Accredited SINTA 2 Ranking
Decree of the Director General of Higher Education, Research, and Technology, No. 158/E/KPT/2021
Validity period from Volume 5 Number 2 of 2021 to Volume 10 Number 1 of 2026

JURNAL RESTI
(Rekayasa Sistem dan Teknologi Informasi)
Vol. 8 No. 6 (2024) 710 - 718 e-ISSN: 2580-0760

Manhattan Distance-based K-Medoids Clustering Improvement for Diagnosing Diabetic Disease

Solikhun[1]*, Muhammad Rahmansyah Siregar[2], Lise Pujiastuti[3], Mochamad Wahyudi[4]
[1,2]Informatics Engineering, STIKOM Tunas Bangsa, Pematangsiantar, Indonesia
[3]Information System, STMIK Antar Bangsa, Tangerang, Indonesia
[4]Computer Science, Faculty of Engineering and Informatics, Universitas Bina Sarana Informatika, Jakarta, Indonesia
[1]solikhun@amiktunasbangsa.ac.id, [2]rahmansyahsiregar77@gmail.com, [3]lise.pujiastuti@gmail.com, [4]wahyudi@bsi.ac.id

Abstract

Diabetes is a metabolic disorder characterized by blood glucose levels above normal limits. Diabetes occurs when the body is unable to produce sufficient insulin to regulate blood sugar levels. As a result, blood sugar management becomes impaired and there is no cure for diabetes. Early detection of diabetes provides an opportunity to delay or prevent its progression into acute stages. Clustering can help identify patterns and groups of diabetes symptoms by analyzing attributes that indicate these symptoms. In this study, researchers are using K-Medoid and Quantum K-Medoid methods for clustering diabetes data. Quantum computing utilizes quantum bits, or qubits, which can represent multiple states at the same time. Compared to classical computers, quantum computing has the potential for an exponential speedup in problem-solving. Researchers conducted a comparison between two methods: the classic K-Medoids method and the K-Medoids method utilizing quantum computing. The researchers found that both Quantum K-Medoid and Classic K-Medoid achieved the same clustering accuracy of 91%. In testing with the Quantum K-Medoids algorithm, it was found that the cost value in the 8th epoch showed a significant decrease compared to the Classical K-Medoids algorithm. This demonstrates that Quantum K-Medoid can be considered a viable alternative for clustering purposes.

Keywords: *Diabetes, K- Medoid; Manhattan Distance; Quantum Computing; Quantum Bit*

How to Cite: Solikhun, M. Rahmansyah Siregar, L. Pujiastuti, and M. Wahyudi, "Manhattan Distance-based K-Medoids Clustering Improvement for Diagnosing Diabetic Disease", *J. RESTI (Rekayasa Sist. Teknol. Inf.)*, vol. 8, no. 6, pp. 710 - 718, Dec. 2024.
DOI: https://doi.org/10.29207/resti.v8i6.5894

1. Introduction

Diabetes is a metabolic disorder characterized by blood glucose levels above normal limits [1]. In diabetes, the human body is unable to create enough insulin to manage blood sugar levels, or the insulin produced is insufficiently utilized [2]–[4]. Currently, diabetes cannot be cured. However, early detection of diabetes provides a crucial opportunity to delay or prevent its progression into acute stages [5]. One of the techniques used in data processing is clustering [6]. Clustering is an unsupervised data mining method used to group data that does not have labels [7]. It can categorize data into several clusters based on similarities [8], [9]. In the context of diabetes, clustering can help identify patterns and groups of diabetes symptoms based on attributes that indicate these symptoms. This can aid in understanding diabetes subtypes, leading to more personalized and effective treatment plans.

In the current study, researchers utilized the k-medoid with the Manhattan distance algorithm, enhanced by quantum computing. The k-medoid algorithm is a clustering algorithm similar to k-means but more resilient against outliers and noise [10]–[13]. Manhattan Distance is used to calculate the absolute difference between the coordinates of a pair of objects [14], [15]. Manhattan Distance is more sensitive to differences between features and more robust to outliers. It generally achieves better accuracy compared to Euclidean Distance, which makes Manhattan Distance work quite well for clustering purposes [16], [17]. Using K-Medoids with Manhattan Distance provides more accurate clustering by minimizing sensitivity to outliers and better capturing differences between features. This combination is especially useful for

Received: 29-06-2024 | Accepted: 28-10-2024 | Published Online: 24-12-2024

datasets with non-uniform distributions or when robustness against extreme values is essential.

Quantum computing offers a promising solution for optimizing healthcare by efficiently executing complex algorithmic instructions [18]. quantum mechanism utilizes quantum bits, or qubits, which can represent multiple states at the same time, unlike classical computers [19], [20]. Classic bits, which can only be 0 or 1, qubits can exist in states other than |0⟩ or |1⟩, making them distinct from conventional [21]–[24]. Compared to classical, quantum computing has the potential for an exponential speedup in problem-solving [25]. The research aimed to utilize the quantum k-medoid algorithm for clustering diabetes data and to assess its performance using quantum mechanics. Several research studies have been conducted on clustering for diabetes:

In the study [26], the researchers compare Manhattan and Euclidean distance metrics to assess the accuracy of k-means clustering in categorizing heart disease data. The study uses a dataset containing cholesterol and age attributes to determine how the choice of distance metric affects clustering outcomes. Using the silhouette coefficient, they found that k-means with Manhattan distance achieved a silhouette score of 0.5374, slightly outperforming Euclidean distance, which scored 0.5355. This indicates that Manhattan distance is marginally more effective.

In the study[27], the researchers aim to improve the accuracy of the K-Medoids algorithm by incorporating a quantum computing approach using the Manhattan distance. The study involves clustering stroke patient data, which includes attributes like age, hypertension, and glucose levels. The researchers compare the classical K-Medoids algorithm with the quantum-enhanced version. The results show that the quantum-enhanced K-Medoids method achieves an accuracy of 64%, outperforming the classical version's 52%. This demonstrates the effectiveness of integrating quantum computing for improved clustering outcomes

Main Reference for this research [28] titled Quality Based Analysis of Clustering Algorithms using Diabetes Data for the Prediction of Disease compares the performance of clustering algorithms k-Means and k-Medoids in analyzing diabetic datasets and predicting diseases. The study focuses on using the final bio-chemistry prescription of diabetic patients for disease identification. The results indicate that the accuracy of the k-means algorithm is 87% and the accuracy of the k-medoids algorithm is 80%, making k-means one of the best techniques for disease prediction.

In study [28], it was found that the accuracy of the K-Medoids algorithm was 80%, which is lower than the K-Means algorithm, which had an accuracy of 87%. The K-Medoid algorithm also has limitations when dealing with noise and outliers. K-Medoid can be affected by unrepresentative data points, as the medoid is selected from the actual data, which can result in suboptimal cluster centers. This contrasts with the Quantum K-Medoid, which employs a quantum-based approach to minimize the impact of outliers and effectively find cluster centers that are more robust to noise.

In this study, researchers conducted a comparison between two methods: the classic K-Medoids method and the K-Medoids method utilizing quantum computing. Both approaches will be assessed based on their accuracy in grouping results for the prediction of diabetes. The gap between this study and the reference study is that the researchers achieved higher clustering accuracy by finding better medoid values from both the K-Medoids and quantum K-Medoids algorithms. The novelty of this study lies in the utilization of quantum computing in conjunction with the K-Medoids method and Manhattan Distance calculations to forecast diabetes.

The contribution from this study is for medical needs, k-medoid placement can provide better results for detecting diabetes, especially with larger data. The results of this approach can provide a short diagnosis that can be used for early detection of diabetes and researchers provide insights into the quantum k-medoid algorithm for diabetes clustering, enabling its application in other clustering tasks.

This study consists of five research structures: introduction, related research, research methods, results, and conclusion. In the related research section, the researchers conduct a literature review to examine the potential of the K-Medoids algorithm for clustering. In the research methods section, the researchers explain the steps undertaken in this study. In the results section, the researchers present the findings of the study, and in the conclusion section, the researchers analyze and summarize the results of the research conducted. The contribution of this research is to enable early detection and identify groups affected by diabetes so that interventions can be quickly implemented to prevent or delay the onset of diabetes.

2. Research Methods

One study titled Clustering and visualization of a high-dimensional diabetes dataset [29]. In this study, researchers explore the use of data clustering and visualization methods to analyze a diabetes dataset. The study evaluates the DBSCAN and K-Means algorithms with various distance measures and identifies several different groups of diabetes patients characterized by different attributes.

In study [30], proposes two algorithms for clustering data, which are the K-medoids and random swap algorithms. The goal of this study is to detect human subjects who share common diseases to potentially predict future illnesses from previous medical history.

This study [31], the study compares the Hierarchical Clustering method and the K-Medoid Clustering method. The objective of this research is to classify Regional Apparatus Work Units. This study tests hierarchical clustering with five methods: Single Linkage, Average Linkage, Complete Linkage, Centroid Method, and Ward Method, comparing them using the correlation coefficient. The Centroid Method achieved the highest value. To determine the best K in K-medoid testing, the researchers used a dendrogram, finding the best K to be 2. The comparison of methods in this study uses the silhouette coefficient, resulting in a silhouette coefficient value of 0.9146 for both Hierarchical and K-medoid methods.

This study [32] utilized the K-Means and K-Medoid algorithms to cluster COVID-19. The Davies-Bouldin Index (DBI) method was used to determine the most optimal data. When testing with K-Means, the best DBI value obtained was 0.139 with k=4. On the other hand, K-Medoid testing produced the best DBI value of 0.197 with k=9. The study concluded that the K-Means algorithm demonstrates better cluster validity compared to the K-Medoid algorithm.

In the study [33] , applied the Divisive Analysis Clustering method to categorize diabetes patients based on their medical records. The study identified two optimal clusters with a silhouette coefficient of 0.468582, indicating a well-structured clustering. Age (X2) and blood sugar level (X8) were found to be the key variables contributing to diabetes, as these variables showed the highest average values across the clusters.

This study [34], the study focuses on determining the optimal value for K to cluster E-Learner using the K-Medoid algorithm. This study uses two methods, namely the Elbow and Silhouette methods. The cluster size returned by the Elbow method for the given dataset is K=5. The cluster size returned by the Silhouette method for the given dataset is K=10. This study concludes that the Silhouette method is the optimal method for validating cluster size for the given dataset.

This study [35], the K-Medoid algorithm and the RapidMiner application were utilized to analyze sales data for transportation equipment. The study aimed to determine the most effective value for K by comparing several options. The accuracy of the clusters was assessed using the Davies-Bouldin Index (DBI) value. The study concluded that the optimal value for K is 2, with a corresponding DBI value of -0.838.

This study [36], the study uses the K-Means and K-Medoid algorithms to analyze the spread of the Covid-19 virus in Indonesia. The Davies-Bouldin Index (DBI) value is used as the evaluation parameter, utilizing Python Version 3 programming language executed on the Jupyter Notebook tool. This study compares several K values from the two methods to find the optimal value. In the test with K-Means, the optimal K found was K=2 with a DBI value of 0.9762331449809145, while for the K-medoid test, the optimal K value was K=2 with a DBI value of 0.9809235412405508. This study concludes that K-Means produces better clustering than the K-Medoid method.

This study [37], the study uses the K-Medoid algorithm to classify internet usage at Wahidin Vocational High School in Cirebon City. The research uses the RapidMiner application to perform clustering. This study compares K values to find the optimal value, with the optimal K found to be K=4 with a DBI value of -0.974.

This study [38], audiometry results were clustered using the K-Medoids algorithm, implemented with the RapidMiner application. The performance of the K-Medoid algorithm in clustering was evaluated using a confusion matrix, which revealed an accuracy of 28.3%, a precision of 64.3%, and a recall of 21.4%.

In the study [39], the study compares various algorithms for solving the NP-complete Hamiltonian Cycle Problem (HCP). The study evaluates deterministic, probabilistic, and quantum approaches, with a focus on bounded-degree graphs. Quantum algorithms demonstrated only a quadratic speedup over classical methods, but significant performance improvements were observed under specific graph constraints. The results emphasize that quantum computing holds promise in reducing time complexity, especially with quantum interference techniques.

This study implements 3 main stages as shown in Figures 1,2,3 and 4: data transformation to Bits, data transformation to Qubits, and clustering process using Quantum K-Medoid and Classic K-Medoid algorithms. The transformation into bits can be seen in Figure 1. Data transformation is performed to convert data into 1s and 0s, with the goal of transforming it into quantum bits. The transformed data must meet certain conditions to be classified as either 0 or 1.

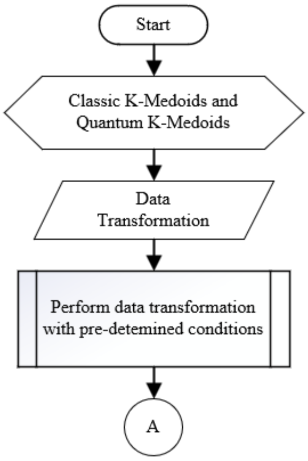


Figure 1. Research Flow Diagram 1

Figure 2, it is shown that if the transformation into bits is successful, the process will proceed to the stage of transforming the data into quantum bits (qubits). In qubits, data is represented as either bra or ket. In this transformation, it will only be represented as ket | >, which will later the state $|0\rangle$ is represented as $\begin{bmatrix}1\\0\end{bmatrix}$ and the state $|1\rangle$ is represented as $\begin{bmatrix}0\\1\end{bmatrix}$.

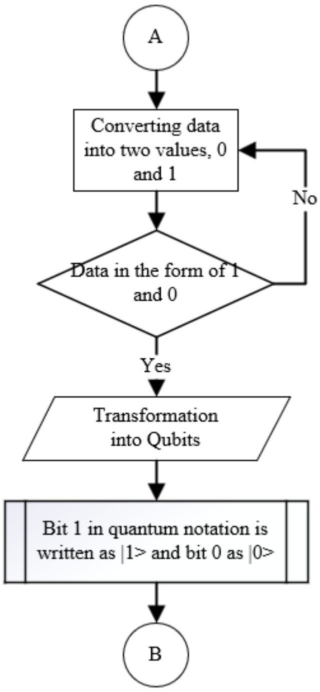


Figure 2. Research Flow Diagram 2

After the data has been successfully converted into qubits and bits, calculations are then performed using classical K-medoids and quantum K-medoids. The calculations are carried out over 14 epochs, with the centroids selected randomly from data points 220 to 246.

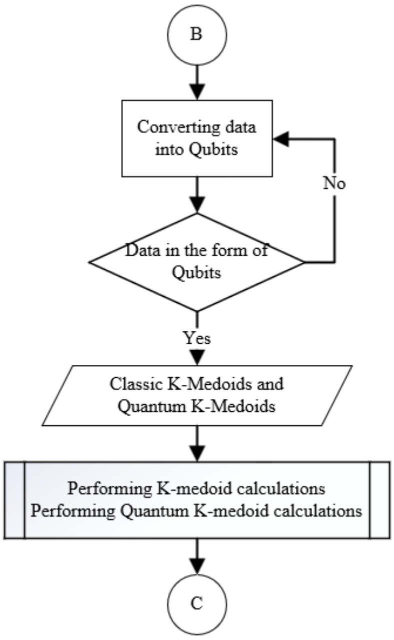


Figure 3. Research Flow Diagram 3

The data from the calculation results using both methods are then compared and evaluated to determine which algorithm is most effective for clustering diabetes data.

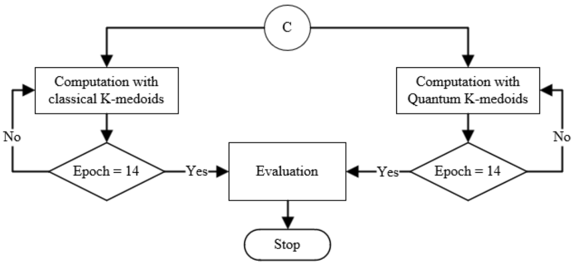


Figure 4. Research Flow Diagram 4

Before presenting the results and discussion, here are some scientific explanations of the methods used

2.1 Data Collection

The data collection stage aims to gather relevant information for clustering diabetes diseases. For this study, the researchers utilized a dataset obtained from the Kaggle website. The focus was on five key attributes related to diabetes: hypertension, heart disease, BMI, HbA1c level, and blood glucose level.

A total of 248 sample data points were used for the clustering process. These attributes were selected to provide a comprehensive analysis of factors associated with diabetes, enabling the identification of patterns and relationships within the dataset.

2.2 Data Transformation

In this stage, the data is transformed into a binary format for further calculations. This process involves converting data, which may have various values, into a format that uses only two values, 0 and 1.

In the diabetes clustering data transformation, BMI is assigned a value of 0 if it falls between 18.5 and 24.9, and a value of 1 if it is greater than 24.9. For HbA1c, a value of 0 is assigned if it is less than 5.7%, and a value of 1 if it exceeds 5.7%. Similarly, blood glucose is assigned a value of 0 if it is below 140, and a value of 1 if it is greater than 140. These rules ensure that diabetes medical record data is standardized and can be used effectively for further analysis and processing in a consistent format

2.3. *Quantum Bit Transformation*

After the data is transformed into a binary format, the next step is to convert it into qubits. Qubits, as the basic unit of information in quantum computing, can store and process information quantumly. This conversion enables the application of the K-Medoids algorithm within a quantum computing system.

The dataset of diabetes undergoes a rule-based transformation process to ensure a standardized representation. In this process, the data is encoded into qubits using values of 0, 1, or both simultaneously, employing Dirac notation with ket (| >) and bra (< |).

2.4 K-Medoid Method

K-Medoid is a clustering method used to group data into several clusters based on the similarity or distance between data points. Unlike other clustering techniques, this method ensures that each cluster is represented by a central point known as a medoid.

A medoid is an actual data point from the dataset that minimizes the total distance to all other points within the same cluster. This characteristic makes K-Medoid robust to outliers and effective for clustering tasks where interpretability and accuracy are important.

2.5 Quantum K-Medoid Method

Quantum K-medoid is a variant of the K-medoid clustering algorithm that leverages the principles of quantum computing. This algorithm aims to cluster data into several groups by selecting data points (medoids) that minimize the total distance to all other points in the cluster. In the quantum context, Quantum K-medoid uses qubits and quantum gates to process information, allowing for parallel exploration of solutions and greater efficiency compared to classical methods. The following is the formula for developing K-Medoids with quantum computing using Manhattan Distance calculations. The Formula of the K-Medoids algorithm with Manhattan Distance is using Equation 1.

$$d(a_x, b_y) = \sum_{z=1}^{n} \left| |a_{az} > - |b_{bz} > \right| \tag{1}$$

Equation 1 represents a quantum distance metric used to measure the difference between two quantum states, $|a_{az} > - |b_{bz} >$.These quantum states are expressed in ket notation, where $|a_{az} >$ represents the quantum state of the z-th element of the vector a_x, and $|b_{bz} >$ represents the quantum state of the z-th element of the medoid vector b_y. The vector a_x, refers to a data point within cluster x, while b_y is the medoid, or central data point, of cluster y, chosen directly from the dataset.

In the medical context, quantum computing can accelerate the analysis of medical data, such as blood test results, to detect diabetes more quickly and accurately. For example, by analyzing various blood parameters (like hemoglobin levels, red blood cell count, etc.) and comparing them with vast medical datasets, quantum computers can assist doctors in diagnosing diabetes faster and determining the specific type of diabetes more precisely.

2.6 Result

At this stage, the results of data clustering are analyzed using both the classical K-Medoids method and the K-Medoids method enhanced with quantum computing. The comparison between these approaches highlights differences in efficiency and accuracy.

This analysis offers valuable insights into the data structure and the patterns identified by the clustering algorithm, helping to understand the impact of quantum computing on clustering performance and revealing potential improvements over the classical method.

2.7 Evaluation

The evaluation stage focuses on assessing the performance of classical methods and the K-Medoids algorithm implemented with quantum computing in clustering diabetes data. This involves comparing the clustering results to evaluate how effectively each method identifies meaningful groupings within the data.

The evaluation utilizes metrics such as accuracy to determine the ability of the clustering model to correctly group diabetes cases based on their similarities or patterns. This helps in understanding the strengths and weaknesses of each approach in capturing relevant data structures.

3. Results and Discussions

3.1 Transformation to Bits

The results of converting the data into binary code can be seen in Table 1.

Table 1. Dataset After Transformation Biner

No	X1	X2	X3	X4	X5	Y
1	0	1	1	1	1	0
2	0	0	1	1	0	0
3	0	0	1	1	1	0
…	…	…	…	…	…	…
…	…	…	…	…	…	…
247	0	0	1	1	1	0
248	0	0	0	1	1	0

3.2 Transformation to Qubits

Transformation into qubits is demonstrated using Table 1. For instance, a sample is taken from dataset number 1, represented by the binary code 011110. This code translates to the following attributes: hypertension = no, heart disease = yes, BMI > 24.9, HbA1c > 5.7%, blood glucose level > 140, and target (diabetes) = no. To convert classical data into qubits, encode each classical bit into the quantum state of the qubit. For example, bit 1 in quantum notation is written as $|1\rangle$ and bit 0 as $|0\rangle$. In matrix notation, a qubit is represented as a column vector. For example, the state $|0\rangle$ is represented as $\begin{bmatrix}1\\0\end{bmatrix}$ and the state $|1\rangle$ is represented as $\begin{bmatrix}0\\1\end{bmatrix}$. then for the first data, namely 011110, it is changed to $\begin{bmatrix}1\\0\end{bmatrix}, \begin{bmatrix}0\\1\end{bmatrix}, \begin{bmatrix}0\\1\end{bmatrix}, \begin{bmatrix}0\\1\end{bmatrix}, \begin{bmatrix}0\\1\end{bmatrix}$ and $\begin{bmatrix}1\\0\end{bmatrix}$. The data in Table 1 is converted into qubit form, as illustrated in Table 2.

Table 2. Quantum Bit Transformation

No	X1	X2	X3	X4	X5	Y
1	$\begin{bmatrix}1\\0\end{bmatrix}$	$\begin{bmatrix}0\\1\end{bmatrix}$	$\begin{bmatrix}0\\1\end{bmatrix}$	$\begin{bmatrix}0\\1\end{bmatrix}$	$\begin{bmatrix}0\\1\end{bmatrix}$	$\begin{bmatrix}1\\0\end{bmatrix}$
2	$\begin{bmatrix}1\\0\end{bmatrix}$	$\begin{bmatrix}1\\0\end{bmatrix}$	$\begin{bmatrix}0\\1\end{bmatrix}$	$\begin{bmatrix}0\\1\end{bmatrix}$	$\begin{bmatrix}1\\0\end{bmatrix}$	$\begin{bmatrix}1\\0\end{bmatrix}$
3	$\begin{bmatrix}1\\0\end{bmatrix}$	$\begin{bmatrix}1\\0\end{bmatrix}$	$\begin{bmatrix}0\\1\end{bmatrix}$	$\begin{bmatrix}0\\1\end{bmatrix}$	$\begin{bmatrix}0\\1\end{bmatrix}$	$\begin{bmatrix}1\\0\end{bmatrix}$
…	…	…	…	…	…	…
247	$\begin{bmatrix}1\\0\end{bmatrix}$	$\begin{bmatrix}1\\0\end{bmatrix}$	$\begin{bmatrix}0\\1\end{bmatrix}$	$\begin{bmatrix}0\\1\end{bmatrix}$	$\begin{bmatrix}0\\1\end{bmatrix}$	$\begin{bmatrix}1\\0\end{bmatrix}$
248	$\begin{bmatrix}1\\0\end{bmatrix}$	$\begin{bmatrix}1\\0\end{bmatrix}$	$\begin{bmatrix}1\\0\end{bmatrix}$	$\begin{bmatrix}0\\1\end{bmatrix}$	$\begin{bmatrix}0\\1\end{bmatrix}$	$\begin{bmatrix}1\\0\end{bmatrix}$

The findings of this research present an alternative model for the K-Medoids clustering method using Manhattan Distance calculations, specifically through a quantum computing approach.

Before implementing the Quantum K-Medoids method, attribute values and medoid values were first converted to ensure compatibility with quantum computing.

3.3 Classic K-Medoid and Quantum K-Medoid Calculation

In this study, diabetes data was clustered using both the Classical K-Medoids and Quantum K-Medoids methods, with Manhattan Distance calculations. The comparison between these methods demonstrated an identical clustering accuracy of 90%. Tables 3 and 4 present the results from testing the data in epoch-1 and epoch-2, offering valuable insights before applying quantum-based clustering techniques.

Table 3. K-Medoids Epoch-1 Test Results

C1	C2	Shortest Distance	Cluster	Data Real	Description
3	1	1	2	0	False
1	1	1	1	0	True
2	0	0	2	0	False
…	…	…	…	…	…
2	0	0	2	0	False
3	1	1	2	0	False
Total Shortest		155	Accuracy		48 %

In the first iteration, using Classical K-Medoids with medoid C1 (data point 220) and medoid C2 (data point 221), the accuracy obtained was 48%. The researcher experimented with 14 medoids to improve the clustering performance, eventually achieving the best results in epoch 14, with an accuracy of 91% and a cost value of 254. The results from the other 14 iterations conducted by the researchers are shown in Table 4.

Table 4. Epochs Performed Using Classical K-Medoid

Epoch	C1	C2	Cost Value	Accuracy
1	220	221	155	48
2	222	223	208	64
3	224	225	290	88
4	226	227	304	91
5	228	229	238	57
6	230	231	196	80
7	232	233	155	52
8	234	235	249	52
9	236	237	196	79
10	238	239	265	38
11	240	241	196	67
12	242	243	283	12
13	244	245	155	52
14	246	247	254	91

As shown in Table 4, although the cost value which measures the assignment of data points to clusters based on the distance between each point and its closest medoid is stable in epochs 1, 2, 3, and 4, the best results are observed in epoch 14. This epoch has a lower cost value than epoch 4, suggesting that the data is relatively well-clustered. The calculations demonstrating the best results of the Classical K-medoids are presented in Table 5.

Table 5. K-Medoids Epoch-2 Test Results

C1	C2	Shortest Distance	Cluster	Data Real	Description
2	2	2	1	0	True
0	2	0	1	0	True
1	1	1	1	0	True
…	…	…	…	…	…
1	1	1	1	0	True
2	2	2	1	0	True
Total Shortest Distance		254	Accuracy		91 %

The calculation for iteration 14 was stopped after the Classical K-Medoids achieved an optimal accuracy of 91% for diabetes clustering. In the quantum calculation, the researchers used the same centroids as in the K-Medoids calculation. The simulation results from testing the K-Medoids algorithm with quantum computing also show an accuracy of 91% with epoch 14. The results of the quantum K-Medoids calculations are presented in Tables 6, 7, and 8.

Table 5. Quantum K-Medoids Epoch-1 Test Results

C1	C2	C1 (Des)	C2 (Des)	Shortest	C	Data Real	Desc
$\begin{bmatrix}3\\3\end{bmatrix}$	$\begin{bmatrix}1\\1\end{bmatrix}$	4.24	1.41	1.41	2	0	False
$\begin{bmatrix}1\\1\end{bmatrix}$	$\begin{bmatrix}1\\1\end{bmatrix}$	1.41	1.41	1.41	1	0	True
$\begin{bmatrix}2\\2\end{bmatrix}$	$\begin{bmatrix}0\\0\end{bmatrix}$	2.83	0.00	0.00	2	0	False
…	…	…	…	…	…	…	…
$\begin{bmatrix}2\\2\end{bmatrix}$	$\begin{bmatrix}0\\0\end{bmatrix}$	2.83	0.00	0.00	2	0	False
$\begin{bmatrix}3\\3\end{bmatrix}$	$\begin{bmatrix}1\\1\end{bmatrix}$	4.24	1.41	1.41	2	0	False
Total Shortest		277.19		Accuracy		48 %	

in the quantum k-medoid calculation in the first iteration, the results are the same as the classical k-medoid, which is 48%, but the total shortest distance value is different. To find out further results, the quantum approach is continued to the 2nd iteration. The researcher will display the calculation results from epochs 1 - 14 using quantum k-medoid in Table 6.

Table 6. The 14 Epochs Performed Using Quantum K-Medoid

Epoch	C1	C2	Cost Value	Accuracy
1	220	221	219.2	48
2	222	223	294.16	64
3	224	225	410.12	88
4	226	227	429.92	91
5	228	229	336.58	57
6	230	231	280.01	80
7	232	233	219.2	52
8	234	235	252.14	52
9	236	237	277.19	79
10	238	239	374.77	38
11	240	241	277.19	67
12	242	243	400.22	12
13	244	245	219.2	52
14	246	247	359.21	91

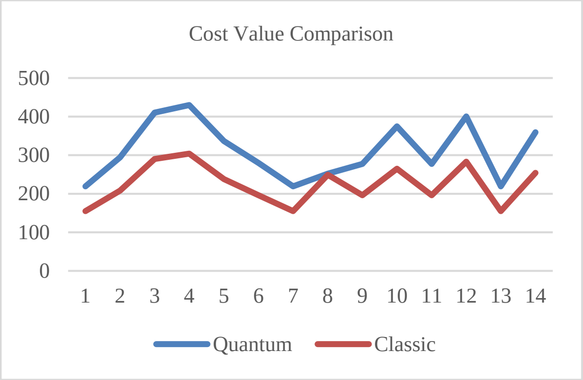
Has a similarity to the Classical K-Medoids, the Quantum K-Medoids also achieves the same accuracy results; however, the cost values differ between the two algorithms. Despite this, the K-Medoids algorithm reaches optimal results at epoch 14, with an accuracy of 91%. The calculations for the best epoch of the Quantum K-Medoids can be found in Table 6.

Table 7. Quantum K-Medoids Epoch-2 Test Results

C1	C2	C1 (Des)	C2 (Des)	Shortest	C	Data Real	Desc
$\begin{bmatrix}2\\2\end{bmatrix}$	$\begin{bmatrix}2\\2\end{bmatrix}$	2.82	2.82	2.82	1	0	True
$\begin{bmatrix}0\\0\end{bmatrix}$	$\begin{bmatrix}2\\2\end{bmatrix}$	0	2.82	0	1	0	True
$\begin{bmatrix}1\\1\end{bmatrix}$	$\begin{bmatrix}1\\1\end{bmatrix}$	1.41	1.41	1.41	1	0	True
…	…						
…	…						
$\begin{bmatrix}1\\1\end{bmatrix}$	$\begin{bmatrix}1\\1\end{bmatrix}$	1.41	1.41	1.41	1	0	True
$\begin{bmatrix}2\\2\end{bmatrix}$	$\begin{bmatrix}2\\2\end{bmatrix}$	2.83	2.83	2.83	1	0	True
Total Shortest		359.21		Accuracy		91 %	

The diagram in Figure 2 compares the cost values between the Quantum K-Medoids and Classical K-Medoids algorithms. The comparison shows similar trends in the increase and decrease of cost values, but in the 8th epoch, the Quantum K-Medoids displays a more significant reduction in cost value compared to the Classical K-Medoids. Unfortunately, the accuracy in the 8th epoch is only 52%, which is not precise enough for effective data clustering.

Figure 2. Cost Value Comparison Diagram



The findings of this research present an alternative model of the K-Medoids method, facilitated by applying quantum computing methodology to Manhattan distance calculations. Simulation results show that the K-Medoids method achieves best accuracy rate of 91% when using quantum computing. The test results confirm this accuracy, consistently showing a 91% accuracy rate. This alternative model can detect diabetes with an accuracy of 91%. From both calculations, the researchers also found that 238 data points entered cluster 1, which is non-diabetes, and 10 data points entered cluster 2, which is diabetes. For medical needs, the results obtained in this study can be used to detect diabetes with several medical attributes such as hypertension, heart disease, BMI, HbA1c level, and blood glucose level.

Although the results obtained by the Quantum K-Medoids algorithm are almost the same as those of the Classical K-Medoids in terms of accuracy, under certain conditions, the Quantum K-Medoids algorithm shows a greater reduction in cost value compared to the Classical K-Medoids. This study reveals that the Quantum K-Medoids algorithm can reduce the cost value more significantly than the Classical K-Medoids, highlighting its potential for more efficient clustering.

4. Conclusions

From calculations using K-Medoid and Quantum K-Medoid, researchers found that 238 data were classified into the non-diabetes cluster and 10 data were classified into the diabetes cluster. The experimental results demonstrate that the Quantum K-Medoids method with Manhattan distance calculations achieves the same level of accuracy as the K-Medoids clustering method without quantum computing, effectively grouping data in a dataset. This study produces a final value of both classical k-medoid and quantum k-medoid calculations of 91%, although the results obtained are sufficient to perform clustering, but the results of classical k-medoid and quantum k-medoid have different costs, where the cost value of the quantum algorithm is higher than the cost value of the classical algorithm, but for the average addition of cost value, the quantum k-medoid algorithm has a fairly stable value in performing classification. The significant decrease in cost value at the 8th epoch for the Quantum K-Medoids algorithm indicates that it can perform clustering with a lower cost value when allowed to run for more iterations. This suggests that, with higher iterations, the Quantum K-Medoids algorithm has the potential to achieve more efficient clustering compared to the Classical K-Medoids. This research marks a significant contribution to the field, introducing a new alternative model that utilizes quantum computing to enhance the K-Medoids data mining method with Manhattan distance calculations. For further research, researchers suggest performing calculations with an application or quantum device to find a more optimal value, or using more complex and larger data to cluster diabetes data to find a more optimal value or find parameters such as medoids that are optimal for both methods.

References

[1] F. Valenzuela, A. García, E. Ruiz, M. Vázquez, J. Cortez, and A. Espinoza, “An IoT-based glucose monitoring algorithm to prevent diabetes complications,” *Applied Sciences (Switzerland)*, vol. 10, no. 3, pp. 1–12, 2020, doi: 10.3390/app10030921.

[2] A. R. Nasser *et al.*, “Iot and cloud computing in health-care: A new wearable device and cloud-based deep learning algorithm for monitoring of diabetes,” *Electronics (Switzerland)*, vol. 10, no. 21, 2021, doi: 10.3390/electronics10212719.

[3] M. O. Edeh *et al.*, “A Classification Algorithm-Based Hybrid Diabetes Prediction Model,” *Frontiers in Public Health*, vol. 10, no. March, pp. 1–7, 2022, doi: 10.3389/fpubh.2022.829519.

[4] A. Paleczek, D. Grochala, and A. Rydosz, “Artificial breath classification using xgboost algorithm for diabetes detection,” *Sensors*, vol. 21, no. 12, 2021, doi: 10.3390/s21124187.

[5] A. Maulana *et al.*, “Machine Learning Approach for Diabetes Detection Using Fine-Tuned XGBoost Algorithm,” *Infolitika Journal of Data Science*, vol. 1, no. 1, pp. 1–7, 2023, doi: 10.60084/ijds.v1i1.72.

[6] Y. Istianto and S. ’Uyun, “Klasifikasi Kebutuhan Jumlah Produk Makanan Customer Menggunakan K-Means Clustering dengan Optimasi Pusat Awal Cluster Algoritma Genetika,” *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 8, no. 5, p. 861, 2021, doi: 10.25126/jtiik.2021842990.

[7] I. Setiaji and A. Z. Fanani, “Optimasi K-means Clustering Dengan Menggunakan Particle Swarm Optimization Untuk Menentukan Jumlah Cluster Pada Kanker Serviks,” *Jurnal Media Informatika Budidarma*, vol. 7, no. 3, pp. 1463–1473, 2023, doi: 10.30865/mib.v7i3.6292.

[8] M. I. Zarkasyi, H. Mawengkang, and O. S. Sitompul, “Optimasi Cluster Pada K-Means Clustering Dengan Teknik Reduksi Dimensi Dataset Menggunakan Gini Index,” *Building of Informatics, Technology and Science (BITS)*, vol. 4, no. 3, 2022, doi: 10.47065/bits.v4i3.2458.

[9] T. A. Mulyadi and D. Purnomo, “Optimasi Pelayanan Kapal Penumpang melalui Clustering Penumpang dengan Metode Silhouette Coefficient,” *Edumatic: Jurnal Pendidikan Informatika*, vol. 7, no. 2, pp. 217–226, 2023, doi: 10.29408/edumatic.v7i2.21067.

[10] M. D. Doi, A. Rusgiyono, and T. Wuryandari, “Analisis K-Medoids Dengan Validasi Indeks Pada Ipm Daerah 3t Di Indonesia,” *Jurnal Gaussian*, vol. 12, no. 2, pp. 178–188, 2023, doi: 10.14710/j.gauss.12.2.178-188.

[11] E. Syafaqoh, “Pengelompokan Provinsi Di Indonesia Berdasarkan Luas Panen, Produksi, Dan Produktivitas Padi Menggunakan Algoritma K-Medoid,” *Fakultas Sains dan Teknologi-Universitas PGRI Kanjuruhan Malang*, vol. 5, no. 3, p. 2023, 2023.

[12] D. A. I. C. Dewi and D. A. K. Pramita, “Analisis Perbandingan Metode Elbow dan Silhouette pada Algoritma Clustering K-Medoids dalam Pengelompokan Produksi Kerajinan Bali,” *Matrix : Jurnal Manajemen Teknologi dan Informatika*, vol. 9, no. 3, pp. 102–109, 2019, doi: 10.31940/matrix.v9i3.1662.

[13] H. Ramdan, A. Gunawan, and G. Gunawan, “Analisis Pengaruh Kardiovaskular Dalam Kasus Covid-19 Terhadap Obesitas Menggunakan Metode K-Medoid,” *Indonesian Journal Computer Science*, vol. 3, no. 1, pp. 16–24, Apr. 2024, doi: 10.31294/ijcs.v3i1.2558.

[14] W. Wahyu Pribadi, A. Yunus, and A. S. Wiguna, “Perbandingan Metode K-Means Euclidean Distance Dan Manhattan Distance Pada Penentuan Zonasi Covid-19 Di Kabupaten Malang,” *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 6, no. 2, pp. 493–500, 2022, doi: 10.36040/jati.v6i2.4808.

[15] T. M. Ghazal *et al.*, “Performances of k-means clustering algorithm with different distance metrics,” *Intelligent Automation and Soft Computing*, vol. 30, no. 2, pp. 735–742, 2021, doi: 10.32604/iasc.2021.019067.

[16] R. F. N. Alifah and A. C. Fauzan, “Implementasi Algoritma K-Means Clustering Berbasis Jarak Manhattan untuk Klasterisasi Konsentrasi Bidang Mahasiswa,” *ILKOMNIKA: Journal of Computer Science and Applied Informatics*, vol. 5, no. 1. Universitas Nahdlatul Ulama Blitar, pp. 31–41, 2023. doi: 10.28926/ilkomnika.v5i1.542.

[17] R. Suwanda, Z. Syahputra, and E. M. Zamzami, “Analysis of Euclidean Distance and Manhattan Distance in the K-Means Algorithm for Variations Number of Centroid K,” *Journal of Physics: Conference Series*, vol. 1566, no. 1, 2020, doi: 10.1088/1742-6596/1566/1/012058.

[18] S. Deshmukh, B. K. Behera, and P. Mulay, “Patient Data Analysis with the Quantum Clustering Method,” *Quantum Reports*, vol. 5, no. 1, pp. 138–155, 2023, doi: 10.3390/quantum5010010.

[19] C. Bauckhage, R. Sifa, and S. Wrobel, “Adiabatic Quantum Computing for Max-Sum Diversification,” in *Proceedings of the 2020 SIAM International Conference on Data Mining*, Philadelphia, PA: Society for Industrial and Applied Mathematics, 2020, pp. 343–351. doi: 10.1137/1.9781611976236.39.

[20] V. Hassija *et al.*, “Present landscape of quantum computing,” *IET Quantum Communication*, vol. 1, no. 2, pp. 42–48, 2020, doi: 10.1049/iet-qtc.2020.0027.

[21] F. Cleri, “Quantum computers, quantum computing, and quantum thermodynamics,” *Frontiers in Quantum Science and Technology*, vol. 3, no. August, pp. 1–29, 2024, doi: 10.3389/frqst.2024.1422257.

[22] A. Naik, E. Yeniaras, G. Hellstern, G. Prasad, and S. K. L. P. Vishwakarma, “From Portfolio Optimization to Quantum Blockchain and Security: A Systematic Review of Quantum Computing in Finance,” *arXiv*, pp. 1–64, Jun. 2023, doi: 10.48550/arXiv.2307.01155.

[23] U. Awan, L. Hannola, A. Tandon, R. K. Goyal, and A. Dhir, “Quantum computing challenges in the software industry. A fuzzy AHP-based approach,” *Information and Software Technology*, vol. 147, no. February, p. 106896, Jul. 2022, doi: 10.1016/j.infsof.2022.106896.

[24] T. Baidawi and Solikhun, “A Comparison of Madaline and Perceptron Algorithms on Classification with Quantum Computing Approach,” *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 8, no. 2, pp. 280–287, Apr. 2024, doi: 10.29207/resti.v8i2.5502.

[25] F. F. C. Silva, P. M. S. Carvalho, and L. A. F. M. Ferreira, “A quantum computing approach for minimum loss problems in electrical distribution networks,” *Scientific Reports*, vol. 13, no. 1, p. 10777, Jul. 2023, doi: 10.1038/s41598-023-37293-9.

[26] N. Febrian and N. Noviandi, “Perbandingan Manhattan dan Euclidean Distance Untuk Pengelompokan Penyakit Jantung Menggunakan Algoritma K-Means,” *ICIT Journal*, vol. 10, no. 1, pp. 61–70, 2024, doi: 10.33050/icit.v10i1.2860.

[27] R. H. Siahaan, S. Purba, J. Siregar, M. F. S. Hutabarat, and R. Sitohang, “Quantum-Enhanced K-Medoids Clustering : Comparative Analysis of,” *RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 5, no. 2, pp. 7–10, 2024, doi: 10.29207/resti.v8i5.5814.

[28] K. Saravananathan and T. Velmurugan, “Quality Based Analysis of Clustering Algorithms using Diabetes Data for the Prediction of Disease,” *International Journal of Innovative Technology and Exploring Engineering*, vol. 8, no. 11S2, pp. 448–452, Oct. 2019, doi: 10.35940/ijitee.K1072.09811S219.

[29] P. Lasek and Z. Mei, “Clustering and visualization of a high-dimensional diabetes dataset,” *Procedia Computer Science*, vol. 159, pp. 2179–2188, 2019, doi: 10.1016/j.procs.2019.09.392.

[30] L. Nigro and P. Fränti, “Two Medoid-Based Algorithms for Clustering Sets,” *Algorithms*, vol. 16, no. 7, pp. 1–17, 2023, doi: 10.3390/a16070349.

[31] Mifthahul Rahma and Muhammad Muhajir, “Komparasi Hierarchical Clustering Dan K-Medoid Clustering Terhadap Pengelompokan Satuan Kerja Perangkat Daerah (SKPD) Di Kabupaten Sleman Tahun 2021,” *Emerging Statistics and Data Science Journal*, vol. 2, no. 1, pp. 41–52, 2024, doi: 10.20885/esds.vol2.iss.1.art5.

[32] U. R. Gurning and M. Mustakim, “Penerapan Algoritma K-Means dan K-Medoid untuk Pengelompokkan Data Pasien Covid-19,” *Building of Informatics, Technology and Science (BITS)*, vol. 3, no. 1, pp. 48–55, Jun. 2021, doi: 10.47065/bits.v3i1.1003.

[33] A. C. Method, “Pengelompokkan Data Rekam Medis pada Penyakit Diabetes menggunakan Metode Divisive Analysis Clustering Clustering Medical Record Data on Diabetes Disease using Divisive,” vol. 13, pp. 1718–1731, 2024.

[34] S. A. P. Raj and Vidyaathulasiraman, “Determining Optimal Number of K for e-Learning Groups Clustered using K-Medoid,” *International Journal of Advanced Computer*

Science and Applications, vol. 12, no. 6, pp. 400–407, 2021, doi: 10.14569/IJACSA.2021.0120644.

[35] W. Aprilyani, K. Kaslani, E. Wahyudin, R. Hamonangan, and R. Herdiana, “Klasterisasi Data Penjualan Alat Transportasi Dengan Rapidminer Menggunakan Metode K-Medoid,” *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 8, no. 2, pp. 1348–1353, 2024, doi: 10.36040/jati.v8i2.8968.

[36] F. Fathurrahman, S. Harini, and R. Kusumawati, “EVALUASI CLUSTERING K-MEANS DAN K-MEDOID PADA PERSEBARAN COVID-19 DI INDONESIA DENGAN METODE DAVIES-BOULDIN INDEX (DBI),” *Jurnal Mnemonic*, vol. 6, no. 2, pp. 117–128, Oct. 2023, doi: 10.36040/mnemonic.v6i2.6642.

[37] Martanto, S. Anwar, C. L. Rohmat, F. M. Basysyar, and Y. A. Wijaya, “Clustering of internet network usage using the K-Medoid method,” *IOP Conference Series: Materials Science and Engineering*, vol. 1088, no. 1, p. 012036, 2021, doi: 10.1088/1757-899x/1088/1/012036.

[38] E. Wahyudi, D. Meidelfi, Nofrizal, Z. Saam, and Juandi, “The Implementation of the K-Medoid Clustering for Grouping Hearing Loss Function on Excessive Smartphone Use,” *International Journal on Informatics Visualization*, vol. 7, no. 4, pp. 2523–2531, 2023, doi: 10.30630/joiv.7.4.1873.

[39] G. Corrente, C. Vincenzo Stanzione, and V. Stanzione, “Comparison among Classical, Probabilistic and Quantum Algorithms for Hamiltonian Cycle Problem,” *Journal of Quantum Computing*, vol. 5, no. 0, pp. 55–70, 2023, doi: 10.32604/jqc.2023.044786.