

# Quantum Circuit Algorithm for Heart Disease Detection

Panuwat Tanapornchinpong and Prabhas Chongstitvatana \*

*Department of Computer Engineering, Chulalongkorn University, Bangkok 10330, Thailand*

(\*Corresponding author's e-mail: [prabhas.c@chula.ac.th](mailto:prabhas.c@chula.ac.th))

*Received: 30 April 2025, Revised: 5 May 2025, Accepted: 12 May 2025, Published: 25 June 2025*

## Abstract

This study investigates hybrid quantum-classical K-Means algorithms for clustering heart disease patient data. Using quantum swap-test circuits for distance calculation, 2 approaches were tested on both noisy and ideal quantum simulators. A real-world dataset of over 1,000 records was preprocessed using standard techniques, including normalization, outlier removal, and dimensionality reduction. The results show that both quantum methods achieve accuracy up to 0.83 and F1-scores comparable to the classical K-Means baseline (0.82 - 0.83), even when executed on quantum simulators with real-device noise models. These findings highlight the practical potential of quantum-enhanced clustering methods. The study supports the feasibility of applying quantum-enhanced clustering in medical analytics, particularly for early-stage heart disease detection and patient risk stratification.

**Keywords:** Quantum, K-means, Machine learning, Unsupervised learning, Swap test, Heart disease dataset

## Introduction

Machine learning has been widely applied across various industries, including healthcare, where it plays an important part in analyzing vast datasets to predict diseases. One significant application is in the early detection of heart disease, a leading cause of death globally where early diagnosis can significantly improve outcomes [1]. Heart disease diagnosis can be predicted using historical medical data from patients. Many studies utilize this data and apply machine learning methods, such as Support Vector Machines (SVM) [2,3], Convolutional Neural Networks (CNN) [4], and ensemble learning approaches [5,6], to enhance prediction accuracy. These advancements enable the automation of diagnostic processes, improving both the speed and accuracy of heart disease detection.

As datasets continue to grow exponentially, handling large-scale and high-dimensional data poses significant challenges. Traditional machine learning methods on classical computers often encounter performance bottlenecks and limitations in scalability. Quantum technology, however, promises to revolutionize computing by addressing these challenges. Quantum computing holds immense potential to

enhance the efficiency of machine learning algorithms. Leveraging quantum properties such as superposition and entanglement, quantum computers can outperform classical systems in specific computational tasks. Unlike classical computers, which process data sequentially, quantum computers utilize quantum parallelism to perform multiple calculations simultaneously [7].

Despite these advantages, current quantum devices fall under the category of Noisy Intermediate-Scale Quantum (NISQ) systems [8], which are constrained by noise and hardware limitations. Nonetheless, these NISQ devices show promise for computationally intensive tasks, particularly in high-dimensional datasets [9]. To address current limitations, hybrid quantum-classical algorithms have been developed. These algorithms assign resource-intensive tasks, such as distance computation or optimization, to quantum systems, while classical systems provide stability and control. This approach enables the practical use of quantum devices in fields such as medical diagnostics and predictive analytics.

Recent research has introduced several quantum-enhanced machine learning algorithms, including Quantum Random Forest (QRF), Quantum K-

Nearest Neighbors (QKNN), Quantum Decision Tree (QDT), and Quantum Gaussian Naïve Bayes (QGNB), where classical systems handle dataset preprocessing and quantum systems accelerate computation. These hybrid models have demonstrated superior performance compared to traditional models in terms of prediction accuracy [10].

In the area of unsupervised learning, clustering plays a vital role in pattern recognition and grouping of similar data points. Quantum-enhanced clustering algorithms emerge as promising tools for efficiently handling high-dimensional datasets. Several studies have applied quantum computing to improve clustering performance and reduce complexity. For example, Poggiali *et al.* [10] proposed 3 strategies for generating shorter-depth quantum circuits to address the limitations of NISQ devices. Their method was tested using the MNIST dataset on both quantum simulators and IBM's real hardware. Similarly, Khan *et al.* [11] introduced a hybrid quantum K-means algorithm that leverages quantum parallelism to compute distances between centroids and data points simultaneously, reducing the complexity of the assignment step. Another study, Quiroga *et al.* [12], employed basic linear algebra subroutines (BLAS) to further reduce clustering time complexity.

In a different direction, Arthur and Date [13] presented a method for solving the balanced K-means clustering problem using quadratic unconstrained binary optimization (QUBO), encoding 8-dimensional classical data into a quantum format and benchmarking their results with Principal Component Analysis (PCA) on Modified National Institute of Standards and Technology (MNIST) dataset. Meanwhile, Kerenidis *et al.* [14] proposed a fully quantum clustering algorithm, though it was not tested on physical hardware. In Urgelles *et al.* [15], a quantum Euclidean distance calculation was implemented on quantum hardware for a 6G indoor positioning scenario, showing potential for use in K-means clustering. Furthermore, real-world applications such as recommender systems, electricity grid classification, and medical prediction for diseases like heart disease and diabetes have demonstrated the practical value of quantum clustering [16-19].

This study introduces a hybrid quantum clustering algorithm applied to a heart disease dataset. The dataset undergoes preprocessing and transformation into

quantum data before being input into a quantum circuit. The algorithm leverages quantum parallelism to optimize the efficiency of distance computations between data points and centroids, a critical step in the K-means clustering process. We propose 2 distinct quantum circuit designs, each employing a different approach to distance calculation, and evaluate their performance on both synthetic and heart disease datasets, comparing the results with classical methods. The designs are implemented and tested using IBM's quantum simulator, Qiskit, which simulates quantum hardware to assess the quality and efficiency of the proposed circuits.

The rest of the paper is structured as follows: materials and methods, which describe the dataset, preprocessing steps, and quantum circuit design; experimental results and discussion, which present performance evaluations and comparisons with classical methods; and conclusion, which summarizes the key findings and discusses potential directions for future research.

## Materials and methods

**Data preparation:** The heart disease dataset is pre-processed to address missing values, standardize features using techniques like standard scaling and PCA, duplicate elimination, and remove outliers to ensure consistency and reliability. The cleaned dataset is then used as input for the classical K-Means algorithm, and its performance is assessed using metrics. These results serve as a reference benchmark for evaluating the quantum implementation.

**Quantum data preparation:** Before implementing the dataset into the quantum algorithm, the classical data must be mapped to quantum states through quantum data encoding.

**Quantum distance calculation:** The proposed algorithm utilizes a quantum computer to calculate distances through a swap test circuit in 2 configurations: Single point to single centroid, and single point to multiple centroids. Once the distance calculations are completed, the centroid update step is executed on a classical computer, following a procedure similar to the classical K-Means algorithm.

**Evaluation:** The performance of both the quantum and classical implementations is compared using the same benchmark metrics established earlier. This

comparison highlights the advantages and limitations of the quantum approach relative to the classical baseline.

### Classical K-means

The K-Means algorithm was first proposed by Lloyd [20]. It is a widely used method for unsupervised clustering [21], designed to partition a dataset  $X = x^{\rightarrow}_1, \dots, x^{\rightarrow}_M$  of  $M$  data points (each datapoint is an  $N$ -dimensional vector) into  $k$  clusters  $c^{\rightarrow}_1, \dots, c^{\rightarrow}_k$ . The algorithm proceeds through 2 main steps until convergence: The 1<sup>st</sup> step is cluster assignment. Firstly, the  $k$  centroid was randomly selected from the dataset. Then, each data point is assigned to the nearest centroid using the Euclidean distance:

$$D(\vec{a}, \vec{b}) = \sqrt{\sum_{i=1}^N (a_i - b_i)^2} \quad (1)$$

where  $a^{\rightarrow}$  and  $b^{\rightarrow}$  are  $N$ -dimensional vectors. The 2<sup>nd</sup> step is centroid update. After all data points are assigned to their nearest centroids, the centroids are recalculated as the mean of all the points assigned to each cluster. Then, the recalculated centroids will be used in the next iteration. The goal of K-Means is to minimize the within-cluster sum of squared errors (WSSE), which is given by:

$$WSSE = \sum_{k=1}^K \sum_{m \in c_k} \|x_m - c_k\|^2 \quad (2)$$

where  $\vec{x}_m$  represents the data points and  $\vec{c}_k$  represents the centroids. The optimal number of clusters  $k$  is often using the elbow method, where the within-cluster sum of squares (WCSS) is plotted against  $k$ , and the ‘‘elbow’’ point, where the rate of decrease in WCSS slows down, is identified [22]. The classical K-Means algorithm has a time complexity of  $O(kMN)$ , where  $k$  is the number of clusters,  $M$  is the number of data points, and  $N$  is the dimensionality of the data. The computation cost can become expensive for large datasets, encouraging the exploration of more efficient alternatives, such as quantum-based methods.

### Quantum data preparation and distance calculation

Quantum K-means differ from classical K-means in how distances are computed. In classical K-means, distances are calculated using coordinates in the feature

space, employing Euclidean distance or other similar metrics. In contrast, Quantum K-means computes distances probabilistically using the quantum states of qubits. Before performing calculations in quantum, all classical data must first be mapped into quantum states through a process called quantum data encoding. Encoding techniques can be done in many different ways [23]. In this paper, amplitude encoding is used for encoding input features into the amplitude of a quantum system

Amplitude encoding is a technique in quantum computing for representing classical data within quantum states. Given a classical vector  $x = (x_0, x_1, \dots, x_{N-1})$ , the encoding maps the elements of  $x$  to the amplitudes of a quantum state, resulting in the following representation:

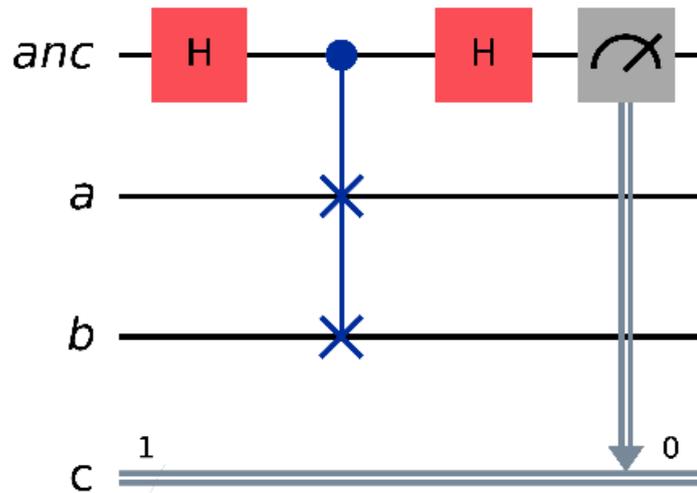
$$|\psi\rangle = \sum_{j=0}^{N-1} x_j |j\rangle \quad (3)$$

where  $x_j$  are the components of the classical vector, and  $|j\rangle$  denotes the binary representation of  $j$  in a quantum system with  $\log_2(N)$  qubits. To ensure that  $|\psi\rangle$  is a valid quantum state, the vector  $x$  must be normalized, satisfying the condition.

$$\sum_{j=0}^{N-1} |x_j|^2 = 1 \quad (4)$$

For example, if  $x = [1, 2]$ , the vector has  $N = 2$  elements, requiring  $n = \log_2 2 = 1$  qubits. The indices  $j$  of the vector elements are then mapped to binary representations using  $n$ -bit strings, such as  $|0\rangle, |1\rangle$  for  $j = 0, 1, 2$  respectively. Next, the vector  $x$  must be normalized to ensure it satisfies the quantum state’s normalization condition. The normalized vector is obtained by dividing each component by the vector’s L2-norm. The norm is  $\sqrt{1^2 + 2^2} = \sqrt{5}$ , resulting in the normalized

vector  $x_{\text{normal}} = \left[ \frac{1}{\sqrt{5}}, \frac{2}{\sqrt{5}} \right]$ . The result quantum state is  $|\psi\rangle = \left[ \frac{1}{\sqrt{5}} |0\rangle, \frac{2}{\sqrt{5}} |1\rangle \right]$



**Figure 1** Swap test circuit.

In quantum computing, distance calculation between vectors can be performed using the SWAP test circuit, illustrated in **Figure 1**. The SWAP test is a widely used quantum subroutine designed to estimate the overlap between 2 quantum states by measuring the probability of the ancillary qubit being in the  $|0\rangle$  state after a controlled quantum operation. Initially, 2 quantum states,  $|a\rangle$  and  $|b\rangle$  are prepared to represent the classical vectors to be compared. These are encoded into separate qubit registers. An ancillary qubit is initialized in the  $|0\rangle$  state and passed through a Hadamard gate, resulting in the combined state:

$$\frac{1}{\sqrt{2}}(|0\rangle + |1\rangle) \otimes |a\rangle \otimes |b\rangle \tag{5}$$

Then, the controlled-SWAP gate is applied between the vector qubits (a and b), controlled by the ancilla qubit (q[0]). The result state is shown as:

$$\frac{1}{\sqrt{2}}(|0\rangle |a\rangle |b\rangle + |1\rangle |a\rangle |b\rangle) \tag{6}$$

Hadamard gate is applied again at the end which result in the state as:

$$\frac{1}{\sqrt{2}} |0\rangle (|a\rangle |b\rangle + |b\rangle |a\rangle) + |1\rangle (|a\rangle |b\rangle - |b\rangle |a\rangle) \tag{7}$$

Finally, the ancillary qubit is measured of probability at the state  $|0\rangle$  shown as:

$$P(0) = \frac{1 + |\langle a|b\rangle|^2}{2} \tag{8}$$

This probability is directly related to the inner product of the 2 states. If the states are orthogonal,  $P(0) = 0.5$ , and if they are identical,  $P(0) = 1$ .

To leverage all features from the dataset in a quantum circuit, this study adopts the Euclidean distance estimation approach proposed in [14]. Specifically, we utilize the quantum circuit design from [14], depicted in **Figure 2** encodes 2 quantum states  $|a\rangle$  and  $|b\rangle$  derived from vectors x and y, respectively, using the following scheme:

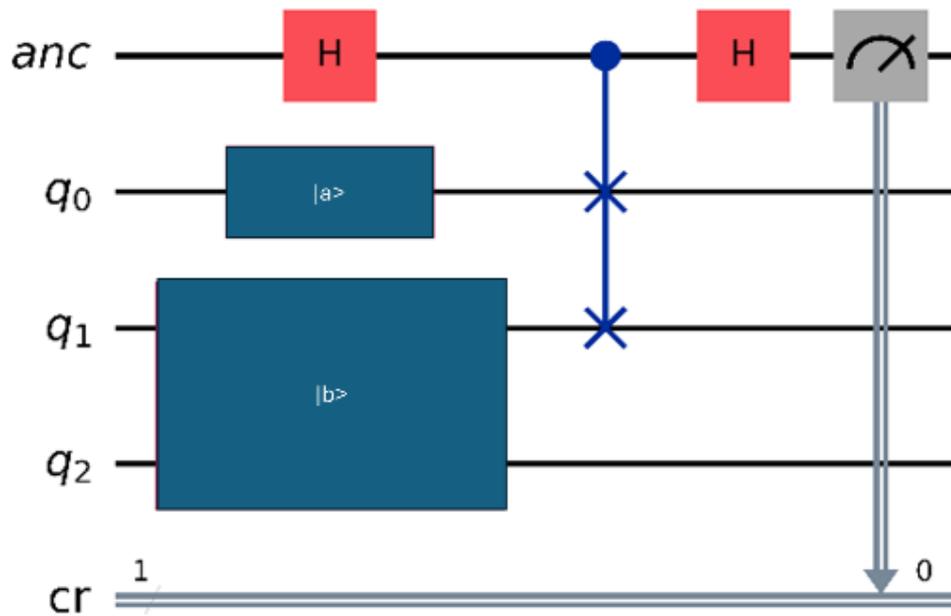


Figure 2 Swap test full circuit.

$$| a \rangle = \frac{1}{\sqrt{2}}(| 0 \rangle \otimes | x \rangle + | 1 \rangle \otimes | y \rangle) \tag{9}$$

$$| b \rangle = \frac{1}{\sqrt{Z}}(| x | 0 \rangle + | x | 1 \rangle) \tag{10}$$

where  $Z = |x|^2 + |y|^2$  (11)

The Euclidean distance D between the 2 vectors can be calculated as:

$$D = \sqrt{2 - Z|\langle a | b \rangle|^2} \tag{12}$$

This technique requires only  $O(\log(N) + 1)$  qubits, where N is the dimension of the classical input vectors. This method optimizes quantum resource usage, making it suitable for implementation on near-term quantum hardware. As explained, data encoding and distance calculations are performed on the quantum computer, while cluster assignments and centroid updates are handled on the classical computer. A detailed

description of the complete algorithm will be provided in the next section.

### Quantum K-mean algorithms

Two approaches to the Quantum K-Means algorithm are introduced. The 1<sup>st</sup> algorithm follows a 1-point-1-centroid strategy (1p1k), where each data point is individually compared against all centroids using quantum computation. The overview pseudocode of the 1<sup>st</sup> approach will show as Algorithm 1. An overview of this approach is presented in Algorithm 1. Its structure closely mirrors that of the classical K-Means algorithm: centroid initialization is performed using the K-Means++ method (Line 1) to improve convergence, and centroid updates are executed classically by averaging assigned data points (Line 14). The key distinction lies in Lines 5 - 9, where quantum computation is employed to estimate distances. In this step, each data point-centroid pair is encoded into quantum states, and the distance is measured using a quantum swap test.

---

**Algorithm 1** Quantum K-Means Algorithm

---

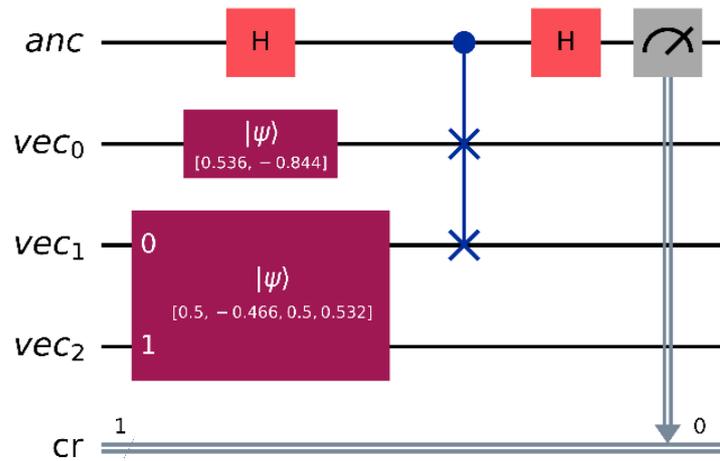
**Require:** Dataset -  $D$ , Clusters -  $k$ , Max iter -  $T$ , quantum shots -  $s$   
**Ensure:** Centroids -  $C$ , Assignments -  $A$

```

1:  $C \leftarrow \text{InitCentroids}(D, k)$  ▷ Random init
2:  $\text{Converged} \leftarrow \text{False}, t \leftarrow 0$ 
3: while not  $\text{Converged}$  and  $t < T$  do
4:    $\text{Dist} \leftarrow []$ 
5:   for each  $x \in D$  do
6:      $d_x \leftarrow []$ 
7:     for each  $c \in C$  do
8:        $Q \leftarrow \text{Encode}(x, c)$  ▷ Quantum Encoding
9:        $d \leftarrow \text{QDist}(Q, s)$  ▷ Quantum swap test
10:      Append  $d$  to  $d_x$ 
11:    end for
12:    Append  $d_x$  to  $\text{Dist}$ 
13:  end for
14:   $A \leftarrow \text{Assign}(\text{Dist})$  ▷ Classical Step
15:   $C' \leftarrow \text{Update}(D, A, k)$ 
16:  if  $C' = C$  then ▷ Check if centroids remain the same
17:     $\text{Converged} \leftarrow \text{True}$ 
18:  else
19:     $C \leftarrow C'$ 
20:  end if
21:   $t \leftarrow t + 1$ 
22: end while
23: return  $(C, A)$ 

```

---



**Figure 3** Quantum circuit of algorithm 1.

This algorithm follows a hybrid quantum-classical model, utilizing quantum hardware exclusively for distance calculation. The swap test component in Lines 5 - 9 operates with a time complexity of  $O(d)$ , where  $d$  is the feature dimension. The overall per-iteration complexity of the algorithm is  $O(Tnkd)$ , where  $T$  is the number of iterations,  $n$  is the number of data points, and  $k$  is the number of centroids. The 2<sup>nd</sup> approach, known as the 1-point-many-centroids strategy (1pMK),

leverages quantum parallelism to compute distances between a single data point and all centroids simultaneously. This significantly reduces the number of quantum circuit executions required per iteration. In this method, a single quantum circuit encodes the data point and all centroids concurrently, and ancilla qubit measurements are used to extract distance-related probabilities for all centroids in parallel.

---

**Algorithm 2** Quantum K-Means Algorithm (1P-MK)

---

**Require:** Dataset -  $D$ , Clusters -  $k$ , Max iter -  $T$ , Quantum shots -  $s$

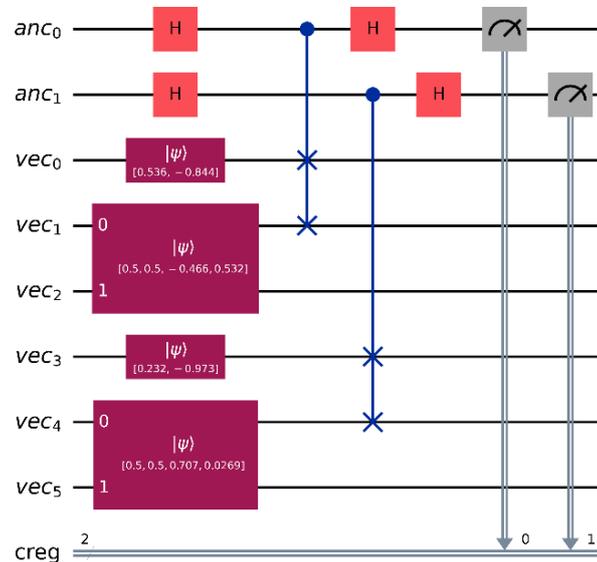
**Ensure:** Centroids -  $C$ , Assignments -  $A$

```

1:  $C \leftarrow \text{InitCentroids}(D, k)$  ▷ Random initialization
2:  $\text{Converged} \leftarrow \text{False}, t \leftarrow 0$ 
3: while not  $\text{Converged}$  and  $t < T$  do
4:    $\text{Dist} \leftarrow []$  ▷ Stores all distances
5:   for each  $x \in D$  do ▷ Iterate over each data point
6:      $Q \leftarrow \text{Encode}(x, C)$  ▷ Quantum encoding of all centroids in parallel
7:      $d_x \leftarrow \text{QDist}(Q, s)$  ▷ Quantum swap test for all centroids
8:     Append  $d_x$  to  $\text{Dist}$ 
9:   end for
10:   $A \leftarrow \text{Assign}(\text{Dist})$  ▷ Classical cluster assignment
11:   $C' \leftarrow \text{Update}(D, A, k)$  ▷ Update centroids based on new assignments
12:  if  $C' = C$  then ▷ Check if centroids remain the same
13:     $\text{Converged} \leftarrow \text{True}$ 
14:  else
15:     $C \leftarrow C'$ 
16:  end if
17:   $t \leftarrow t + 1$ 
18: end while
19: return  $(C, A)$ 

```

---



**Figure 4** Quantum circuit of algorithm 2.

### Experiment setting

Two proposed Quantum K-Means algorithms are implemented and tested on 2 different quantum platforms: A noisy quantum simulator and an ideal quantum simulator. The experiments are conducted using real-world datasets. The performance of the quantum algorithms is evaluated by comparing them with each other and with the traditional K-Means algorithm.

**Dataset:** The dataset used in this study originates from a public health database [24] and consists of 13 attributes, including 1 target variable. A total of 1,025 patient records were available. For clustering analysis, 8 features were selected based solely on their categorical

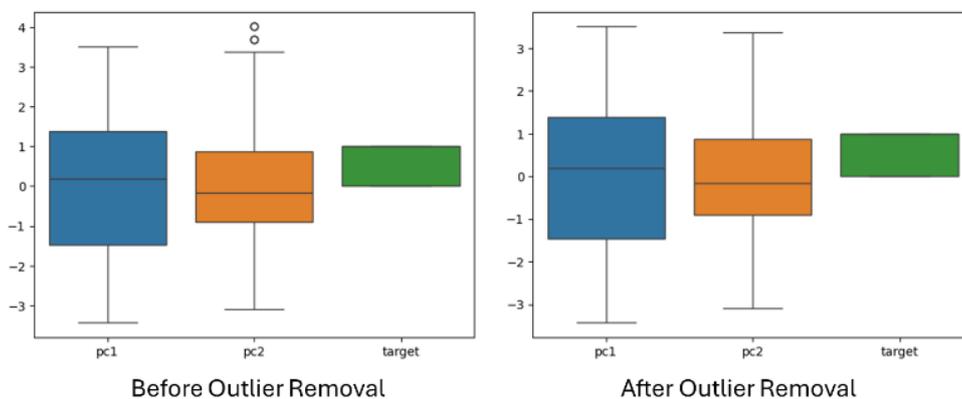
data type. Numerical and continuous variables were excluded to ensure compatibility with categorical-based encoding schemes. The selected features are illustrated in **Figure 5**. Prior to applying clustering algorithms, the dataset underwent several preprocessing steps. First, feature values were normalized using standard scaling to prevent attributes with larger numerical ranges from dominating the clustering process. Outliers were removed using the Interquartile Range (IQR) method to enhance robustness by eliminating extreme values as shown in **Figure 6**. Duplicate records were also removed to minimize bias and reduce computational overhead. Lastly, dimensionality reduction was performed using Principal Component Analysis (PCA).

PCA was applied to project the data onto 2 principal components, both to improve clustering efficiency and to enable visualization in a lower-dimensional space. After preprocessing, the dataset was reduced to

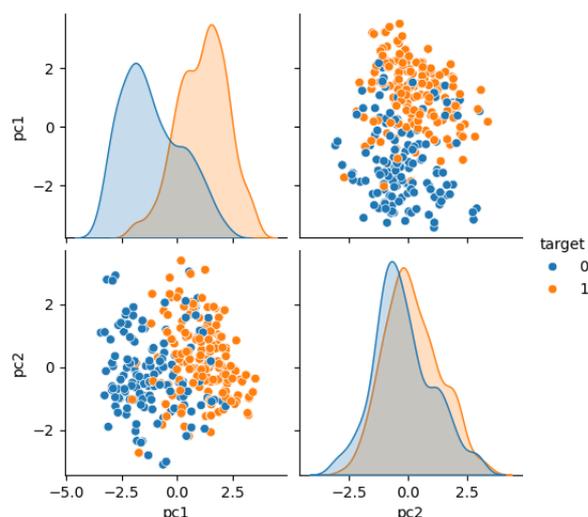
300 clean and representative records. The distribution of the transformed data is visualized in **Figure 7** using a pair plot.

Feature	Category → Encoded Value
Sex	Male → 1, Female → 2
Chest Pain Type	Typical angina → 1, Non-anginal pain → 2, Atypical angina → 3, Asymptomatic → 4
Fasting Blood Sugar	Lower than 120 mg/ml → 1, Greater than 120 mg/ml → 2
Rest ECG	ST-T wave abnormality → 1, Normal → 2, Left ventricular hypertrophy → 3
Exercise Induced Angina	No → 1, Yes → 2
Slope of ST Segment	Flat → 0, Downsloping → 1, Upsloping → 2
Vessels Colored by Fluoroscopy	Zero → 1, One → 2, Two → 3, Three → 4, Four → 5
Thalassemia	Fixed Defect → 0, Reversible Defect → 1, Normal → 2, No → 3

**Figure 5** Selected 8 categorical features.



**Figure 6** Outlier removal process.



**Figure 7** Pair Plot of pre-processed data.

Quantum platforms: AerSimulator is used to emulate a real quantum system with noise. A custom noise model based on the real IBM backend `ibm_brisbane` is applied to simulate realistic quantum behavior. Additionally, the `qasm_simulator` is used to provide ideal (noise-free) quantum simulation results, serving as a baseline for comparison.

Evaluation parameters: The performance comparison between traditional and quantum clustering approaches was evaluated using 3 clustering-quality metrics:

Accuracy: Measures the proportion of correctly assigned cluster labels compared to the ground truth labels.

$$\text{Accuracy} = \frac{\text{Predictions}}{\text{Total Predictions}} \quad (13)$$

Precision: Evaluates the quality of positive cluster assignments. It measures the proportion of true positive assignments among all points assigned to a specific cluster.

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}} \quad (14)$$

Recall: Measures the completeness of the clustering by calculating the proportion of true positives that were correctly identified out of all actual positives.

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}} \quad (15)$$

F1-score: The F1-score is the harmonic mean of precision and recall. It provides a balanced metric that considers both false positives and false negatives.

$$\text{F1} = 2 \left( \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \right) \quad (16)$$

Sum of Squared Errors (SSE): Measures the compactness within clusters by summing squared distances from each data point to its assigned centroid. Lower SSE indicates more compact clusters.

$$\text{SSE} = \sum_{i=1}^k \sum_{x \in C_i} \|x - \mu_i\|^2 \quad (17)$$

Silhouette Score(sil): Measures how well-separated clusters are from each other. Values range from  $-1$  to  $1$ , with higher scores indicating better-defined clusters.

$$\text{sil} = \frac{b-a}{\max(a,b)} \quad (18)$$

where  $a$  is the mean distance between a record and all other records in the same cluster, and  $b$  is the mean distance between a record and all other records in the nearest neighbouring cluster.

V-measure (V-meas): Combines homogeneity (each cluster contains only data points of a single class) and completeness (all data points from a given class belong to the same cluster). It ranges from  $0$  to  $1$ , with higher scores indicating better clustering consistency.

$$\text{V} = \frac{2(\text{homogeneity} + \text{completeness})}{\text{homogeneity} \cdot \text{completeness}} \quad (19)$$

## Result and discussion

As illustrated in **Figure 8**, the dataset comprises 300 records, each reduced to 2 principal components and labeled with ground truth. The classical K-Means result serves as a baseline for comparison with the quantum clustering results. **Figure 9** presents experimental results obtained using a quantum simulator with the noise model from the real quantum device `ibm_brisbane`. Both Algorithm 1 (1p1k) and Algorithm 2 (1p1k) yield identical cluster assignments, closely matching those of classical K-Means. Despite this, both quantum approaches exhibit some misclassifications when compared to the true labels. These discrepancies are not due to quantum computation errors but are likely caused by data points lying near cluster boundaries, where classification is inherently ambiguous.

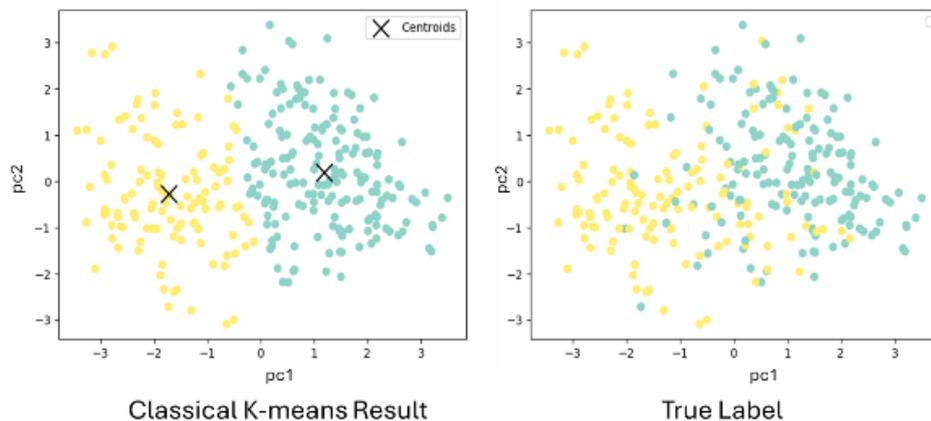


Figure 8 Classical K-means result and true label.

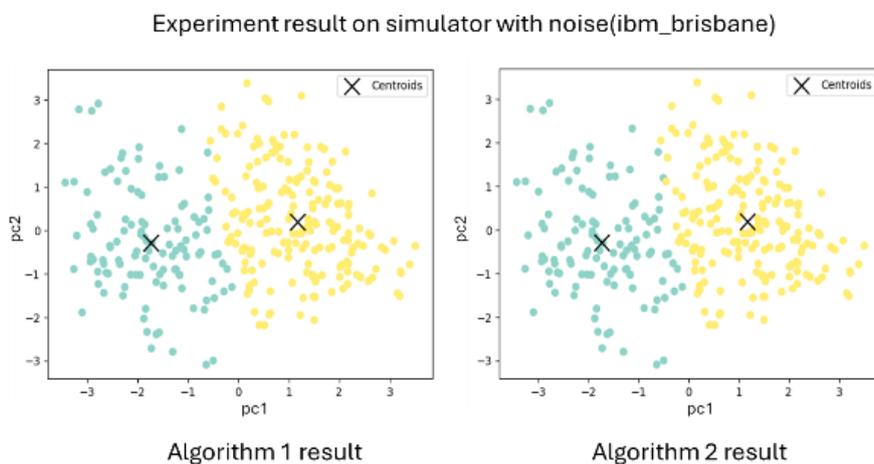


Figure 9 Experiment result on simulator with noise(ibm\_brisbane).

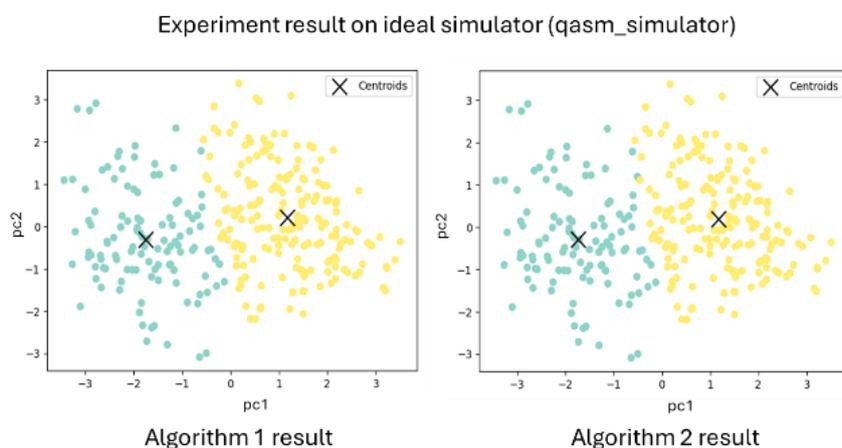


Figure 10 Experiment result on ideal simulator(qasm\_simulator).

When the experiments were conducted on an ideal quantum simulator (*qasm\_simulator*), as shown in **Figure 10**, the results remained consistent. Both

algorithms once again produced identical cluster assignments and maintained high similarity with classical K-Means.

**Table 1** Summary of evaluation metrics.

Algorithm	Accuracy	Precision	Recall	F1	SSE	Sil	V-meas
Classical K-mean	0.82	0.83	0.82	0.82	665.06	0.433	0.334
1p1k (ibm_brisbane)	0.82	0.83	0.82	0.83	665.17	0.433	0.328
1pmk (ibm_brisbane)	0.82	0.83	0.82	0.83	665.17	0.434	0.353
1p1k(qasm_simulator)	0.83	0.83	0.82	0.83	665.17	0.435	0.346
1pmk(qasm_simulator)	0.82	0.83	0.82	0.83	665.17	0.433	0.328

**Table 2** Comparison with other work.

Algorithm	Accuracy	Precision	Recall	F1
XGboost [6]	0.97	0.95	0.97	0.93
Correlation Based Feature Selection [25]	0.85	0.80	0.85	0.79
Quantum Circuit [18]	0.96	0.97	0.94	0.96
Classical K-mean	0.82	0.83	0.82	0.82
1p1k(ibm_brisbane)	0.82	0.83	0.82	0.83
1pmk(ibm_brisbane)	0.82	0.83	0.82	0.83
1p1k(qasm)	0.83	0.83	0.82	0.83
1pmk(qasm)	0.82	0.83	0.82	0.83

The results summarized in **Table 1** indicate that the proposed quantum K-Means algorithms (1p1k and 1pmk) achieve clustering performance closely comparable to the classical K-Means baseline across all key evaluation metrics. Specifically, the quantum models show similar Accuracy scores of 0.82 - 0.83, Precision, Recall, and F1-score values, with only minor variations. Additionally, clustering quality metrics such as SSE, Silhouette Score, and V-measure remain consistent between classical and quantum approaches, even when executed on real quantum hardware (ibm\_brisbane) with noise present. These results demonstrate the robustness of quantum clustering techniques under practical hardware conditions.

In **Table 2**, a comparison with other methods on the same heart disease dataset is presented. Advanced models such as XGBoost [6] achieve higher Accuracy (~0.97) through supervised learning, which leverages label information to optimize decision boundaries. However, when focusing on unsupervised clustering approaches, the proposed quantum K-Means algorithms

show competitive performance. Compared to Correlation-Based Feature Selection [25], which achieved an Accuracy of 0.85, the proposed quantum models maintain comparable results while operating under the stricter constraints of unsupervised learning. Furthermore, when compared to Quantum Circuit Clustering [18], which reported an Accuracy of 0.96, the performance of the proposed method remains respectable, especially considering the simplicity of the circuits (one qubit, angle encoding) and the execution under real hardware noise.

However, a key factor contributing to the observed clustering performance (~0.82 - 0.83) is the inherent overlap between the 2 true label groups, as illustrated in **Figure 8**. The 2 classes are not cleanly separable in the reduced feature space. This overlap limits clustering performance, as unsupervised models cannot exploit label information to adjust decision boundaries. Therefore, the results should be interpreted not as a shortcoming of the quantum clustering approach but rather as a reflection of the data complexity.

As the problem complexity increases (e.g., clustering datasets with more than 2 features), each quantum K-Means algorithm faces different scalability challenges. Algorithm 1 (1p1k), which processes each data-centroid pair individually, is more scalable for larger datasets because it maintains a lower qubit requirement and iteratively computes distances. In contrast, Algorithm 2 (1pmk) encodes all centroids simultaneously into a single quantum circuit, leveraging quantum parallelism to reduce circuit executions. While efficient for small datasets, Algorithm 2 demands significantly more qubits, increasing susceptibility to hardware limitations and noise as the number of centroids or feature dimensions grows.

This trade-off between quantum parallelism and hardware resource availability should be carefully considered when selecting the appropriate quantum K-Means strategy for larger and more complex datasets.

### Conclusions

This study explored the application of hybrid quantum-classical K-Means algorithms for clustering patient data related to heart disease. By incorporating quantum distance computation using swap test circuits, the study demonstrated how quantum computing can enhance computational efficiency and parallelism within a traditional clustering workflow. Two proposed algorithms, namely Algorithm 1 (1 Point to 1 Centroid) and Algorithm 2 (1 Point to Many Centroids), were evaluated using a real-world public health dataset containing over 1,000 patient records. Experimental results from both noisy and ideal quantum simulators showed that the clustering outcomes were closely aligned with those produced by the classical K-Means algorithm, with comparable performance across standard evaluation metrics.

Quantum clustering faces limitations due to current hardware constraints, including limited qubits, circuit depth, and noise. Algorithm 2 is especially impacted, as encoding multiple centroids in one circuit limits scalability for high-dimensional data or large cluster sizes.

Despite current challenges, the results show strong potential for quantum clustering in medical data analysis. With advancements in quantum hardware and error correction, it could support early disease detection and risk assessment. Future work should focus on circuit

optimization, scalable encoding, and testing on larger datasets using real quantum devices.

### Acknowledgements

The author would like to thank Professor Prabhas Chongstitvatana for valuable guidance throughout this research.

### Declaration of Generative AI in Scientific Writing

Generative AI tools were used solely to improve the clarity and language of the manuscript. All content was created, verified, and reviewed by the authors, who remain fully responsible for the scientific accuracy and integrity of the work.

### CRediT author statement

Panuwat Tanapornchinpong: Conceptualization; Methodology; Software; Writing - Original draft; Visualization. Prabhas Chongstitvatana: Conceptualization; Writing - Review & Editing; Supervision.

### References

- [1] SMS Shah, FA Shah, SA Hussain and S Batool. Support vector machines-based heart disease diagnosis using feature subset, wrapping selection and extraction methods. *Computers & Electrical Engineering* 2020; **84**, 106628.
- [2] A Salau, T Admassu, G Chhabra, K Kaushik and SL Braide. Heart disease detection model using support vector machine with feature selection. *In: Proceedings of the 2<sup>nd</sup> International Conference on Advancement in Computation & Computer Technologies (InCACCT)*, Gharuan, India. 2024, p. 204-207.
- [3] AA Ahmad and H Polat. Prediction of heart disease based on machine learning using jellyfish optimization algorithm. *Diagnostics* 2023; **13(14)**, 2392.
- [4] M Porumb, E Iadanza, S Massaro and L Pecchia. A convolutional neural network approach to detect congestive heart failure. *Biomedical Signal Processing and Control* 2020; **55**, 101597.
- [5] ID Mienye, Y Sun and Z Wang. An improved ensemble learning approach for the prediction of heart disease risk. *Informatics in Medicine Unlocked* 2020; **20**, 100402.

- [6] H El-Sofany, B Bouallegue and YM Abd El-Latif. A proposed technique for predicting heart disease using machine learning algorithms and an explainable AI method. *Scientific Reports* 2024; **14(1)**, 23277.
- [7] J Biamonte, P Wittek, N Pancotti, P Rebentrost, N Wiebe and S Lloyd. Quantum machine learning. *Nature* 2017; **549(7671)**, 195-202.
- [8] J Preskill. Quantum computing in the NISQ era and beyond. *Quantum* 2018; **2**, 79.
- [9] M Cerezo, A Arrasmith, R Babbush, SC Benjamin, S Endo, K Fujii, JR McClean, K Mitarai, X Yuan, L Cincio and PJ Coles. Variational quantum algorithms. *Nature Reviews Physics* 2021; **3(9)**, 625-644.
- [10] A Poggiali, A Berti, A Bernasconi, GMD Corso and R Guidotti. Guidotti. Quantum clustering with k-Means: A hybrid approach. *Theoretical Computer Science* 2024; **992**, 114466.
- [11] SU Khan, AJ Awan and G Vall-llosera. K-Means clustering on noisy intermediate scale quantum computers. *ArXiv* 2019, 1-12.
- [12] D Quiroga, P Date and R Pooser. Discriminating quantum states with quantum machine learning. *In: Proceedings of the 2021 IEEE International Conference on Quantum Computing and Engineering (QCE)*, Broomfield, CO, USA. 2021, p. 56-63.
- [13] D Arthur and P Date. Balanced k-means clustering on an adiabatic quantum computer. *Quantum Information Processing* 2021; **20(9)**, 294.
- [14] I Kerenidis, J Landman, A Luongo and A Prakash. q-means: A quantum algorithm for unsupervised machine learning. *ArXiv* 2018, p. 1-11.
- [15] H Urgelles, P Picazo-Martínez and JF Monserrat. Application of quantum computing to accurate positioning in 6G indoor scenarios. *In: Proceedings of the ICC 2022 - IEEE International Conference on Communications*, Seoul, Korea. 2022, p. 643-647.
- [16] O Ouedrhiri, O Banouar, S El Hadaj and S Raghay. Intelligent recommender system based on quantum clustering and matrix completion. *Concurrency and Computation: Practice and Experience* 2022; **34(15)**, e6943.
- [17] S DiAdamo, C O'Meara, G Cortiana and J Bernabé-Moreno. Practical quantum k-means clustering: Performance analysis and applications in energy grid classification. *IEEE Transactions on Quantum Engineering* 2022; **3(11)**, 1-16.
- [18] SS Kavitha and N Kaulgud. Quantum K-means clustering method for detecting heart disease using quantum circuit approach. *Soft Computing* 2022; **27(18)**, 13255-13268.
- [19] H Gupta, H Varshney, TK Sharma, N Pachauri and OP Verma. Comparative performance analysis of quantum machine learning with deep learning for diabetes prediction. *Complex & Intelligent Systems* 2022; **8(4)**, 3073-3087.
- [20] S Lloyd. Least squares quantization in PCM. *IEEE Transactions on Information Theory* 1982; **28(2)**, 129-137.
- [21] A Likas, N Vlassis and JJ Verbeek. The global k-means clustering algorithm. *Pattern Recognition* 2003; **36(2)**, 451-461.
- [22] C Shi, B Wei, S Wei, W Wang, H Liu and J Liu. A quantitative discriminant method of elbow point for the optimal number of clusters in clustering algorithm. *EURASIP Journal on Wireless Communications and Networking* 2021; **2021(1)**, 31,
- [23] M Rath and H Date. Quantum data encoding: A comparative analysis of classical-to-quantum mapping techniques and their impact on machine learning accuracy. *EPJ Quantum Technology* 2024; **11(1)**, 72.
- [24] K Gangal. Heart disease dataset (UCI). UCI, Irvine , California, 2025.
- [25] AM Kuruvilla and NV Balaji. Heart disease prediction system using correlation based feature selection with multilayer perceptron approach. *IOP Conference Series: Materials Science and Engineering* 2021; **1085(1)**, 012028.