix, it focuses on quantum system assessment utilizing the states of quantum. QPCA basically derives the principal components estimation from a dataset, whose execution period for certain terms is faster exponentially. The dataset representation assumptions are in the form of a density (or covariance) matrix, and quantum computers can prepare copies of it efficiently. The logical calculations of QPCA are helpful in cases like quantum machine learning, big data, and eigenvector estimations in polylogarithmic time. QPCA is an accelerated version of PCA, that is theoretical and scalable quantum hardware is awaited.

Keywords: Quantum · Quantum Computing · Quantum states · Tomography · PCA

1 Introduction

The emergence of quantum industries is witnessed after one hundred years of its foundational works. In 1925, Werner Heisenberg, Pascual Jordan, and Max Born, working at the local university in Germany, developed the foundational works of quantum mechanics [1, 2]. They worked on the experimental phenomena of the preceding decades like - double slit experiment, whose explanation can't be based upon classical laws. Again, in 1925, Austrian Erwin Schrodinger, affiliated with the University of Zurich, used a partial differential approach for the description via wave function of a quantum mechanical object that eventually got published in 1926 [3].

Quantum computing's basic principle is to manage challenging situations in a shorter period. Its fast computation process has affected worldwide progress, that too before the deployment of quantum mechanics. After analyzing the revolutionizing capacity of Quantum technology, various nations and industries

are investing in this challenging field that involves investments from both the commercial and public sectors [4]. Quantum Computing technology solves the problem that conventional CMOS technology faces in achieving better performance and density, in applications. This technology's revolutionary capacity attracts researchers in large numbers, but due to the existence of properties based on quantum, it leads to difficulty for beginners in this field. Therefore, authors have focused on revealing the quantum computing aspects to scientists, quantum engineers, and researchers [5]. Quantum computer limitations are not limited to 2 states; it can be both 0 and 1 at the same time (which is called superposition), contrasting bits that can be either 0 or 1, and it can be encoded as qubits, also known as quantum bits. If we head towards the quantum computer, at the atomic level electrons, ions, photons, and atoms can be realized physically, and the control devices acting correspondingly collectively work as computer memory and processor [5]. The qubits in quantum computation represent physical bits and can be treated as abstract mathematical objects. Qubit in contrast to classical bits has two states -1 or 0 and it can be in a state other than 0 or 1. It can also form states of linear combinations also termed as superpositions [6]. A quantum algorithm that depends upon quantum computation for its execution is the most pervasive model, being the model of computation based on quantum circuits [7]. In quantum computers, the computation procedure is a step-by-step procedure performed in quantum computers, in contrast to classical computers, where every step is performed on classical computers. Classical algorithms can perform on quantum computers also, but quantum computers are utilized for quantum algorithms, inherently, for computers that utilize the significant features of quantum calculation like quantum entanglement or quantum superposition [8]. The renowned quantum computing algorithms are Shor's algorithm and Grover's algorithm. Shor's algorithm explores an integer's prime factors, and Grover's algorithm, also known as the quantum search algorithm, searches an unordered list or an unstructured database.

Principal Component Analysis (PCA) is a statistical technique that achieves linear mapping for re-referencing and identification of the data. Principal Components can be termed as the formation of uncorrelated variables in smaller numbers, which are achieved after the transformation of correlated variables. The first principal component, or the primary axis, is calculated and contributes to the data variability in the greatest quantity. The remaining components, or the subsequent axes, can be ignored as they contribute to dimensionality reduction without compromising the data representation accuracy. PCA is an adaptive data analysis technique in which principal components or new variables are reduced to solve the eigenvector problem, and the dataset at hand, not a priori, defines it [9]. Quantum PCA represents an algorithm that performs the reduction of dimension in quantum data. It achieves the speedups exponentially, when dealing with datasets based on quantum as compared to the traditional PCA algorithm. The computation becomes infeasible when achieved through classical methods. The data analysis process of entanglement and superposition are the properties of quantum mechanics leveraged by QPCA that lead to the

discovery of insights and patterns not achieved via classical methods. QPCA grips the quantum mechanism principle for targeting the results achieved after the application of efficient and quick PCA methods, as compared to classical methods. It is designed for the principal component extraction of a quantum state. Those principal components are eigen-vector values in the correspondence of the largest eigen-values. An unknown quantum state's attributes reveal usual ways, for which a system is copied in that state. This is achieved by the statistical measurements of observations performed and its results [10].

The reconstruction of quantum states is done using the same quantum states that ensembles identically. The tomography states sources may be any system or device that leads quantum states into mixed states (i.e., general) or into quantum pure states. To achieve the state identification uniquely, tomography measurements must be complete. The measurements must provide all the state's information, and these types of observations can sometimes be termed as quorum. The tomography term was first introduced in the year 1993, which is related to homodyne tomography [11]. The determination of quantum states for any source output is achieved through quantum tomography on that system source. Quantum tomography determines the quantum states before the measurements are done, in contrast to a single system, in which the system states are achieved after the measurements themselves. Quantum tomography can be used for quantum information theory and quantum computing in order to achieve the qubits' actual states [12, 13]. It can also be used for optical signals characterization, including the loss and gain signals of optical devices [14]. A light beam quantum state can be shown as an infinite-dimensional density matrix or also as the plane density, termed as Wigner function. Quantum tomography can be described as a statistics problem with inverse representations that measure the results of similar quantum systems leading to its data, and the unknown parameter leads to its state [15].

The quantum system energies and the PCA principal values are the same mathematical thing [16, 17]. Quantum mechanics is all about operators and vectors, in which vectors are state vectors and operators are linear operators representing the physical measurement. For example, if we want to compute the position of a quantum particle then the position operator should be applied on the particle's associated state vector. Likewise, if there is a requirement to measure the quantum particle's energy, then an energy operator should be applied to it. The last thing to notice for the matching of PCA and quantum mechanics is that linear operators after the choosing of basis can be visualized with matrix representation. A quantum system common basis for the energy operators are the energy operator's eigenvectors. In this the matrix of energy operators is diagonal and the diagonal terms are the system's energy for the variety of energy(eigen) states. So the final interpretation is that the energy value sum corresponds to the energy operator's trace that cannot be amended under a change of basis. PCA principal values correspond to the system's energy that corresponds to it. Quantum PCA actually supports the concept of the quantum and machine learning process. PCA is an elemental method in the machine learning technology and accelerates the prospective of quantum process. The

quantum PCA process extracts the principal components and its (i.e., quantum) state tomography leads to the extraction of eigenvectors from the output state [22]. Quantum state tomography focusses on the notion of complex amplitudes addressing challenges in the perspective of non-integer eigenvalues. Thus, PCA focuses on the low-rank structure of ρ and tomography rebuilds all of ρ . Quantum tomography acts as a forerunner to PCA, like if there is a complete tomography we can execute classical PCA on the rebuild density matrix, but this is inefficient for larger systems. Quantum PCA provides a way to grasp only the significant components of the density matrix (ρ) without rebuilding all of it, therefore PCA can be utilized as an alternative approach to tomography which is an expensive approach to reconstruct density matrix. Now the reflection of thought must be generating, therefore QPCA weightage is greater than tomography concept, so the reason is clear that though QPCA provides partial knowledge but its useful (dominant eigen-structure), which is enough in the practical applications of feature extraction, data compression, outliers reduction, etc. Tomography is the costly approach since it provides complete knowledge of the state.

2 Existing Approaches

In 2013 Lloyd et al. [10] stated that quantum states can perform a role dynamically in the analysis done by itself. Thus, the state multiple copies ρ can be utilized for the unitary operator's implementation $e^{-i\rho t}$. The implementation of density matrix exponentiation ($e^{-i\rho t}$) enables the quantum phase estimation, eigenvalue extraction, and identification of principal components, all without full tomography. The density matrix exponentiation implementation reflects that Hamiltonian H (the energy operator) is the time evolution and states transformation generator, in quantum mechanics. Thus, the revealing of matrix eigenvectors and eigenvalues is achieved by density matrix multiple copies. The density matrix exponentiation, with further measurements and analysis, can provide quantum tomography benefits. The author's analysis to perform QPCA is the detection of an unknown low rank density matrix for the construction of eigenvectors in correspondence to the state's (the principal components) largest eigenvalues in an exponential speed-up time $O(\log d)$, as compared to existing algorithms.

$$\begin{aligned} \text{tr}_p(e^{-iS\Delta t}(\rho \otimes \sigma)e^{iS\Delta t}) &= \cos^2(\Delta t)\sigma + \sin^2(\Delta t)\rho - i\sin(\Delta t)\cos(\Delta t)[\rho, \sigma] \\ &= \sigma - i\Delta t[\rho, \sigma] + O(\Delta t^2). \end{aligned} \quad (1)$$

In the above equation (1), the authors applied the unitary transformation generation method by a density matrix ρ onto another state σ : $U = e^{-iS\Delta t}$, with $S = \rho \otimes \sigma$. After this, the partial trace over the system p is calculated as $\text{tr}_p(e^{-iS\Delta t}(\rho \otimes \sigma)e^{iS\Delta t})$. This operation describes how, under the "generator" ρ , the "data state" σ evolves. Expanding the result gives $(\cos^2 \Delta t)\sigma + (\sin^2 \Delta t)\rho - i\sin \Delta t \cos \Delta t[\rho, \sigma]$, where the cosine-squared term keeps part of σ , the sine-squared term mixes in part of ρ , and the commutator term produces Hamiltonian-like evolution. Finally, considering the small-time approximation ($\Delta t \ll 1$) and

using Taylor expansions, $\cos^2 \Delta t \approx 1$, $\sin^2 \Delta t \approx O(\Delta t^2)$, and $\sin \Delta t \cos \Delta t \approx \Delta t$. So, we get:

$$\sigma - i\Delta t[\rho, \sigma] + O(\Delta t^2). \quad (2)$$

This final equation reveals the short-time evolution driven by ρ ; the state σ evolves as:

$$\sigma \mapsto \sigma - iAt[\rho, \sigma]. \quad (3)$$

which is precisely the Heisenberg motion equation in which the commutator with ρ behaves like a Hamiltonian, such as the expansion of σ in the eigen basis, its eigenvectors/eigenvalues will be picked after its evolution. Thus, this leads to the foundation of Quantum PCA, since application repetition helps to efficiently extract the principal components (eigenvalues/eigenvectors).

The exponentiation of the density matrix symbolizes a tool for the properties of unknown density matrix analysis. QPCA shapes the classical procedures which consume polynomials in the system's dimension to a procedure based on quantum, that utilizes the polynomial approach in the dimension's logarithm. The advantages of this article mainly depend on the computation of the proposed quantum algorithm $O(R \log d)$, whereas, $R = \text{rank of matrix}$ and $d = \text{dimension of the data}$, and it's much better than the classical PCA which is $O(d^3)$. QPCA utilization leads to the analysis of big sized datasets with high dimensions, such as machine learning datasets, quantum chemistry and big data. The disadvantages involve the requirements of the multiple density matrix copies as an input for the implementation and preparation of many quantum states is extremely difficult. It works best for the low rank matrices and the outputs are in the quantum states not in the classical data format. QPCA can be utilized in quantum chemistry, quantum communication, quantum state tomography, machine learning, clustering algorithms, etc. QPCA is still theoretical and it should be implemented on real devices such as- Google quantum processors, IBM quantum computers, Righetti Computing quantum hardware.

In 2016, Ian T. Jolliffe and Jorge Cadima [9] have paid attention to the PCA and its numerous usages in various disciplines. The study includes new methodological outcomes, adaptations, in addition to its applications. The research work analyzes the difficulties of interpretation of large datasets, which lead to focus on PCA, that allow the interpretation easier for huge datasets as well as it minimize the loss of information. The authors refer to PCA to be an adaptive data analysis since the generation of the principal components reduces the eigenvector/eigenvalues problem solving, and the definition of the new variables are decided by the hand datasets not a priori. The technique variant nature development shaped the various structures and the data types that also leads to its adaptive characteristic. PCA is a statistical technique that has been developed much by the statisticians. PCA preserves the possible variability that creates the variables which are the original dataset's linear functions. The newly created variables are uncorrelated to each other and are successively maximize the variance. The basis of PCA is either on covariance matrix or on correlation matrix. The

results of the datasets i.e., the creation of new variables depend on the dataset itself, therefore PCA is also adaptive in nature since its results are not based upon any pre-defined functions. The main utilization of PCA is not inferential, it's actually descriptive. The authors also analyzed the PCA sensitivity to the presence of outliers, which gives rise to define the robust variants of PCA, titled as RPCA (Robust Principal Analysis). The research article highlights the wide usage of PCA, they are - simple mathematically, efficient computationally, easier interpretation. Along with the advantages, there are limitations of PCA also which involve - information loss, interpretation difficulty, scaling sensitivity, linear relationships capture, outlier sensitivity. Information loss - Some information in the data is inevitably lost due to PCA dimensionality reduction. Interpretation difficulty - Interpretations are difficult since Principal components are linear combinations of original variables, like $PC = 0.1x_1 + 0.2x_2 - 0.3x_3$; grasping such combinations of physical meaning can be challenging. Scaling sensitivity - Data standardization is mandatory since measurement of variables are in different units and results of PCA may become misleading. Linear relationships capture- PCA assumes linear relationships between variables, as it fails to capture the non-linear patterns in the data. Outlier sensitivity - Outliers or unwanted data can strongly influence the principal components. The research gap of this article includes: Handling noisy data via robust PCA, whereas in traditional PCA contains heavy noise, outliers, missing values; Development of sparse PCA, to create principal components with few non-zero variables that lead to easier interpretation, and this aim requires to develop a better algorithm; PCA methods based upon non-linear, since linear relationships were captured by standard PCA, therefore there is a requirement of developing kernel PCA, dimensional reduction techniques based on non-linear relationships; PCA requirement for high dimensional data, modern datasets may have few observations but thousands of variables, therefore this gap requires algorithms that work efficiently for low-sample datasets and of high dimension; Interpretability by PCA, sometimes interpretation through principal components remains difficult, in spite of dimensional reduction by PCA, therefore this research needs methods that improves the interpretation.

In 2018, Meng Yuanyuana and Liu Xiyu [23] used quantum concepts to detect the hidden communities in complex networks. The detection of hidden community structure was achieved by the two proposed algorithms named QIEA and QIEA-net. The detection process involves the optimization of modularity. The proposed algorithms don't require any intervention by humans, and no prior knowledge of the number of communities. QIEA follows the concept of superposition states of quantum computing's quantum bits. It belongs to the evolutionary algorithms (EA) domain. QIEA represents the basic units of information by qubits. A qubit string can define the quantum individuals. The qubit's representation can be abundant compared to the classical representation from a population diversity point of view. QIEA utilized the Q- gate (quantum gates) as an evolutionary operator in order to generate the quantum individuals of the next generation. The quantum individuals over time became a single state and also the property of diversity gradually disappears, in the evolution process. Thus, the QIEA and

QIEA-net algorithms were used for detecting the communities without adjusting many parameters. The research paper represents the optimal number of communities automatically determined by maximizing the value of the modularity. The representation makes the algorithm able to search for many candidate solutions simultaneously; it has a better ability to optimize, and also the exploration goes to large search spaces. The algorithms avoid premature convergence by maintaining population diversity that eventually improves the quality of the solutions. The optimization process is accelerated by choosing the results of an algorithm as the guiding solution and that algorithm is based upon the classical partition, and thus it results in faster convergence and better performance. The authors tested the proposed algorithms on the five real world social networks and revealed the competitive performance comparison with the existing methods of community detection. The limitations of the algorithms include the evolutionary algorithm's requirement of the convergence of many iterations that results in high computational cost, especially for the large sized networks. The algorithm's objective function relies on modularity but that limits the resolution and also fails to detect the small communities. Despite the automatic detection of the number of the communities the performance is still affected by other performance like rotation angle parameters, size of the population, mutation probability, improper parameter selection. The proposed algorithms were tested on five datasets, while the results are unsure for any dynamic networks and also for huge networks.

The algorithm can be applied in many fields like in: Social network analysis for the detection of communities in collaboration networks, Facebook, Twitter etc.; Used in Biological networks for the utilization of protein interaction networks that helps in the identification of functional biological modules; In communication networks that detect the clusters in the internet topology and also in the communication systems; It can be utilized for recommendation systems since community detection helps to find people with same preferences; It can help in financial systems such as in stock market relationships, economic networks. The research gap identified for this research article are: scalability for large networks - it struggles with large-sized networks; Real world networks changed over time often, since the current methods worked on static networks only, therefore the future work relies upon the designing of dynamic community detection algorithms; The utilization of objective function is only on modularity based, not on any other objectives such as stability and community density; Real world networks contains the overlapping communities but the proposed algorithms detects the non-overlapping communities only.

In 2021, Li et al. [25] analyzed the quantum version of PCA (QPCA) to an unknown low rank density matrix by revealing its principal components rapidly. The process of computing the principal components involves selecting large eigenvalues and discarding the one which has values below the threshold level. The QPCA's experimental implementation is challenging because of its substantial resource requirements. Therefore, the authors developed a resonance-based QPCA, i.e., RQPCA which requires a minimum resource for the ancillary qubits. The experimental procedure is achieved by analyzing a 4×4 density matrix

in which the first principal components were distilled. It minimizes the magnitude of ancillary qubits requirements while in QPCA the number of requirements of ancillary qubits are much larger for achieving the equivalent work accuracy. The requirement of smaller scale quantum circuits in Resonant Quantum Principal Component Analysis (RQPCA) enables the achievement of high valued efficiency and high valued fidelity.

RQPCA exploits resonance phenomena in a coupled quantum system to identify dominant eigen components of a data-encoded quantum state. The amplification of principal components is done selectively by resonance. The encoding of the covariance matrix is done by a mixed state of quantum or through the data qubits acted upon by the Hamiltonian. Each data eigenvalue matches an energy level. The proposed approach [25] depends upon resonance rather than quantum phase estimation on which traditional QPCA used to be depended upon. The disadvantages of the approach include limited scalability because of small quantum systems; resolution limits since peaks in the resonance may overlap and it's hard to distinguish, eigenvalues estimation is limited because of precision. The approach is sensitive to noise since its experimental setup affects the decoherent and signal noise that ultimately affects the accuracy. Its requirement is efficient quantum state preparation and data encoding, which is the major bottleneck. This approach is hardware-specific which is designed for NMR systems and it may not directly translate to trapped ions or superconducting qubits. The research gap includes the larger systems scalability for large datasets and high dimensional matrices requirements of extension; implementation on cross platforms such as on IBM quantum devices and Google quantum processors; Noise-resilient techniques that requires integration with error mitigation and quantum computing fault tolerant; It can also combine classical preprocessing and resonant quantum refinement, which is not explored in depth; the eigenvalues extraction is in main focus in this approach, eigenvectors extraction is an open problem; Requirement of stronger benchmarking with classical PCA is there. The applications includes quantum machine learning, feature extraction process in clustering and classification process; signal processing in resonance based detection utilized in noise filtering and analysis of frequency; in quantum chemistry by analyzing molecular eigenstates and energy spectra; big-data analytics by dimensional reduction in high dimensional datasets; in medical imaging by PCA based de-noising and compression; experimental quantum computing devices through quantum devices comparison and testing system response and its coherency.

In 2022, Gordon et al. [21] worked on the gap of covariance matrix encoding in the density matrix, which lacks a concrete protocol. The authors have proven the preparation of covariance matrix by centering the data. The authors exploit a global phase symmetry in quantum dataset and proven the existence of covariance matrix with the formula $\bar{\rho} = \sum_i p_i |\psi_i\rangle\langle\psi_i|$. There, according to the centred dataset, the covariance matrix is precisely $\bar{\rho}$. Global phase symmetry was utilised for the quantum datasets to prove the consistency of the centred dataset with $\bar{\rho}$. For uncentered data, it's interpreted as a PCA from a symmetrized dataset, for the method of PCA without centering. "PCA and PCA without centering"

contrast was proven by the authors through a graph image, after the analysis was achieved for it. In the comparison graph, for PCA an ellipsoid fitting into the data was shown and for PCA without centering the interpretation was shown by a symmetrized version of the dataset in which an ellipsoid was fitting into the symmetrized dataset, means the dataset is duplicated with a minus sign of a duplicated copy. Therefore, the authors argued that the PCA without centering can also be called “PCA on a symmetrized version of the dataset”. The authors focus on the matrix in which the eigenvalues and eigenvectors were worked upon, and keep only the principal components of the matrix. Therefore, PCA without centering is relevant to the proposed approach of quantum PCA. The experimental analysis includes both the classical and quantum datasets, whereas, in classical datasets (stored in classical computers) each datapoint information is about global phase preservation. Therefore, the factors of multiplicative have a non-trivial effect on each datapoint of the dataset whereas, in the quantum dataset each datapoint information about global phase is erased, therefore factors multiplicative have a non-physical (or trivial) effect on the dataset.

The advantages of this research article solve the QPCA key bottleneck. It just converts the classical data into a quantum covariance matrix which makes the QPCA more realistic and practical. It works for noisy quantum computers and limited qubits, which makes it compatible with the hardware, such as Google quantum systems, IBM. This paper gives introduction of the techniques to encode covariance matrix efficiently and also reduces complexity of the circuits. Integration of the approach can be done with different quantum algorithms and also with quantum-classical hybrid algorithms.

The approach also leads towards the practical implementation of QPCA, while earlier approaches are mostly theoretical. Although it is an improved approach, there is a requirement of deep circuits and multiple qubits, and also, it's not fully scalable. The covariance preparation quantum circuits are sensitive to noise and are decoherent affective, which lessens the accuracy of the real device. The current implementations are not yet proven for large scale datasets, its only tested on small data sets and simplified examples. A major bottleneck is the requirements of resources and time, for encoding the classical data into quantum states. This work improved the preparation of covariance, but still QPCA is not end-to-end i.e., pipeline optimization is still incomplete. Therefore, the research gap includes the problem of data loading efficiently, large datasets scalability, covariance preparation noise resilient, full QPCA pipeline integration, validations experimentally. The applications of this research work involve quantum machine learning-based dimensionality reduction, feature extraction; big data analysis based dominant features extraction, high-dimensional datasets handling; quantum chemistry based studying molecular systems, quantum states analyzing, signal processing based on pattern recognition, filtration of noise, finance and optimized risk modeling, portfolio analysis.

In 2023, Emmanuele Dri et al. [22] have utilized the quantum state tomography for the eigenvectors extraction and have applied the concept of QPCA interest rate risk in a practical quantum finance use case. The authors have first prepared

the quantum state via amplitude encoding of the input matrix as given in the below equations.

$$A \in \mathbb{R}^{N \times N}, \quad \text{tr}(A) = 1. \quad (8)$$

$$|\psi_A\rangle = \sum_{i=1}^N \sum_{j=1}^N A_{ij} |i\rangle |j\rangle = \sum_{k=1}^r \sigma_k |u_k\rangle |u_k\rangle. \quad (9)$$

Whereas, r = the rank of the matrix, σ_k = the singular values, u_k = the singular vectors of A . According to the above equation, if $\text{tr}[A] \neq 1$, then it will be followed up with matrix normalization in pertaining to its trace, after adding up with a classical preprocessing step, after which it leads to the output preprocessing that includes obtained eigenvalues multiplication with the trace itself. The author's next focus is on the unitary transformation step, which is the core process for phase estimation. Its assumption is based on the multiple copies of A , that allows efficient generating of the discussed step. After the previous step, the phase estimation algorithm is followed to generate the eigenvalues and input matrix eigenvector, in the subsequent steps. Thus, determining the precision for eigenvalues, principal components extraction is done with the help of the state tomography process. So, the quantum state tomography technique, helps to determine a system's quantum states. Thus, the authors worked on the gap of principal components extraction with the help of QPCA, which focuses on the tomography technique of pure state reconstruction with real amplitudes. If, in this process, there is a generation of non-integers eigenvalues then a precise representation is not possible, and in that case extension of tomography is must, in order to fit in the new constraints. Other research gaps, including the non-exploration of fault tolerant QPCA; requirement of quantum data encoding efficiently are still; further analyzation of classical PCA framework of hybrid quantum; limitation of the benchmarking against in terms to methods of advanced classical PCA; it still lacks the huge industrial datasets applications. The workflow has gained advantage when compared with previous classical PCA - speed gains exponentially for the huge datasets; high dimensional data compact representation; quantum machine learning-based algorithms natural integration; end-to-end workflow demonstration a realistic. The workflow has also gained advantage when compared with QPCA [10] implementation, which is more realistic; data-loading costs are considered; hardware limitations are included; tests are implemented on real platforms. Despite these many advantages there is a disadvantage - bottleneck of data loading due to the expensiveness of classical data into quantum states.

In 2025, Nhat A. Nghiem [18] proposed another approach for the calculation of Quantum PCA in which the computation of covariance matrix was done in a new method, on a quantum computer for a given classical dataset. Thus, a quantum state was revealed by an efficient framework which also maintains the original QPCA philosophy as the analyzing of the states can be achieved itself. The authors got motivation from the QSVT (quantum singular value transformation) framework, which makes it possible to achieve a unified description of earlier

quantum techniques [19]. QSVT allows dealing with computational problems with stronger means, which leads to the construction of the Petz recovery channel [20]. QPCA marked a very significant cornerstone in the perspective of machine learning algorithms as PCA is one of the central tools within big data science and machine learning. In quantum computers, certain assumptions allow the key features that reveal the covariance matrix principal components, that refers to data features correlation. The features of quantum state can be revealed by quantum PCA. Thus, it also allows for self-tomography, which is generated by relying on measurements of quantum state and also it is different from the standard tomography approaches. The authors [18] have constructed a covariance matrix which is better than that developed by Gordon et al.[21] since the previous one can naturally deal with uncentered databases and also there is no loss of contribution from the centroids. An empirical mean of the data is the so-called centroid:

$$\mu = \frac{1}{N} \sum_{i=1}^N x^i. \quad (4)$$

Whereas, $\{x^i\}_{i=1}^N$ is a set of N data points in the perspective of classical PCA, where each point $x^i \in \mathbb{C}^n$ is a vector of dimensions n , that corresponds to the states of the quantum in an n dimensional Hilbert space. This is particularly for the uniform distribution setting and, the mean of the data points for the certain distribution can be achieved by equation (5).

$$\mu = \sum_{i=1}^N p^i x^i. \quad (5)$$

The dataset, which is centred, can be computed by the new coordinate definition for all data points.

$$b^i = x^i - \mu. \quad (6)$$

The covariance matrix can be defined as:

$$A_{jk} = \sum_{i=1}^N p^i b_j^i (b_k^i)^* = \sum_{i=1}^N p^i x_j^i (x_k^i)^* - \mu_j \mu_k. \quad (7)$$

whereas $*$ refers to the complex conjugate, and the rest of the parameters are explained already. The diagonalization of matrix A is the most crucial step of PCA, and the foremost eigenvectors/eigenvalues of A are the principal components analysis. Precisely, it can be inferred that PCA treatment should be in full generality, which means it should be centered (i.e., it should include the contribution from the centroids). In Gordon et al.[21], they worked on the previous research gap of covariance matrix encoding on the density matrix, for which the dataset should be centered (zero value for all the features' mean value).

However, the proposed algorithm [18] has pros and cons that can be measured through benchmarking. QRAM dependency has been reduced in this proposed

algorithm as compared to the earlier QPCA. The practical feasibility has been improved through the proposed covariance preparation method. In specific regimes, the proposed algorithm performs better when the situation of identifying conditions of outperforming traditional QPCA and its significance lies, when most quantum speedups are not universal, it's actually conditional. Preparation of the improved covariance matrix is possible since data representation is more efficient because of the construction of quantum states, and the covariance preparation is quantum machine learning's biggest bottleneck.

The framework flexibility is achieved since it moves beyond the approach of only density matrices, and also this flexibility allows the unification of newer quantum techniques. The approach's nearness to quantum computing potentially makes it more adaptable to forthcoming quantum devices. Unlike most QML algorithms, it requires validation on an actual quantum system; it lacks implementation on real hardware. Speedup conditions are limited, since performance sometimes may match classical PCA, or sometimes it is less efficient, while benefits of it sometimes exist in some specific regime. Data loading into quantum states is still a bottleneck and costly. The approach is sensitive to noise; it's affected by gate errors and decoherence and cannot fully remove the issues related to noise. The implementation complexity also exists since its implementation is hard and requires expertise related to deep circuit design. Thus research gap can be concluded after the discussion of cons, that experimental validation lacking that requires real devices implementation and benchmarking; analysis on scalability i.e., more analysis required on large datasets, high -dimensional inputs, and focus on the benefits issue on the high scale; design is non-resilient, focus on fault tolerance is weak, therefore, noise resilient quantum algorithms integration is required; the paper also lacks in integration of classical preprocessing with quantum PCA, so hybrid models are required to be more practical in nearness quantum computing; lacks practical on datasets such as financial, medical, image datasets etc. The applications of the approach involved in the quantum machine learning which helps in the feature extraction, clustering that enhances quantum neural networks; in the dimensionality reduction of big data where classical PCA seems to be expensive; in the signal and image processing such as in noise and compression filtration in medical imaging, satellite data etc.; analysis of risk and finance on market pattern detection, portfolio optimization, risk modeling using PCA, etc.; in the quantum chemistry and anomaly and cybersecurity detection, like detection of some unusual patterns, detection of frauds etc.

In 2026 [4], Gill et al. outlined the situations and benefits of the quantum era. The authors have diagrammed the quantum computing overview with an explanation of the root concepts of quantum computing named superposition, entanglement, and decoherence. Superposition exploits the feature of quantum particles being present at once in several states on the subatomic scale. Quantum devices follow the qubits whose value lies between 0 and 1 (any floating-point number), and actually its presence can be at any time or it can also co-exist, thanks to quantum mechanics Procopio et al. [24]. Entanglement- the next basic attribute, in which qubits enable the bits arrangement in an entangled state, unlike

Table 1. Comparison of existing approaches.

No.	Title	Year	Focus
1	Quantum principal component analysis	2013	Development of an algorithm that efficiently detects the principal components of a quantum density matrix, providing substantial acceleration over classical PCA for suitable quantum data.
2	Principal component analysis: a review and recent developments	2016	PCA for multivariate data analysis, with emphasis on interpretation, robustness, big-data scalability, and non-linear extensions.
3	Quantum Inspired Evolutionary Algorithm for Community Detection in Complex Networks	2018	Quantum-inspired qubits use superposition so that a single chromosome can represent multiple states of a qubit.
4	Resonant quantum principal component analysis	2021	Replaces the phase-estimation approach with resonance to reduce circuit complexity and resource requirements.
5	Covariance Matrix Preparation for Quantum Principal Component Analysis	2022	Preparation of a covariance matrix on a quantum computer using quantum circuits, measurement-based techniques, and data-encoding strategies.
6	Towards An End-To-End Approach For Quantum Principal Component Analysis	2023	Performs the complete PCA workflow on quantum hardware.
7	New Quantum Algorithm for Principal Component Analysis	2025	Improves traditional QPCA from the perspective of practical feasibility and data preparation.
8	Transforming research with quantum computing	2026	Discusses quantum-computing principles and their potential to address challenging problems in shorter computation times.

in conventional bits, where the bits assignment is in the independent manner. The entangled qubits may be separated but the characteristics of the qubits can be altered after just one of them is probed and one associated global state is maintained. Quantum machine creation in large scale faces unique obstruction, but quantum machines make it possible to tackle the complex problems in a simple way because of the parameter space massive expansion. The main challenge is the quantum states “decoherence” labelling that used to encode qubits. The qubits might lose their coherent characteristics whenever it tries to communicate with the outside world. Therefore, challenges are faced in the creation of quantum gadgets across a wide scope. The authors have outlined the capabilities of quantum computing with the outlined causes. The capability of solving complex problems in the area of financial modelling, supply chains, complicated chemical structures, risk evaluation etc., by quantum AI are much quicker as compared to conventional systems. In business opportunities, startups utilized the quantum algorithms without the quantum hardware. Thus, this technical democratization is a game changer leading the startups, quantum algorithms exploration in their respective areas. The discussion of the applications of Quantum Computing have made this article [4] more fruitful focuses on - security of the data, AI boosting, modelling of the urban traffic flow accuracy of the improved climate prediction, and cutting-edge medicines designing. The discussion ends by summarizing the trends of next-generation quantum computing research fields and their open challenges.

All the existing approaches are compared and summarised in Table 1.

3 Conclusion

Quantum-based principal component analysis paves the way towards developing a fruitful algorithm that enables quantum dimensionality reduction by extracting principal eigenvalues and eigenvectors. The limitation of converting classical datasets into quantum states needs to be addressed. The analysis of a quantum state that gives information about a quantum system should be done properly. The efficiency of a quantum state should be measured using quantum tomography, which performs multiple measurements on as many identical copies of that state in the process of quantum state reconstruction. Principal component analysis is a technique that reduces the dimension statistically while preserving the information as much as possible; thus, the quantum-based dimensionality reduction feature is maintained. The main challenge still lies in the interpretation and information loss, while achieving the aim of QPCA. Surveying the research papers, future study can be focused on QPCA with fewer qubits, shallow quantum circuits, depolarizing noise, effect of significant eigenvalues on QPCA, identification of dominant components through the generation of quantum states.

Code and Data Availability

This work is based on secondary data and doesn't involve any code or dataset.

References

1. Fundamentals first, technology second. *Nature Nanotechnology* 20 (2025). <https://doi.org/10.1038/s41565-025-02002-2>
2. Heisenberg, W.: Über quantentheoretische Umdeutung kinematischer und mechanischer Beziehungen. *Zeitschrift für Physik* 33(1), 879–893 (1925)
3. Schrödinger, E.: An undulatory theory of the mechanics of atoms and molecules. *Physical Review* 28(6), 1049–1070 (1926)
4. Gill, S.S., Buyya, R.: Transforming research with quantum computing. *Journal of Economy and Technology* 4, 1–8 (2026). <https://doi.org/10.1016/j.ject.2024.07.001>
5. Bhat, H.A., Kaushik, B.K., Khan, F.A., Bashir, F., Shah, K.A.: Quantum computing: Fundamentals, implementations and applications. *IEEE Access* 12, 14532–14574 (2024). <https://doi.org/10.1109/ACCESS.2024.3353654>
6. Nielsen, M.A., Chuang, I.L.: *Quantum Computation and Quantum Information*. Cambridge University Press, Cambridge (2000)
7. Mosca, M.: Quantum algorithms. *arXiv preprint arXiv:0808.0369* (2008)
8. Lanzagorta, M., Uhlmann, J.K.: *Quantum Computer Science*. Morgan & Claypool Publishers, San Rafael (2009)
9. Jolliffe, I.T., Cadima, J.: Principal component analysis: A review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 374(2065), 20150202 (2016). <https://doi.org/10.1098/rsta.2015.0202>
10. Lloyd, S., Mohseni, M., Rebentrost, P.: Quantum principal component analysis. *Nature Physics* 10(9), 631–633 (2014). <https://doi.org/10.1038/nphys3029>

11. Smithey, D.T., Beck, M., Raymer, M.G., Faridani, A.: Measurement of the Wigner distribution and the density matrix of a light mode using optical homodyne tomography: Application to squeezed states and the vacuum. *Physical Review Letters* 70(9), 1244–1247 (1993). <https://doi.org/10.1103/PhysRevLett.70.1244>
12. Blume-Kohout, R.: Optimal, reliable estimation of quantum states. *New Journal of Physics* 12(4), 043034 (2010). <https://doi.org/10.1088/1367-2630/12/4/043034>
13. Lvovsky, A.I., Raymer, M.G.: Continuous-variable optical quantum state tomography. *Reviews of Modern Physics* 81(1), 299–332 (2009). <https://doi.org/10.1103/RevModPhys.81.299>
14. D’Ariano, G.M., De Laurentis, M., Paris, M.G.A., Porzio, A., Solimeno, S.: Quantum tomography as a tool for the characterization of optical devices. *Journal of Optics B: Quantum and Semiclassical Optics* 4(3), S127–S132 (2002). <https://doi.org/10.1088/1464-4266/4/3/366>
15. Artiles, L.M., Gill, R.D., Guta, M.I.: An invitation to quantum tomography. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 67, 109–134 (2005)
16. Jauregui, J.: Linear algebra of PCA. Available at: <https://www.math.union.edu/~jauregui/PCA.pdf>
17. Jaffe, R.L.: The postulates of quantum mechanics. Available at: <https://web.mit.edu/8.05/handouts/jaffe1.pdf>
18. Nghiem, N.A.: New quantum algorithm for principal component analysis. arXiv preprint arXiv:2501.07891 (2025). <https://doi.org/10.48550/arXiv.2501.07891>
19. Gilyén, A., Su, Y., Low, G.H., Wiebe, N.: Quantum singular value transformation and beyond: Exponential improvements for quantum matrix arithmetics. In: *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing (STOC 2019)*, pp. 193–204 (2019). <https://doi.org/10.1145/3313276.3316366>
20. Gilyén, A., Lloyd, S., Marvian, I., Quek, Y., Wilde, M.M.: Quantum algorithm for Petz recovery channels and pretty good measurements. *Physical Review Letters* 128(22), 220502 (2022). <https://doi.org/10.1103/PhysRevLett.128.220502>
21. Gordon, M.H., Cerezo, M., Cincio, L., Coles, P.J.: Covariance matrix preparation for quantum principal component analysis. *PRX Quantum* 3(3), 030334 (2022). <https://doi.org/10.1103/PRXQuantum.3.030334>
22. Dri, E., Aita, A., Fioravanti, T., Franco, G., Giusto, E., Ranieri, G., Corbelleto, D., Montrucchio, B.: Towards an end-to-end approach for quantum principal component analysis. In: *2023 IEEE International Conference on Quantum Computing and Engineering (QCE)*, pp. 1–6 (2023). <https://doi.org/10.1109/QCE57702.2023.10175>
23. Meng, Y., Liu, X.: Quantum inspired evolutionary algorithm for community detection in complex networks. *Physics Letters A* 382, 2305–2312 (2018). <https://doi.org/10.1016/j.physleta.2018.05.044>
24. Procopio, L.M., Moqanaki, A., Araújo, M., et al.: Experimental superposition of orders of quantum gates. *Nature Communications* 6, 7913 (2015)
25. Li, Z., Chai, Z., Guo, Y., Ji, W., Wang, M., Shi, F., Wang, Y., Lloyd, S., Du, J.: Resonant quantum principal component analysis. *Science Advances* 7(34), eabg2589 (2021). <https://doi.org/10.1126/sciadv.abg2589>