Quantum K-means clustering method for detecting heart disease using quantum circuit approach

S S Kavitha · Narasimha Kaulgud

Abstract
The development of noisy intermediate-scale quantum computers is expected to signify the potential advantages of quantum computing over classical computing. This paper focuses on quantum paradigm usage to speed up unsupervised machine learning algorithms particularly the K-means clustering method. The main approach is to build a quantum circuit that performs the distance calculation required for the clustering process. This proposed technique is a collaboration of data mining techniques with quantum computation. Initially, extracted heart disease dataset is preprocessed and classical K-means clustering performance is evaluated. Later, the quantum concept is applied to the classical approach of the clustering algorithm. The comparative analysis is performed between quantum and classical processing to check performance metrics.

Keywords
Qubits · Quantum machine learning (QML) · Quantum K-means clustering · IBMQ

1 Introduction
Machine learning (ML) plays a vital role in solving various highly challenging tasks due to its ability to handle computational improvements. Although computing power has grown rapidly in recent decades and new algorithms have hit the market regularly, the growth in data has outpaced the growth in computing power. As a result, in the field of artificial intelligence, which in many cases relies on massive data, a lack of computational capacity eventually becomes a shortcoming. As the dataset becomes huge, it requires fast and efficient algorithms to process. Polynomial-time algorithms are not always thought to be the most efficient as it takes more time to process it Xu and Tian (2015). One of the fields, where it is becoming difficult to handle the increasing data management, is the healthcare sector.

Despite significant medical improvements, heart failure continues to pose a significant hazard to patients’ health. Currently, doctors must manually diagnose and anticipate the condition based on their knowledge, experience, and observations. It poses a challenge for healthcare practitioners since it results in extremely high rates of mortality and morbidity (Abdel-Basset et al. 2020). Developing healthcare systems based on patient data models and technologies can significantly improve the decision-making process. Despite the fact that the medical industry collects a large amount of data, it must be processed accurately and rapidly to provide an early and automated diagnosis. Furthermore, various machine learning (ML) algorithms have demonstrated promising results in the diagnosis of heart disorders, and they were used to solve a variety of complex issues with efficient and accurate outcomes (Rani et al. 2021; Katarya and Meena 2021; Bakhsh 2021). The ability of ML algorithms to learn the relationship between input data and output class using standard probability and logic theories is one of the fundamental issues. However, even with faster ML methods, there exists a computational bottleneck on handling large data for traditional computers. Hence to enhance the computational power, a new paradigm quantum computers came into existence.

Quantum mechanics (QM) has also been shown to outperform classical theory-based models in a variety of disciplines (including classification, object detection, object tracking,
prediction, and optimization) (Shao et al. 2019; Biamonte et al. 2017). With a burgeoning amount of data, existing machine learning algorithms are finding it difficult to tackle computationally intensive tasks. In this aspect, quantum computing power may provide a competitive advantage in machine learning applications (Möller and Vuik 2017). As a result, quantum computing and medical systems provide hardware sophisticated results that allow the healthcare industry to evaluate and treat complicated medical disorders. Thus, it is concluded that both classical and quantum machine learning methods contribute to the earlier detection of heart diseases and the identification of predictive traits in various pathological states (Kumar et al. 2021). Based on these, the heart disease dataset chosen has to be properly preprocessed before applying machine learning algorithms (Yang and Nataliani 2017; Wetschereck et al. 1997). The various state of the art carried out to understand and find a new approach to diagnose heart disease at early stage is discussed in next section

1.1 Related work and motivation

Various researchers had made different approach to implement clustering QML algorithm. Bharill et al. (2015) use an improved quantum-inspired evolutionary fuzzy C-means (EQIE-FCM) algorithm to calculate the global optima of the parameters. The author uses “quantum K-means” and “quantum fuzzy C-means” clustering approach to predict diabetes. The limitation was imposed on calculating the processing time required to form the clustering. Aïmeur et al. (2013) propose divisive clustering, k-medians clustering methods to form clustering on random data. Grover search approach is used for distance calculation and form cluster. The author only proposes the theoretical concepts on how the black-box model is used to query the distance between pairs of points. Lacks to construct the quantum circuit that has the same functionality as the black box to perform clustering. Casaña-Eslava et al. (2020) focus of the work is on developing a new algorithm for cluster allocation with autonomous hyper-parameter selection, and efficient computation has not been analyzed deeply. For cluster identification by density, the authors adopt the Schrödinger equation to deal with local length parameters. The probabilistic quantum clustering (PQC) was implemented on real and synthetic datasets.

Yao et al. (2008) propose the main benefit and drawback of the quantum clustering algorithm (QC). The authors explored the improved algorithm—exponent distance-based quantum clustering algorithm (EQDC) based on its shortcomings observed in QC. This (EQDC) approach was implemented on the IRIS dataset. No preprocessing was applied to data before analyzing EQDC. The processing time required to form the clustering and other performance factors is not mentioned. Singh and Bose (2021) in this paper, “fast forward quantum optimization algorithm (FFQOA)” along with “K-means clustering (KMC) algorithm” was used to segment chest CT images of COVID-19 patients. The FFQOA technique was only tested with chest images. The future work was imposed to work with distinct medical images, such as X-rays and MRIs, as well as a variety of engineering design challenges. Li et al. (2011) integrate the concept of quantum walks (QWs) with the problem of data clustering and develop two clustering algorithms based on one-dimensional discrete-time QWs. The proposed QML algorithms are implemented on Soybean, Iris, Sonar, Glass, Ionosphere, and Breast cancer Wisconsin datasets. Simulated only on classical machines. No quantum processing operations are performed.

Arthur (2021) present a quantum approach to solve the balanced K-means clustering training by exploring quadratic unconstrained binary optimization (QUBO). The balanced K-means clustering method was imposed on the IRIS dataset. The dataset chosen is not so computationally expensive as quantum can process large datasets at a faster speed. Farouk (2017) uses a labeled dataset of Arabic vowels to be applied on the QC algorithm. The accuracy and processing time of non-hierarchical kernel techniques for unsupervised clusterings, such as K-means, self-organizing map (SOM), and fuzzy C-means (FCM), are then computed. The K-means, SOM, and FCM methods are applied to process speech data. Implemented in MATLAB with no intervention of quantum processor. The usage of all available data on Arabic vowels is limited due to processing time constraints. Kumar et al. (2021) propose a comparison of classical and quantum-enhanced machine learning techniques to predict patients’ diabetes. The quantum-enhanced machine learning was implemented on the PIDD dataset. Improper rotation of quantum logic gates in quantum computers is prone to induce inaccuracies, which resulted in erroneous output. Khan et al. (2019) build K-means clustering based on (i) quantum interference, (ii) negative rotations, and (iii) destructive interference. The proposed K-means clustering method was implemented using the IRIS dataset, MNIST dataset. Lacks to address qubit’s coherence times and noise which reduces the effect on solving problems with high accuracy.

Inspired by the work carried out by various researchers in the field of quantum unsupervised learning, the proposed work highlights the use of the quantum K-means approach to predict heart disease with significant improvements in clustering performance compared to classical machines.

The main motivation of this work is—the existing ML models are generally characterized by a sophisticated quantum of data that leads to process overhead (Haq et al. 2018; Mohan et al. 2019). Most of the healthcare datasets available are observed to be very huge as it has to store all the information in order to have a proper prediction. This high dimensionality of the data makes it difficult to visual-
ize even in 3D, referred to as the curse of dimensionality (Keogh and Mueen 2017). As a result, performing operations on this data requires a lot of memory, and the data can grow exponentially at times, leading to overfitting. The weighting features are utilized to eliminate redundancy in the heart disease dataset, which helps to speed up the execution time and thereby predict the disease precisely at early stages. Here the quantum computer comes into the picture which has a high speed of computing due to its parallelism feature and can solve specific algorithms with distinct problems (Kerenidis et al. 2018; Sarma et al. 2019). To execute machine learning algorithms on a quantum computer, the classical algorithms must be redesigned to adhere to quantum mechanics principles (Ramezani et al. 2020). Quantum primitive-based algorithms for supervised (Acampora 2019) and reinforcement learning (Lamata 2021) have already been developed. However, with the exception of a quantum technique for the minimal spanning tree, limited work has been done on unsupervised learning related to the healthcare domain. Several quantum clustering algorithms use Grover’s search and show a speedup over their classical counterparts, but they are unable to increase the quality of the clustering process (Rubio et al. 2017). From this motivations, the contribution of the proposed work is summarized in the next subsection.

### 1.2 Contribution

In this work, we contribute three major aspects of quantum K-means method an unsupervised learning with the focus on quantum circuit approach.

- First, we propose different quantum data preparation method to convert classical data into quantum state. From this, the best data preparation method based on performance metrics is chosen for further processing.
- Second, we provide quantum circuit approach to calculate distance between centroid and datapoints. We repeat the iteration of centroid and cluster updation until the convergence is reached.
- Third, we show the effectiveness of quantum clustering with classical clustering with experimental results, which exhibits the efficacy of the proposed method.
- Finally, the performance evaluation of the proposed work is compared with other state-of-the-art methods. The improvement in the proposed method results is due to the less time taken by quantum processor to form the clusters with proposed data preparation and distance calculation methods.

Unsupervised quantum algorithm implementation on NISQ computers based on quantum circuits can provide better results considering the number of gates required to build the circuit. Executing these ML algorithms on a quantum system reduces processing overhead and also improves/matches with performance metrics of classical systems. Therefore, this work focuses on how well the quantum computing paradigm can be utilized to perform clustering and speed up the processing time of unsupervised learning and also other performance measures such as accuracy, precision, F1-score, sensitivity, and specificity.

The rest of this article is organized in the following order. The methodology incorporated to predict heart disease using classical and quantum approach is discussed in Sect. 2. In Sect. 3, we explain the classical K-means algorithm to cluster the heart disease dataset. Next, we explain the quantum K-means approach for clustering the heart disease dataset in Sect. 4. We present experimental results and comparisons of quantum K-means using quantum circuits with classical in Sect. 5 and finally, Sect. 6 presents the conclusion with future work.

## 2 Methodology

Methodology carried out to predict heart disease in classical and quantum context using K-means clustering method is shown in Fig. 1.

1. Classical data preparation, preprocessing and K-means approach: Heart disease dataset chosen is preprocessed to check NULL values, and use relevant normalization techniques such as standard scalar and PCA scalar to scale all features and identify and eliminate outliers to obtain better performance metrics. Finally classical K-means algorithm is implemented to evaluate accuracy, precision, sensitivity, specificity, and F1-score.
2. Quantum data preparation: All the classical data are converted to quantum state by applying different data preparation methods. The best data preparation method is chosen for further processing.
3. Distance calculation using quantum swap test circuit and cluster and centroid updation: After data preprocessing (normalization and outlier rejection) quantum K-means algorithm is implemented to evaluate all performance measures as that of classical to relate the comparative analysis.
4. Finally, the quantum cluster formed should be relatively or equally comparable with the classical cluster formed, along with this performance metrics and processing time between two machines are recorded.
5. Also state-of-the-art comparison is carried out to exhibit the performance of the proposed method.
2.1 Heart disease prediction data description

The data content of the heart disease prediction consists of a collection of records, representing the diagnosis of the patient heart disease (Al-Yarimi et al. 2021). The extracted dataset was the Public Health dataset (Gangal 2021). It contains 13 features including the predicting attribute. Since the chosen dataset contains 8 categorical data, it is handled by converting to integer during preprocessing stage as shown in Table 1. The “target” field indicates whether or not the patient has cardiac disease. Target = “0” refers to no disease and Target = “1” refers presence of disease. A total of
1025 patients record are summarized for the prediction of cardiac disease. The first two columns tabulate the “age” and “sex” of the patient. The rest of the columns describe the type of chest pain, resting blood pressure value, cholesterol level, fasting blood sugar value, resting electrocardiographic results, maximum heart rate, exercise-induced angina, Oldpeak referring to ST depression, the slope of ST segment (Bharti et al. 2021). Hence it is difficult for the classical machines to process the huge data, and it has to be preprocessed before applying K-means algorithm. Next section explains various preprocessing steps such as feature selection and reduction techniques, outlier removal techniques, and data visualization.

2.2 Dimensionality reduction technique and feature selection

Increasing dimensionality results in computationally expensive and poor performance of clustering algorithms. The principal component analysis (PCA) and standard MinMax scaler are used to compress higher-dimensional data for dimensionality reduction of the dataset. Setting the dimensionality to 2 has the added benefit that clusters can be visualized appropriately. Standard scaling and MinMax scaling are used to scale the data to a fixed range from +1 to -1. As a result, the standard deviations are minimized.

2.3 Outlier removal technique

As the dataset is not properly distributed and consists of many outliers, data have to undergo preprocessing. Initially, the amount of outliers present in the dataset is detected. These outliers are removed from data using the mathematical function Inter Quartile Range(IQR). Using the quartile method find Q1 and Q3 values, which are then subtracted (Q3–Q1) to obtain IQR. Lower bound(Q1 - 1.5 * IQR) and upper bound (Q3 + 1.5 * IQR) values are used to check outlier and this outlier can be easily removed from data as shown in Fig. 2.

2.4 Data visualization

During preprocessing, the pairs plot visualization technique is used to see the distribution of single variables and the relationship between two variables. This helps to find the most separated clusters. Based on the dataset analysis oldpeak, vessels_colored_by_fluoroscopy, resting_blood_pressure, rest_ecg features are dropped. Figure 3 shows the correlation among features and hence chosen for subsequent analysis. After the data preprocessing, classical K-means approach to form the cluster is applied. Next section details the algorithm used to implement classical K-means clustering.

3 Classical K-means clustering method

K-means clustering (Na et al. 2010) is an unsupervised clustering method first proposed by Lloyd (1982). It is one of the most used clustering techniques. Let $X = x_1, x_2, \ldots, x_N \in \mathbb{R}^M, 1 \leq n \leq N$ consists of “M” number of features. Clustering of datapoints is driven by a set of centroids $W = w_1, w_2, \ldots, w_K, w_n \in \mathbb{R}^M, 1 \leq n \leq K$ generated from the dataset X. Thus set of datapoints is divided into K-clusters $C = C_1, C_2, \ldots, C_K$ by reducing “weighted sum of squared errors” (WSSE) given by “Elbow method” Shi et al. (2021). As shown in Fig. 4, the k value is selected as 2.
The classical K-means algorithm (Huang 1998) described below iterates between two major steps: 1) updating of the clusters and 2) updating of the centroids.

**Algorithm 1 Classical K-means algorithm**

**Input** Datapoints \( X = x_1, x_2, \ldots, x_N \in \mathbb{R}^M \), \( 1 \leq n \leq N \) consists of “M” number of features, set initial centroids \( W = w_1, w_2, \ldots, w_K \), \( w_n \in \mathbb{R}^M, 1 \leq n \leq K \).

**Output** Number of K-clusters consisting of vectors \( C_k \)

repeat

Clustering updation: New datapoint is allocated to the cluster \( C_{k^*} \), \( k^* \) is calculated by:

\[
k^* = \text{argmin}_{1,2,\ldots,K} \| x_n - w_k \|^2
\]

Centroid updation: Centroid \( w_k \) is updated for each cluster \( C_k \) by:

\[
w_k = \frac{1}{\sqrt{|C_k|}} \sum_{n=1}^{N} g_{nk} x_n
\]

until a stopping condition is reached

The time complexity of the algorithm is specified by \( O(LNpK) \), where \( L \) = number of iterations taken to form clusters, \( N \) = number of datapoints, \( p \) = dataset dimension, \( K \) = centroids. Since the time taken to process such a huge data is more in classical machines, we now analyze the working of clustering processing using quantum systems. Next section brieflys the quantum K-means clustering approach in detail.

### 4 Quantum K-means clustering

The initial step to achieve the quantum machine learning process is to convert the data points into quantum states (Ben-lamine et al. 2020). Following that, the distance between each state’s centroids is computed. In terms of quantum learning, estimating the distance between quantum states is completely different from classical learning. The distance computation in the classical version is performed using coordinates, whereas in quantum learning the distance calculation is exhibited in terms of the probabilistic nature of qubits. Although quantum learning allows for the measurement and quantification of notions such as phase differences and amplitudes of different probabilities, the calculation of the distance between two states is unfortunately not achievable due to quantum noise and instability (Gong et al. 2021).

Set \( | \phi \rangle \) to be the centroid cluster in the chosen dataset and \( | \psi \rangle \) to be the new datapoint used to find which cluster it belongs to. To find the distance between these two vectors, one more ancillary qubit in state \( | 0 \rangle \) is used. A Hadamard gate is appended to qubit \( | 0 \rangle \) to bring it to a superposed state. Using swap gate controlled on the auxiliary qubit, the distance from the state \( | \phi \rangle \) to the entangled state \( | \psi \rangle \) of the ancillary qubit is computed. The ancillary qubit is then subjected to another Hadamard gate before being measured. After determining the distance between the states \( | \phi \rangle \) and \( | \psi \rangle \) and the centroid, a swap test circuit is used to assign each state to its nearest cluster. Iterative steps are carried out by updating new centroids of each cluster till the stop criterion is satisfied. Quantum K-means clustering is concluded in three subroutines: 1) distance calculation using swap test circuit, 2) cluster updation, and 3) centroid updation.

After preprocessing steps (similar to classical method), data preparation and distance calculation is followed to convert classical data to quantum state and process. In quantum systems, due to the limited memory, the original data cannot be used directly to process clustering. Therefore the data is normalized and then data preparation methods are applied. Next sections explains the different data preparation methods and distance calculation method to update cluster and centroid.

### 4.1 Data preparation and distance calculation using swap test circuit

Prior to computing the distance between two vectors, it is necessary to encode the data from classical to quantum state. By using the rotation gate each data point is assigned to a specific point on the Bloch-sphere. On chosen heart disease dataset, many state-preparation approaches are applied and tested. Various types of state preparation methods are-

1. **U3 rotation gate (one rotation angle):** Where rotation angle \( \theta \) is determined by the values of data point features (Shende et al. 2006).

\[
\tan \theta = \frac{\alpha}{\beta} \Rightarrow \theta = \arctan \left( \frac{\alpha}{\beta} \right)
\]

Then rotations are applied to \( U_3(\theta, 0, 0) \) to match the datapoints.

\[
U_3(\theta, 0, 0) = \begin{pmatrix}
\cos \frac{\theta}{2} & \sin \frac{\theta}{2} \\
\sin \frac{\theta}{2} & \cos \frac{\theta}{2}
\end{pmatrix}
\]

2. **U3 rotation gate (two rotation angles):** Rotation angles \( \theta \) and \( \varphi \) are used to map the data points. (Liu et al. 2010).

\[
\theta = (\alpha + 1) \frac{\pi}{2}, \varphi = (\beta + 1) \frac{\pi}{2}
\]
Then rotations are applied to $U_3(\theta, \varphi, 0)$ to encode these angles into qubits.

$$U_3(\theta, \varphi, 0) = \begin{pmatrix} \cos \frac{\theta}{2} & \sin \frac{\theta}{2} e^{i\varphi} \\ e^{i\varphi} \sin \frac{\theta}{2} & e^{i\varphi} \cos \frac{\theta}{2} \end{pmatrix}$$

(5)

3. Ry and Rz rotations: To transform all the qubit-state $|\chi\rangle$ to the $|0\rangle$, initial rotation is applied to Z-axis and then to Y-axis.

$$U|\chi\rangle = R_y(\theta)R_z(\varphi)|\chi\rangle = re^{i\theta}|0\rangle \tag{6}$$

This analogue encoding task of mapping data into the large Hilbert space is referred as “Angle encoding” given by

$$X = \prod_{i=1}^{n} \left[ \begin{array}{c} x_1 \\ x_2 \\ \vdots \\ x_2^n \end{array} \right] = \cos(x_i)|0\rangle + \sin(x_i)|1\rangle \tag{7}$$

In a quantum context, datapoint assignment to the relative cluster is achieved by the overlap function, which measures...
the state of overlap between two vectors (Zhang and Ni 2020). Outputting “1” indicates two vectors are identical, and “0” indicates they are orthogonal or no overlap found. Turning this analogy to quantum part requires a swap test that results in a squared probability of the overlap shown in Fig. 5.

Initially, two quantum states \(|\psi\rangle\) and \(|\phi\rangle\) and one ancillary qubit \(|0\rangle\) are chosen, and its outer product is given as

\[ |S_0\rangle = |0, |\psi\rangle, |\phi\rangle\] = |0\rangle \otimes |\psi\rangle \otimes |\phi\rangle \tag{8} \]

Applying Hadamard gate to the ancillary qubit, a superposition state is obtained as:

\[ |S_1\rangle = \frac{1}{\sqrt{2}} (|0, |\psi\rangle, |\phi\rangle + |1, |\psi\rangle, |\phi\rangle) \tag{9} \]

The controlled “SWAP gate” is then applied to the states \(|\psi\rangle\) and \(|\phi\rangle\) in order to swap the states \(|\psi\rangle\) and \(|\phi\rangle\) with the target qubit \(|1\rangle\). This results in

\[ |S_2\rangle = \frac{1}{\sqrt{2}} (|0, |\psi\rangle, |\phi\rangle + |1, |\phi\rangle, |\psi\rangle) \tag{10} \]

Another “Hadamard gate” is placed in the ancillary qubit at the end, resulting in:

\[ |S_3\rangle = \frac{1}{\sqrt{2}} \left( |0, |\psi, \phi\rangle + |1, |\phi, \psi\rangle \right) \tag{11} \]

Expanding the above equation results in:

\[ |S_3\rangle = \frac{1}{2} |0\rangle (|\psi, \phi\rangle + |\phi, \psi\rangle) + \frac{1}{2} |1\rangle (|\psi, \phi\rangle - |\phi, \psi\rangle) \tag{12} \]

Finally, the measurement is applied to the ancillary qubit (Benlamine et al. 2020). The probability of getting state \(|0\rangle\) is given by,

\[ P(|0\rangle) = \left| \frac{1}{2} (|\psi, \phi\rangle + |\phi, \psi\rangle) \right|^2 + \left| \frac{1}{2} (|\psi, \phi\rangle - |\phi, \psi\rangle) \right|^2 \]

\[ = \frac{1}{4} (|\langle \psi, \phi \rangle + \langle \phi, \psi \rangle|^2) \]

\[ = \frac{1}{4} \left( (\langle \phi | \psi \rangle + \langle \psi | \phi \rangle) (|\psi, \phi\rangle + |\phi, \psi\rangle) + (\langle \psi | \phi \rangle + \langle \phi | \psi \rangle) (|\phi, \psi\rangle + |\psi, \phi\rangle) \right) \tag{13} \]

If the probability value \(P(|0\rangle) = 0.5\), the two states \(\phi\) and \(\psi\) are orthogonal or have no overlap, and if \(P(|0\rangle) = 1\), the two states are identical. Similarly, \(P(|1\rangle)\) is specified by

\[ P(|1\rangle) = \frac{1}{2} - \frac{1}{2} \left| \langle \psi | \phi \rangle \right|^2 \tag{14} \]

Furthermore, distance between 2 states is computed by

\[ Dist = d^2 r_{max}^4 (2P(|0\rangle - 1)) \tag{15} \]

where \(d\) = input nonzero data points, \(r_{max}\) = maximum limit of any feature in the data. Thus the overlap function leads a foundation to perform clustering tasks. To form the final cluster, after every distance calculation cluster and centroid
Cluster updation are performed iteratively. This cluster and centroid updation is similar to classical method summarized in next subsections.

4.2 Cluster updation

The nearest-neighbor function is used to allocate the new data point to the closest cluster after computing the distance from the new data point to each cluster centroid (Sergioli et al. 2018). Cluster assignment is given by

\[ C_k = \text{Cluster}(X_n) = \arg \min_{k=1}^{K} \sum_{n \in C_k} g_{nk} || x_n - w_k ||^2 \]

Centroid updation: Centroid \( w_k \) of each cluster is updated by

\[ w_k = \frac{1}{\sqrt{|C_k|}} \sum_{n=1}^{N} g_{nk} x_n \]

These three phases are performed until the cluster centroid’s position does not change, indicating convergence. The algorithm for Q K-means is summarized below:

Algorithm 2 Quantum K-means algorithm

**Input:** Convert input datapoints \( X = x_1, x_2, ..., x_N, x_n \in \mathbb{R}^M, 1 \leq n \leq N \) into quantum states \( |\psi \rangle \) using angle encoding. Initial cluster centroid: \( |w_1 \rangle, |w_2 \rangle, ..., |w_K \rangle \).

**Output:** \( K \) clusters consisting of vectors \( C_k \)

**Distance calculation using swap test:**

\[ D_{\text{qubits}} = d^2 r_{\text{max}} (2 P(0) - 1) \]

**repeat**

**Clustering updation:** Nearest-neighbor function is used to assign each datapoint to nearest cluster. Neighbour function:

\[ C_k = \text{Cluster}(X_n) = \arg \min_{k=1}^{K} \sum_{n \in C_k} g_{nk} || x_n - w_k ||^2 \]

**Centroid updation:** Centroid \( w_k \) of each cluster is updated by

\[ w_k = \frac{1}{\sqrt{|C_k|}} \sum_{n=1}^{N} g_{nk} x_n \]

**until** Convergence is reached

5 Result and analysis

The performance measures between the two approaches were evaluated in terms of accuracy, precision, specificity, and F1-score. The evaluating parameters are given below:

Accuracy is the measure of the system’s ability to make correct predictions. It is given by

\[ \text{Accuracy} = \left( \frac{\text{Correct predictions}}{\text{Total predictions}} \right) \times 100 \]

The system ability to make correct positive predictions is given by sensitivity and correct negative predictions are given by specificity

\[ \text{Sensitivity} = \left( \frac{\text{True positives}}{\text{True positives} + \text{False negatives}} \right) \times 100 \]

\[ \text{Specificity} = \left( \frac{\text{True negatives}}{\text{True negatives} + \text{False positives}} \right) \times 100 \]

Precision evaluates the system capability to produce only relevant results.

\[ \text{Precision} = \left( \frac{\text{True positives}}{\text{True positives} + \text{False positives}} \right) \times 100 \]

F1-score gives the harmonic mean of precision and sensitivity. A higher F1-score represents better modeling.

\[ F1\text{ score} = 2 \times \left( \frac{\text{Sensitivity} \times \text{Precision}}{\text{Sensitivity} + \text{Precision}} \right) \]

Initially Classical K-means approach for predicting heart disease was implemented on the original dataset and results are tabulated. The normalization steps such as standard scaling, min-max scaling, and PCA are applied clean and high-quality data with reduced features, which promises good performance metrics. Finally, to have better clustering outliers rejection (OR) is performed and clustering is evaluated to check the performance. The results obtained are shown in Table 2.

Similarly, to check the performance of clustering of the proposed method, the methodology flow shown in Sect. 2 is followed. The number of gates that can be executed and the number of qubits that can be used in a given algorithm are both limited in IBM Q experience. This dataset is preprocessed before distance calculation for clustering. By using principal component analysis (PCA), the heart disease dataset consisting of 1025 features are reduced to 2 features. Quantum computers could also be used for dimensionality reduction algorithms such as (Cong and Duan 2016; Lloyd et al. 2014). Implementation and execution of the
Table 2 Classical K-means performance for heart disease dataset

| Data Preparation Method | Accuracy | Precision | Sensitivity | Specificity | F1-score |
|-------------------------|----------|-----------|-------------|-------------|----------|
| Original                | 85       | 87        | 85          | 86          | 86       |
| After normalization     | 93       | 93        | 95          | 94          | 94       |
| After outlier rejection | 94       | 95        | 94          | 94          | 93       |

Table 3 Performance measure for different data preparation methods

| Data Preparation Method | Accuracy | Precision | Sensitivity | Specificity | F1-score |
|-------------------------|----------|-----------|-------------|-------------|----------|
| Rz,Ry                   | 90       | 91        | 88          | 90          | 91       |
| U3(θ,0,0)               | 93       | 92        | 91          | 93          | 92       |
| U3(θ,ϕ,π)               | 95       | 95        | 94          | 92          | 94       |

Table 4 Quantum K-means performance for heart disease dataset

| Data Preparation Method | Accuracy | Precision | Sensitivity | Specificity | F1-score |
|-------------------------|----------|-----------|-------------|-------------|----------|
| After normalization     | 95       | 95        | 94          | 92          | 94       |
| After outlier rejection | 96.4     | 96.8      | 95          | 95          | 96.4     |

Table 5 Preprocessing time between classical and quantum K-means

| Data Preparation Method | Classical K-means | Quantum K-means |
|-------------------------|-------------------|-----------------|
| Original                | 1.49              | –               |
| After normalization     | 1.40              | 0.68            |
| After outlier rejection | 1.38              | 0.45            |

The processing time in quantum is faster compared to classical machines. Indeed, traditional machine learning approaches take polynomial time to process and classify huge set of vector numbers in high-dimensional domains. In contrast, quantum computers may alter high-dimensional vectors in the tensor product spaces in logarithmic time (Khan et al. 2019). Table 5 shows the comparison of two algorithms based on processing time.

From running the quantum K-means method using different rotation methods, the best obtained performance measure is selected for further processing. Due to the limitations imposed on quantum computers, initially, data have to be preprocessed by normalization and then Q-K means is applied. Same as classical, quantum K-means method is applied to data after normalization and after outlier rejection, and results obtained in quantum laboratories from Qiskit are tabulated in Table 4.

Finally, the classical and quantum K-means approach is compared to present the performance evaluation shown in Fig. 6.

The proposed approach outperforms all compared state-of-the-art methods in the detection of heart disease. For the chosen heart disease prediction, the original data scatter and processed cluster steps are shown in Fig. 7; simi-
Quantum K-means clustering method for detecting heart disease using quantum circuit approach

Table 6 Comparative analysis of proposed work with state-of-the-art methods

| State-of-the-art           | Methodology                                | Accuracy | Precision | Sensitivity | F1-score |
|----------------------------|--------------------------------------------|----------|-----------|-------------|----------|
| Kuruvilla and Balaji (2021)| Correlation based feature selection (CBFS) | 84.9     | 80.5      | 84.9        | 79.1     |
| Kannan and Vasanthi (2019) | Logistic regression (LR)                   | 86.8     | 81.2      | 84.3        | 82.4     |
| Gao et al. (2021)          | Bagging ensemble learning with DT and PCA   | 83       | 88        | 81          | 85       |
| Kumar et al. (2021)        | Quantum KNN (QKNN)                         | 88       | 87        | 90          | 89       |
| Gupta et al. (2021)        | QML-VQC                                    | 86.3     | 74.4      | 85.2        | 79.2     |
| Proposed work              | Classical K-means                          | 94       | 95        | 94          | 94       |
| Proposed work              | Quantum K-means                            | 96.4     | 96.8      | 94.3        | 96.4     |

Fig. 7 Classical K-means clustering. a original scattered dataset, b clustering process iteration 1, c clustering process iteration 2, d final clustered data

larly, the clustering process carried out in the quantum system is shown in Fig. 8.

Clustering achieved between classical and quantum shows relatively comparable results. Quantum k-means executed in the Qasm simulator exhibits less execution time compared to the classical system. This proves the supremacy of quantum computers in terms of speed. As the dataset becomes huge, quantum systems perform better than classical in terms of speed. The quantum circuit used to perform clustering is shown in Fig. 9.

Some of the limitations imposed while processing the data from classical to quantum are-

- Number of available qubits and gate operations are limited. Therefore the original data is preprocessed and then fed into the system.
- Lack of noise immunity in quantum operations. This small imperfections in the input data or gate operations leads to erroneous output.
- Rotational gates used are more likely prone to error. Any wrong rotations can cause error in the output.
- Quantum decoherence caused by heat and light results in loss of qubit’s quantum entanglement property thereby losing the stored data.
These limitations of the quantum computers provide inadvertent results for quantum algorithms with a larger number of quantum operations. To resolve this limitations, one has to design the quantum circuits with shallow depth. Therefore, the proposed quantum K-means approach to detect heart disease is implemented with simpler quantum circuits resulting in better results even on NISQ computers.

6 Conclusion

The possibility of quantum computers to give exponential speedups has prompted substantial research into quantum algorithms. Quantum computers currently have some constraints in terms of quantum noise and qubit coherence times. These impediments limit their ability to solve problems with reasonable precision. In this article, heart disease prediction is presented using the classical and quantum K-means method. Quantum K-means clustering is performed utilizing a set of quantum gates consisting of less number quantum operations thereby resulting in significant improvement in terms of accuracy and execution time. In contrast to traditional K-means clustering, which takes $O(LNpK)$ time, quantum mechanical K-means clustering takes only $O(LNK)$ time. This might save a lot of time in data mining applications, where data dimensions are typically in the hundreds.

In the future, the developed Q K-means approach can be extended to explore other healthcare datasets. Based on the
analysis quantum machine learning offers a wide range of applications in the medical industry. Despite the fact that quantum machine learning increases computing performance and can handle data stored in a variety of ways, it still has limits. Qubits lose their quantum entanglement property due to quantum decoherence produced by heat and light, resulting in data loss. Also, rotations of various quantum gates are prone to generate inaccuracy and produce erroneous output. Likewise, quantum algorithms are constrained by certain simulations, which limit their use. With the fast-moving progress in technology, it is optimistic that this technique will find a lot of advancement beyond its classical counterparts.

Author Contributions All authors contributed to the study conception and design.

Funding The work presented came from the Ph.D. research undertaken at Department of Electronics and Communication Engineering, the National Institute of Engineering, Mysuru, Karnataka, India.

Availability of data and material Yes, data and material for transparency will be made available.

Code Availability N/A.

Conflict of interest There is no conflict of interests that are directly or indirectly related to the work submitted for publication.

References

Abdel-Basset M, Gamal A, Manogaran G, Son LH, Long HV (2020) A novel group decision making model based on neurosophistic models for heart disease diagnosis. Multimedia Tools Appl 79(15):9977–10002

Acampora G (2019) Quantum machine intelligence. Springer

Aimeur E, Brassard G, Gambs S (2013) Quantum speed-up for unsupervised learning. Mach Learn 90(2):261–287

Aleksandrowicz G, Alexander T, Barkoutos P, Bello L, Ben-Haim Y, Bucher D, Cabrera-Hernández FJ, Carballo-Franquis J, Chen A, Chen C-F, et al (2019) Qiskit: An open-source framework for quantum computing. Accessed on: Mar, 16

Al-Yarimi FAM, Munassar NMA, Bamashmos MHM, Ali MYS (2021) Feature optimization by discrete weights for heart disease prediction using supervised learning. Soft Comput 25(3):1821–1831

Arthur D et al (2021) Balanced k-means clustering on an adiabatic quantum computer. Quant Inf Process 20(9):1–30

Bakhsh AA (2021) High-performance in classification of heart disease using advanced supercomputing technique with cluster-based enhanced deep genetic algorithm. J Supercomput , pp 1–22

Benlamine, K., Bennani, Y., Grozavu, N., Matei, B. (2020) Quantum collaborative k-means. In: 2020 International joint conference on neural networks (IJCNN), pp 1–7, IEEE

Bharti R, Khamparia A, Shahab M, Dhiman G, Pande S, Singh P (2021) Prediction of heart disease using a combination of machine learning and deep learning. Comput Intell Neurosci

Biamonte J, Wittek P, Pancotti N, Rebentrost P, Wiebe N, Lloyd S (2017) Quantum machine learning. Nature 549(7671):195–202

Casaña-Eslava RV, Lisboa PJ, Ortega-Martorell S, Jarman IH, Martín-Guerrero JD (2020) Probabilistic quantum clustering. Knowl Based Syst 194:105567

Cong I, Duan L (2016) Quantum discriminant analysis for dimensionality reduction and classification. New J Phys 18(7):73011

Farouk MH (2017) On the application of quantum clustering on speech data. Int J Speech Technol 20(4):891–896

Gangal K (2021) Heart disease dataset uci (2021). https://www.kaggle.com/ketangangal/heart-disease-dataset-uci?select=HeartDiseaseTrain-Test.csv

Gao X-Y, Amin Ali A, Shahab Hassan H, Anwar EM (2021) Improving the accuracy for analyzing heart diseases prediction based on the ensemble method. Complexity

Gong C, Dong Z, Gani A, Qi H (2021) Quantum k-means algorithm based on trusted server in quantum cloud computing. Quant Inf Process 20(4):1–22

Gupta H, Varshney H, Sharma TK, Pachauri N, Verma OP (2021) Comparative performance analysis of quantum machine learning with deep learning for diabetes prediction. Complex Intell Syst, pp 1–15

Haq AU, Li JP, Memon MH, Nazir S, Sun R (2018) A hybrid intelligent system framework for the prediction of heart disease using machine learning algorithms. Mobile Intell Syst

Huang Z (1998) Extensions to the k-means algorithm for clustering large data sets with categorical values. Data Min Knowl Disc 2(3):283–304

Kannan R, Vasanthi, V (2019) Machine learning algorithms with roc curve for predicting and diagnosing the heart disease. Soft Comput Med Bioinf, pp 63–72. Springer

Katarya R, Meena SK (2021) Machine learning techniques for heart disease prediction: a comparative study and analysis. Health Technol 11(1):87–97

Keogh EJ, Mueen A (2017) Curse of dimensionality. Encycl Mach Learn Data Min 2017:314–315

Kerenidis I, Landman J, Luongo A, Prakash A (2018) q-means: a quantum algorithm for unsupervised machine learning. arXiv preprint arXiv:1812.03584

Khan SU, Awan AJ, Vall-Llosa G (2019) K-means clustering on noisy intermediate scale quantum computers. arXiv preprint arXiv:1909.12183

Khan M, Hussain I, Jamal SS, Amin M (2019) A privacy scheme for digital images based on quantum particles. Int J Theor Phys 58(12):4293–4310

Kumar Y, Koul A, Sisodia PS, Shafi J, Kavita V, Gheisari M, Davoodi MB (2021) Heart failure detection using quantum-enhanced machine learning and traditional machine learning techniques for internet of artificially intelligent medical things. Wireless Commun Mobile Comput

Kuruvilla AM, Balaji N (2021) Heart disease prediction system using correlation based feature selection with multilayer perceptron approach. In: IOP conference series: materials science and engineering, vol 1085, p 012028 (2021). IOP Publishing

Lamata L (2021) Quantum reinforcement learning with quantum photonics. In: Photonics, vol 8, p 33 . Multidisciplinary Digital Publishing Institute

Li Q, He Y, Jiang J-p (2011) A hybrid classical-quantum clustering algorithm based on quantum walks. Quant Inf Process 10(1):13–26

Liu, W., Chen, H., Yan, Q., Liu, Z., Xu, J., Zheng, Y (2010) A novel quantum-inspired evolutionary algorithm based on variable angle-distance rotation. In: IEEE congress on evolutionary computation, pp 1–7 (2010). IEEE
Lloyd S (1982) Least squares quantization in pcm. IEEE Trans Inf Theory 28(2):129–137
Lloyd S, Mohseni M, Rebentrost P (2014) Quantum principal component analysis. Nat Phys 10(9):631–633
Mohan S, Thirumalai C, Srivastava G (2019) Effective heart disease prediction using hybrid machine learning techniques. IEEE Access 7:81542–81554
Möller M, Vui C (2017) On the impact of quantum computing technology on future developments in high-performance scientific computing. Ethics Inf Technol 19(4):253–269
Na S, Xumin L, Yong G (2010) A new clustering algorithm: An improved k-means clustering algorithm. In: 2010 Third international symposium on intelligent information technology and security informatics, pp 63–67. IEEE
Ramezani SB, Sommers A, Manchukonda HK, Rahimi S, Amiratifi A (2020) Machine learning algorithms in quantum computing: A survey. In: 2020 International joint conference on neural networks (IJCNN), pp 1–8. IEEE
Rani P, Kumar R, Ahmed NMS, Jain A (2021) A decision support system for heart disease prediction based upon machine learning. J Reliable Intell Environ, pp 1–13
Rubio E, Castillo O, Valdez F, Melin P, Gonzalez CI, Martinez G (2017) An extension of the fuzzy possibilistic clustering algorithm using type-2 fuzzy logic techniques. Adv Fuzzy Syst
Sarma A, Chatterjee R, Gili K, Yu T (2019) Quantum unsupervised and supervised learning on superconducting processors. arXiv preprint arXiv:1909.04226
Sergioli G, Santucci E, Didaci L, Miszczak JA, Giuntini R (2018) A quantum-inspired version of the nearest mean classifier. Soft Comput 22(3):691–705
Shao C, Li Y, Li H (2019) Quantum algorithm design: techniques and applications. J Syst Sci Complex 32(1):375–452
Shende VV, Bullock SS, Markov IL (2006) Synthesis of quantum-logic circuits. IEEE Trans Comput Aided Des Integrated Circuits Syst 25(6):1000–1010
Shi C, Wei B, Wei S, Wang W, Liu H, Liu J (2021) A quantitative discriminant method of elbow point for the optimal number of clusters in clustering algorithm. EURASIP J Wireless Commun Netw 2021(1):1–16
Singh P, Bose SS (2021) A quantum-clustering optimization method for COVID-19 CT scan image segmentation. Exp Syst Appl 185:115637
Wettschereck D, Aha DW, Mohri T (1997) A review and empirical evaluation of feature weighting methods for a class of lazy learning algorithms. Artif Intell Rev 11(1):273–314
Xu D, Tian Y (2015) A comprehensive survey of clustering algorithms. Ann Data Sci 2(2):165–193
Yang M-S, Nataliani Y (2017) A feature-reduction fuzzy clustering algorithm based on feature-weighted entropy. IEEE Trans Fuzzy Syst 26(2):817–835
Yao Z, Peng W, Gao-yun C, Dong-Dong C, Rui D, Yan Z (2008) Quantum clustering algorithm based on exponent measuring distance. In: 2008 IEEE international symposium on knowledge acquisition and modeling workshop, pp 436–439. IEEE
Zhang Y, Ni Q (2020) Recent advances in quantum machine learning. Quant Eng 2(1):34

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.