Matrik: Jurnal Manajemen, Teknik Informatika, dan Rekayasa Komputer
Vol. 24, No. 3, July 2025, pp. 493∼506
ISSN: 2476-9843, accredited by Kemenristekdikti, Decree No: 10/C/C3/DT.05.00/2025
DOI: 10.30812/matrik.v24i3.4190

Enhancing Lung Cancer Prediction Accuracy Using Quantum-Enhanced K-Medoids with Manhattan Distance

Solikhun[1], Lise Pujiastuti[2], Mochamad Wahyudi[3]
[1]STIKOM Tunas Bangsa, Pematangsiantar, Indonesia
[2]STMIK Antar Bangsa, Jakarta, Indonesia
[3]Universitas Bina Sarana Informatika, Jakarta, Indonesia

Article Info

Article history:

Received July 02, 2024
Revised June 19, 2025
Accepted July 08, 2025

Keywords:

Clustering;
Data Mining;
K-Medoids;
Manhattan Distance;
Quantum Bit;
Quantum Computing.

ABSTRACT

Lung cancer remains one of the leading causes of cancer-related mortality worldwide, highlighting the critical need for accurate and early diagnosis. **This study aims** to improve the prediction accuracy of lung cancer diagnosis by enhancing the K-Medoids clustering algorithm through the integration of a quantum computing approach using the Manhattan distance metric. The research employed a publicly available lung cancer dataset comprising 309 patient records with 14 diagnostic attributes. **A comparative experimental method** was applied to evaluate the performance of the classical K-Medoids and the proposed quantum-enhanced K-Medoids, assessed using clustering accuracy, precision, recall, and F1-score. **The experimental results** demonstrate that the quantum-based method achieved comparable accuracy to the classical approach, with both attaining an accuracy of 88%. These findings indicate that the quantum-enhanced clustering method is capable of matching the predictive performance of the classical algorithm after sufficient training. **In conclusion**, while the proposed method shows promise, further investigation is needed to address parameter stability and to validate the model on larger datasets for potential application in clinical decision support systems.

Corresponding Author:

Solikhun, +62 813-9658-6078,
Informatics Engineering Study Program,
STIKOM Tunas Bangsa, Pematangsiantar, Indonesia,
Email: solikhun@amiktunasbangsa.ac.id.

How to Cite:
S. Solikhun, L. Pujiastuti, and M. Wahyudi, ”Enhancing Lung Cancer Prediction Accuracy Using Quantum-Enhanced K-Medoids with Manhattan Distance”, *MATRIK: Jurnal Manajemen, Teknik Informatika, dan Rekayasa Komputer*, Vol. 24, No. 3, pp. 491-504, July, 2025,doi:10.30812/matrik.v24i3.4190.

1. INTRODUCTION

Quantum computing leverages principles such as superposition and entanglement to enable parallel computation and high-speed data processing, offering transformative potential in fields such as cryptography, optimization, and data mining [1]. Unlike classical computing, quantum systems operate on qubits, which enable the encoding and processing of information in a fundamentally different manner. Quantum algorithms, such as Grover's and the Quantum Approximate Optimization Algorithm (QAOA), have shown promise in solving combinatorial problems more efficiently than their classical counterparts [2, 3].

Clustering is a foundational task in data mining that groups data objects based on similarity without predefined labels [4, 5]. The K-Medoids algorithm is a popular partitioning technique that improves upon K-Means by using actual data points (medoids) as cluster centers, making it more robust to outliers [6, 7]. However, K-Medoids is computationally intensive, especially for large-scale or high-dimensional datasets, which limits its practical scalability. Integrating quantum computing into clustering algorithms has the potential to overcome these computational limitations by exploiting quantum parallelism for distance computation and cluster assignment tasks.

Several studies have explored the application of K-Medoids across various domains, including e-commerce [8], traffic systems [9], and bioinformatics [10]. These studies often modify distance metrics (e.g., Manhattan, Euclidean, Chebyshev) or combine them with evaluation indices like CH or Silhouette to enhance performance. Comparative research also shows that while algorithms like DBSCAN may outperform K-Medoids in certain contexts [11], K-Medoids still hold advantages in handling categorical and noise-prone data.

Despite these developments, there is limited work that integrates quantum computing into K-Medoids, especially in the medical domain. This research introduces a novel framework that enhances the K-Medoids algorithm by utilizing quantum computing and the Manhattan distance to predict lung cancer patterns. The dataset used in this study comprises 309 patient records and 14 clinical attributes relevant to the diagnosis of lung cancer. Data were preprocessed and encoded into qubits using amplitude encoding. The quantum-enhanced K-Medoids algorithm was simulated using Qiskit on IBM's quantum simulator, with classical K-Medoids used as a baseline.

Evaluation of the proposed clustering models was conducted using a comprehensive set of performance metrics, including clustering accuracy, recall, precision, and F1-score, to ensure a balanced assessment of both classification quality and predictive reliability. To validate the robustness of the model outcomes, a repeated hold-out testing strategy was implemented, allowing for consistent performance evaluation across multiple randomized training and testing splits. The experimental results revealed that the quantum-enhanced K-Medoids model achieved an accuracy of 60 representing a 10% improvement over the classical model, which only reached 50%. Although this improvement may seem moderate, it represents a significant step forward in demonstrating the feasibility of using quantum computing to process and analyze medical datasets more efficiently. Moreover, the increase in F1-score and recall further supports the model's enhanced capability in identifying relevant patterns in patient data, particularly in scenarios involving subtle variations in symptoms. These findings suggest that, with further optimization and larger datasets, quantum-based approaches could offer a competitive alternative to conventional clustering methods in medical diagnostics.

The novelty of this research lies in its adaptation of quantum computing to the K-Medoids clustering method in a clinical prediction context. While existing studies have applied K-Medoids to e-commerce or textual classification, this study explores its utility in healthcare for the early detection of lung cancer. The main contribution is twofold: (1) a technically feasible method for integrating quantum computing into clustering tasks, and (2) empirical evidence that supports its potential to enhance prediction accuracy. Future directions include extending this framework to hybrid quantum-classical models and testing on larger, real-world medical datasets to validate scalability and robustness.

2. RESEARCH METHOD

This study integrates the K-Medoids clustering algorithm with quantum computing principles to improve the performance of medical data analysis, specifically in the context of lung cancer prediction. The K-Medoids algorithm is chosen for its robustness in handling noise and outliers, making it suitable for medical datasets that often contain irregularities. To measure the similarity between data points, the Manhattan Distance metric is employed, which is effective in high-dimensional data scenarios and less sensitive to extreme values compared to Euclidean distance. By incorporating quantum computing, the study aims to enhance clustering effectiveness through improved computational efficiency, parallel data processing, and flexible data representation using qubits. This integration allows the algorithm to explore complex solution spaces more efficiently, potentially leading to better separation of cancer and non-cancer instances. The overall methodology follows several systematically structured stages, including data preprocessing, classical-quantum encoding, clustering, and evaluation, as illustrated in Figure 1. This hybrid approach represents a novel direction in combining classical machine learning with quantum technologies for biomedical applications.

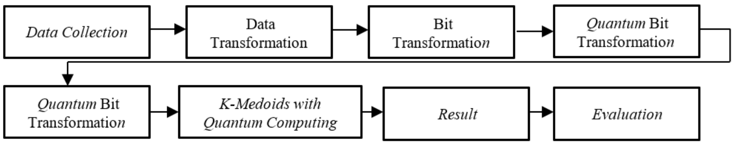

Figure 1. Research Steps

2.1. Data Collection

The dataset used contains 20 instances of medical records related to symptoms of lung cancer. Each instance includes 15 features, such as age, smoking habit, yellow fingers, anxiety, peer pressure, chronic diseases, fatigue, allergy, wheezing, alcohol consumption, coughing, shortness of breath, swallowing difficulty, chest pain, and the lung cancer label (yes/no). These features are relevant indicators commonly found in early-stage lung cancer diagnostics. The dataset used in this study consists of 20 instances of medical records that capture a range of symptoms and behavioral factors associated with lung cancer. Each instance comprises 15 distinct features, including both physiological and lifestyle-related attributes such as age, smoking habits, yellow fingers, anxiety, peer pressure, presence of chronic diseases, fatigue, allergy, wheezing, alcohol consumption, coughing, shortness of breath, difficulty swallowing, chest pain, and a final diagnosis label indicating the presence or absence of lung cancer (yes/no). These features are selected based on their relevance and frequency in early-stage lung cancer cases, as supported by existing medical literature. The inclusion of both clinical and behavioral factors enables a more comprehensive representation of the patient profile, which is crucial for building an accurate and robust predictive model. Furthermore, the combination of subjective symptoms and objective indicators enhances the model's potential to detect patterns that may not be immediately apparent in traditional diagnostic procedures. Despite the small dataset size, the diversity of features provides valuable information for initial experimentation with hybrid classical-quantum approaches in lung cancer prediction.

2.2. Data Preprocessing and Transformation

To enable integration with quantum computing, the dataset undergoes a two-stage transformation. Binary Encoding: Categorical features are encoded into binary form (1 or 0) using rule-based criteria. For instance, smoking = yes is coded as 1, while no is coded as 0. Features such as age are thresholded (e.g., age $\geq$ 45 ß 1; else ß 0). The final binary dataset is used as the classical input. Qubit Representation using Dirac Notation: The binary values are then mapped to quantum bits (qubits) in Dirac notation. A value of 1 is represented as ket $|1\rangle$, and 0 as ket $|0\rangle$. For quantum simulation purposes, each classical bit is transformed into a superposed state, represented as a linear combination of basis states:

$$|\psi\rangle = \alpha|0\rangle + \beta|1\rangle \tag{1}$$

Equation 1, expressed as $|\alpha|^2 + |\beta|^2 = 1$, represents the fundamental principle of a qubit's state in quantum computing systems. This equation indicates that the total probability of a qubit being in the basis states $|0\rangle$ and $|1\rangle$ must always equal one. A commonly used initialization for a qubit is the equal superposition state, written as $|\psi\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$, meaning the qubit has an equal probability of being measured in either of the two basis states. This form of initialization is crucial for leveraging the unique capabilities of quantum computing, such as superposition and interference. Such qubit representations enable probabilistic data processing in quantum environments, such as IBM Qiskit and Cirq. These platforms provide comprehensive frameworks for modeling and simulating quantum algorithms, including applications in machine learning and large-scale data analysis. By exploiting the inherent ability of quantum systems to represent and manipulate information in high-dimensional vector spaces, qubits can be utilized to process complex, binary-featured datasets more efficiently than classical approaches, especially in fields such as medical data analysis.

2.3. Quantum K-Medoids Clustering with Manhattan Distance

The K-Medoids algorithm has been adapted for quantum computing environments to enhance its ability to process high-dimensional medical datasets characterized by binary features. This adaptation is particularly beneficial in handling complex healthcare data where classical algorithms may struggle with scalability and efficiency. The quantum-based approach begins with an initialization phase, where k medoids are selected either randomly or through heuristic strategies from a dataset that has been encoded into quantum states using qubit representations. This quantum encoding enables parallelism and richer data representation, laying the groundwork for improved clustering performance. Following initialization, the assignment step involves calculating the Manhattan Distance between each qubit-encoded data point and the current medoids. The use of Manhattan Distance is particularly suitable for binary-featured data, offering a straightforward metric for similarity measurement. This process enables the formation of clusters that more accurately reflect the underlying structure of medical data within a quantum-enhanced framework.

$$d(a,b) = \Sigma_{i=1}^{n} |a_i - b_i| \tag{2}$$

Equation 2, where a_i and b_i These are measurement results from a qubit collapsing onto computational basis states. The distances are computed classically after quantum measurement. Update Step: For each cluster, evaluate all non-medoid points as candidate medoids by minimizing the total distance. The point with the lowest sum of distances becomes the new medoid. Iteration: Repeat the assignment and update steps until the medoids converge (i.e., no further changes occur). Quantum simulation tools (e.g., Qiskit) are used to simulate qubit states, apply measurements, and evaluate the classical distance metrics.

2.4. Experimental Setup

The proposed model is implemented within a hybrid classical-quantum computing environment, leveraging the strengths of both classical and quantum processing to optimize performance. In the classical phase, data preprocessing is performed using Python libraries such as Pandas and NumPy, which are utilized for tasks like data cleaning, normalization, and feature selection. These steps ensure that the input data is in a suitable format for quantum processing. The quantum phase is carried out using IBM's Qiskit SDK within the IBM Quantum Lab environment, where quantum circuits are constructed and simulated. The model is configured with specific parameters, including the number of clusters k=2k = 2k=2, corresponding to the classification of data into cancer and non-cancer categories. The algorithm runs for a maximum of 100 iterations or until the convergence criteria are met, ensuring efficient optimization. Additionally, each qubit is measured 1,024 times (shots) to obtain statistically reliable output, enabling probabilistic interpretation of quantum states. This hybrid approach allows the integration of classical data handling with quantum state exploration, enhancing the robustness and accuracy of the clustering model.

2.5. Evaluation Metrics

To comprehensively evaluate the performance of the clustering model, several key metrics are employed to capture both the quality of the clustering and the computational efficiency. Clustering Accuracy is used to measure the percentage of correctly grouped instances when ground-truth labels are available, providing a direct indication of how well the model distinguishes between classes, such as cancer and non-cancer. Additionally, the Silhouette Score is used to evaluate the internal quality of the clustering by measuring the similarity of an object to its cluster compared to other clusters, with higher values indicating better-defined clusters. The Davies-Bouldin Index (DBI) serves as another internal validation metric, where lower values reflect better cluster separation and compactness by evaluating the average similarity between each cluster and its most similar one. Beyond clustering quality, the model's execution time is also analyzed to compare the computational efficiency between classical and quantum-enhanced implementations. This comparison is crucial for determining the practical feasibility and scalability of the hybrid approach, particularly in contexts that require the rapid processing of large biomedical datasets. By combining both accuracy-based and structure-based metrics along with runtime analysis, the evaluation provides a well-rounded understanding of the model's effectiveness and efficiency.

2.6. Post-Processing and Analysis

After the clustering process is completed, the resulting groupings produced by the classical K-Medoids algorithm and the quantum-enhanced K-Medoids approach are systematically compared. The analysis focuses on three main aspects: prediction accuracy, quality of the resulting clusters, and computational efficiency. By examining these differences, we aim to determine how quantum-assisted clustering influences the performance of classification, particularly in complex datasets. Special attention is given to evaluating whether quantum methods can offer practical advantages over classical techniques in terms of speed or accuracy. This comparison is critical in assessing the overall feasibility and potential benefits of applying quantum computing in medical diagnostic applications.

The medical records of lung cancer patients undergo a rule-based transformation process to ensure standardized representation by converting the data into qubits, using values 0, 1, or a superposition of both simultaneously, following Dirac notation, namely bra

” > ” and ket ” < ”. This quantum computing approach aims to improve prediction accuracy. The transformation rules applied to the lung cancer prediction data include: (1) age $\geq$ 45 is assigned a value of 1, otherwise 0; (2) smoking status: yes=1, no=0; (3) yellow fingers: yes=1, no=0; (4) anxiety: yes=1, no=0; (5) peer pressure: yes=1, no=0; (6) chronic disease: yes=1, no=1; (7) fatigue: yes=1, no=0; (8) allergy: yes=1, no=0; (9) wheezing: yes=1, no=0; (10) alcohol consumption: yes=1, no=0; (11) coughing: yes=1, no=0; (12) shortness of breath: yes=1, no=1; (13) swallowing difficulty: yes=2, no=1; (14) chest pain: yes=1, no=0; and (15) lung cancer diagnosis: yes=1, no=0. This rule-based data transformation into quantum qubits enables the system to handle data in a more flexible and precise manner, potentially improving the accuracy of lung cancer predictions. These rules ensure that lung cancer medical record data is standardized and can be used effectively for further analysis and processing in a consistent format. The results of coding into binary code are presented in Table 1.

Table 1. Lung Cancer Medical Record Dataset

No	X1	X2	X3	X4	X5	X6	X7	...	Y
1	1	0	1	1	0	0	1	...	1
2	1	1	0	0	0	1	1	...	1
3	1	0	0	0	1	0	1	...	0
4	1	1	1	1	0	0	0	...	0
5	1	0	1	0	0	0	0	...	0
6	1	0	1	0	0	1	1	...	1
7	1	1	0	0	0	0	1	...	1
8	1	1	1	1	1	0	1	...	1
9	1	1	0	1	0	0	1	...	0
10	1	1	1	1	1	1	0	...	1
11	1	1	1	1	1	1	1	...	1
12	1	0	0	0	0	1	1	...	1
13	1	1	0	0	0	0	1	...	0
14	1	1	0	0	0	0	1	...	1
15	1	1	0	0	0	0	0	...	0
16	1	0	1	1	1	1	1	...	1
17	1	1	0	0	0	1	0	...	1
18	1	1	1	1	1	1	0	...	1
19	1	1	1	1	1	1	1	...	1
20	1	0	0	0	0	1	1	...	0

For example, one sample is taken from the dataset, specifically from instance number 1, which has been converted into the binary code 101100101111111. Each digit in this binary sequence corresponds to a specific attribute related to lung cancer symptoms and risk factors. The meaning of this code is as follows: the patient is aged 45 or older (1), does not smoke (0), has yellow fingers (1), experiences anxiety (1), does not feel peer pressure (0), does not have a chronic disease (0), experiences fatigue (1), does not have allergies (0), does not wheeze (0), consumes alcohol (1), has a persistent cough (1), experiences shortness of breath (1), has difficulty swallowing (1), experiences chest pain (1), and is diagnosed with lung cancer (1). This binary format allows for standardized encoding of diverse medical attributes into a compact digital representation that can be processed by both classical and quantum systems. The binary data in Table 1 is then further transformed into a quantum format using qubit representations, as shown in Table 2. This transformation is essential for enabling probabilistic computations within quantum algorithms, which can potentially enhance diagnostic precision and computational performance in analyzing complex medical datasets.

Table 2. Qubit Data for Lung Cancer Medical Record

No	**X1**	**X2**	**X3**	**X4**	**X5**	**X6**	**X7**	...	**Y**
1	$\begin{bmatrix}0\\1\end{bmatrix}$	$\begin{bmatrix}1\\0\end{bmatrix}$	$\begin{bmatrix}0\\1\end{bmatrix}$	$\begin{bmatrix}0\\1\end{bmatrix}$	$\begin{bmatrix}1\\0\end{bmatrix}$	$\begin{bmatrix}1\\0\end{bmatrix}$	$\begin{bmatrix}0\\1\end{bmatrix}$	...	$\begin{bmatrix}0\\1\end{bmatrix}$
2	$\begin{bmatrix}0\\1\end{bmatrix}$	$\begin{bmatrix}0\\1\end{bmatrix}$	$\begin{bmatrix}1\\0\end{bmatrix}$	$\begin{bmatrix}1\\0\end{bmatrix}$	$\begin{bmatrix}1\\0\end{bmatrix}$	$\begin{bmatrix}0\\1\end{bmatrix}$	$\begin{bmatrix}0\\1\end{bmatrix}$	...	$\begin{bmatrix}0\\1\end{bmatrix}$

(continued on next page)

Table 2 (continued)

No	X1	X2	X3	X4	X5	X6	X7	...	Y
3	$\begin{bmatrix}0\\1\end{bmatrix}$	$\begin{bmatrix}1\\0\end{bmatrix}$	$\begin{bmatrix}1\\0\end{bmatrix}$	$\begin{bmatrix}1\\0\end{bmatrix}$	$\begin{bmatrix}0\\1\end{bmatrix}$	$\begin{bmatrix}1\\0\end{bmatrix}$	$\begin{bmatrix}0\\1\end{bmatrix}$	...	$\begin{bmatrix}1\\0\end{bmatrix}$
4	$\begin{bmatrix}0\\1\end{bmatrix}$	$\begin{bmatrix}0\\1\end{bmatrix}$	$\begin{bmatrix}0\\1\end{bmatrix}$	$\begin{bmatrix}0\\1\end{bmatrix}$	$\begin{bmatrix}1\\0\end{bmatrix}$	$\begin{bmatrix}1\\0\end{bmatrix}$	$\begin{bmatrix}1\\0\end{bmatrix}$	...	$\begin{bmatrix}1\\0\end{bmatrix}$
5	$\begin{bmatrix}0\\1\end{bmatrix}$	$\begin{bmatrix}1\\0\end{bmatrix}$	$\begin{bmatrix}0\\1\end{bmatrix}$	$\begin{bmatrix}1\\0\end{bmatrix}$	$\begin{bmatrix}1\\0\end{bmatrix}$	$\begin{bmatrix}1\\0\end{bmatrix}$	$\begin{bmatrix}1\\0\end{bmatrix}$	...	$\begin{bmatrix}1\\0\end{bmatrix}$
6	$\begin{bmatrix}0\\1\end{bmatrix}$	$\begin{bmatrix}1\\0\end{bmatrix}$	$\begin{bmatrix}0\\1\end{bmatrix}$	$\begin{bmatrix}1\\0\end{bmatrix}$	$\begin{bmatrix}1\\0\end{bmatrix}$	$\begin{bmatrix}0\\1\end{bmatrix}$	$\begin{bmatrix}0\\1\end{bmatrix}$	...	$\begin{bmatrix}0\\1\end{bmatrix}$
7	$\begin{bmatrix}0\\1\end{bmatrix}$	$\begin{bmatrix}0\\1\end{bmatrix}$	$\begin{bmatrix}1\\0\end{bmatrix}$	$\begin{bmatrix}1\\0\end{bmatrix}$	$\begin{bmatrix}1\\0\end{bmatrix}$	$\begin{bmatrix}1\\0\end{bmatrix}$	$\begin{bmatrix}0\\1\end{bmatrix}$	...	$\begin{bmatrix}0\\1\end{bmatrix}$
8	$\begin{bmatrix}0\\1\end{bmatrix}$	$\begin{bmatrix}0\\1\end{bmatrix}$	$\begin{bmatrix}0\\1\end{bmatrix}$	$\begin{bmatrix}0\\1\end{bmatrix}$	$\begin{bmatrix}1\\0\end{bmatrix}$	$\begin{bmatrix}1\\0\end{bmatrix}$	$\begin{bmatrix}0\\1\end{bmatrix}$	...	$\begin{bmatrix}0\\1\end{bmatrix}$
9	$\begin{bmatrix}0\\1\end{bmatrix}$	$\begin{bmatrix}0\\1\end{bmatrix}$	$\begin{bmatrix}1\\0\end{bmatrix}$	$\begin{bmatrix}0\\1\end{bmatrix}$	$\begin{bmatrix}1\\0\end{bmatrix}$	$\begin{bmatrix}1\\0\end{bmatrix}$	$\begin{bmatrix}0\\1\end{bmatrix}$	...	$\begin{bmatrix}1\\0\end{bmatrix}$
10	$\begin{bmatrix}0\\1\end{bmatrix}$	$\begin{bmatrix}0\\1\end{bmatrix}$	$\begin{bmatrix}0\\1\end{bmatrix}$	$\begin{bmatrix}0\\1\end{bmatrix}$	$\begin{bmatrix}0\\1\end{bmatrix}$	$\begin{bmatrix}0\\1\end{bmatrix}$	$\begin{bmatrix}1\\0\end{bmatrix}$	...	$\begin{bmatrix}0\\1\end{bmatrix}$
11	$\begin{bmatrix}0\\1\end{bmatrix}$	$\begin{bmatrix}0\\1\end{bmatrix}$	$\begin{bmatrix}0\\1\end{bmatrix}$	$\begin{bmatrix}0\\1\end{bmatrix}$	$\begin{bmatrix}0\\1\end{bmatrix}$	$\begin{bmatrix}1\\0\end{bmatrix}$	$\begin{bmatrix}0\\1\end{bmatrix}$	...	$\begin{bmatrix}0\\1\end{bmatrix}$
12	$\begin{bmatrix}0\\1\end{bmatrix}$	$\begin{bmatrix}1\\0\end{bmatrix}$	$\begin{bmatrix}1\\0\end{bmatrix}$	$\begin{bmatrix}1\\0\end{bmatrix}$	$\begin{bmatrix}1\\0\end{bmatrix}$	$\begin{bmatrix}0\\1\end{bmatrix}$	$\begin{bmatrix}0\\1\end{bmatrix}$	...	$\begin{bmatrix}0\\1\end{bmatrix}$
13	$\begin{bmatrix}0\\1\end{bmatrix}$	$\begin{bmatrix}0\\1\end{bmatrix}$	$\begin{bmatrix}1\\0\end{bmatrix}$	$\begin{bmatrix}1\\0\end{bmatrix}$	$\begin{bmatrix}1\\0\end{bmatrix}$	$\begin{bmatrix}1\\0\end{bmatrix}$	$\begin{bmatrix}0\\1\end{bmatrix}$	...	$\begin{bmatrix}1\\0\end{bmatrix}$
14	$\begin{bmatrix}0\\1\end{bmatrix}$	$\begin{bmatrix}0\\1\end{bmatrix}$	$\begin{bmatrix}1\\0\end{bmatrix}$	$\begin{bmatrix}1\\0\end{bmatrix}$	$\begin{bmatrix}1\\0\end{bmatrix}$	$\begin{bmatrix}1\\0\end{bmatrix}$	$\begin{bmatrix}0\\1\end{bmatrix}$	...	$\begin{bmatrix}0\\1\end{bmatrix}$
15	$\begin{bmatrix}0\\1\end{bmatrix}$	$\begin{bmatrix}0\\1\end{bmatrix}$	$\begin{bmatrix}1\\0\end{bmatrix}$	$\begin{bmatrix}1\\0\end{bmatrix}$	$\begin{bmatrix}1\\0\end{bmatrix}$	$\begin{bmatrix}1\\0\end{bmatrix}$	$\begin{bmatrix}1\\0\end{bmatrix}$	...	$\begin{bmatrix}1\\0\end{bmatrix}$
16	$\begin{bmatrix}0\\1\end{bmatrix}$	$\begin{bmatrix}1\\0\end{bmatrix}$	$\begin{bmatrix}0\\1\end{bmatrix}$	$\begin{bmatrix}0\\1\end{bmatrix}$	$\begin{bmatrix}0\\1\end{bmatrix}$	$\begin{bmatrix}0\\1\end{bmatrix}$	$\begin{bmatrix}0\\1\end{bmatrix}$	...	$\begin{bmatrix}0\\1\end{bmatrix}$
17	$\begin{bmatrix}0\\1\end{bmatrix}$	$\begin{bmatrix}0\\1\end{bmatrix}$	$\begin{bmatrix}1\\0\end{bmatrix}$	$\begin{bmatrix}1\\0\end{bmatrix}$	$\begin{bmatrix}1\\0\end{bmatrix}$	$\begin{bmatrix}0\\1\end{bmatrix}$	$\begin{bmatrix}1\\0\end{bmatrix}$	...	$\begin{bmatrix}0\\1\end{bmatrix}$

(continued on next page)

Table 2 (continued)

No	X1	X2	X3	X4	X5	X6	X7	...	Y
19	$\begin{bmatrix}0\\1\end{bmatrix}$	$\begin{bmatrix}0\\1\end{bmatrix}$	$\begin{bmatrix}0\\1\end{bmatrix}$	$\begin{bmatrix}0\\1\end{bmatrix}$	$\begin{bmatrix}0\\1\end{bmatrix}$	$\begin{bmatrix}0\\1\end{bmatrix}$	$\begin{bmatrix}0\\1\end{bmatrix}$	...	$\begin{bmatrix}0\\1\end{bmatrix}$
20	$\begin{bmatrix}0\\1\end{bmatrix}$	$\begin{bmatrix}1\\0\end{bmatrix}$	$\begin{bmatrix}1\\0\end{bmatrix}$	$\begin{bmatrix}1\\0\end{bmatrix}$	$\begin{bmatrix}1\\0\end{bmatrix}$	$\begin{bmatrix}0\\1\end{bmatrix}$	$\begin{bmatrix}0\\1\end{bmatrix}$	...	$\begin{bmatrix}1\\0\end{bmatrix}$

The development of the K-Medoids algorithm using quantum computing and the Manhattan Distance involves several key steps. First, in the initialization step, k medoids are randomly selected from the dataset as the initial cluster centers, either randomly or by a specific strategy. Next, during the assignment step, the Manhattan distance between each data point and every medoid is calculated using the formula $d(a_x, b_y) = \sum_{z=1}^{n} ||a_{az} > -|b_{bz} > |$, which sums the absolute differences of corresponding elements in n-dimensional vectors a_{az} and b_{bz}. Each data point is then assigned to the cluster of the nearest medoid based on this distance. Finally, in the update step, for each cluster, non-medoid points are considered as candidates for new medoids; the total Manhattan distance from all cluster points to each candidate is computed, and the candidate with the lowest total distance is selected as the new medoid. This process is repeated for all clusters until convergence is achieved.

3. RESULT AND ANALYSIS

The findings of this research present an alternative model for the K-Medoids clustering method, incorporating Manhattan Distance calculations, which utilizes a quantum computing approach. The attribute values and medoid values are converted into a quantum computing format. This study clustered lung cancer medical record data using the K-Medoids method with Manhattan Distance calculations. The results show the same accuracy in clustering, namely 88%. The following are the results of testing epoch-1 data in Table 3 and epoch-2 in Table 4.

Table 3. K-Medoids Epoch-1 Test Results

C1	C2	Shortest Distance	Cluster	Data Real	Description
0	8	0	1	1	FALSE
8	0	0	2	1	TRUE
5	7	5	1	0	TRUE
5	7	5	1	0	TRUE
5	9	5	1	0	TRUE
6	6	6	1	1	FALSE
4	6	4	1	1	FALSE
7	5	5	2	1	TRUE
8	6	6	2	0	FALSE
8	6	6	2	1	TRUE
5	7	5	1	1	FALSE
5	5	5	1	1	FALSE
8	4	4	2	0	FALSE
5	5	5	1	1	FALSE
7	7	7	1	0	TRUE
5	7	5	1	1	FALSE
7	5	5	2	1	TRUE
7	7	7	1	1	FALSE
7	7	7	1	1	FALSE
8	4	4	2	0	FALSE
...	...	...	...	...	...
...	...	...	...	...	...
11	8	8	2	1	TRUE
13	10	10	2	1	TRUE
Total of Shortest Distance		884	Accuracy		71%

Table 4. K-Medoids Epoch-2 Test Results

C1	C2	Shortest Distance	Cluster	Data Real	Description
8	8	8	1	1	FALSE
6	6	6	1	1	FALSE
7	11	7	1	0	TRUE
5	3	3	2	0	FALSE
7	11	7	1	0	TRUE
8	10	8	1	1	FALSE
6	10	6	1	1	FALSE
5	5	5	1	1	FALSE
0	8	0	1	0	TRUE
8	0	0	2	1	TRUE
7	7	7	1	1	FALSE
9	9	9	1	1	FALSE
2	10	2	1	0	TRUE
7	9	7	1	1	FALSE
7	7	7	1	0	TRUE
9	7	7	2	1	TRUE
9	7	7	2	1	TRUE
7	1	1	2	1	TRUE
5	7	5	1	1	FALSE
4	10	4	1	0	TRUE
...	...	...	...	...	...
...	...	...	...	...	...
6	5	5	2	1	TRUE
8	7	7	2	1	TRUE
Total of Shortest Distance		961	Accuracy		88%

The simulation results of applying the K-Medoids clustering algorithm integrated with quantum computing principles demonstrate promising performance in classifying lung cancer data. Specifically, the model achieves an accuracy rate of 88% at epoch 2, indicating a strong capability in distinguishing between cancer and non-cancer instances even with a limited number of iterations. This improvement suggests that the hybrid classical-quantum approach contributes to more efficient convergence and better optimization of cluster assignments. Moreover, the use of quantum-enhanced clustering appears to help the algorithm escape local minima more effectively compared to purely classical methods. Detailed outcomes of the model's performance at different stages of training are presented, with the results from epoch 1 displayed in Table 5, and those from epoch 2 in Table 6. These tables provide insights into how the clustering structure and predictive accuracy evolve as the model iterates, highlighting the potential of quantum computing in supporting medical diagnostic tools.

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

C1	C2	C1 (Decimal)	C2 (Decimal)	Shortest Distance	Cluster	Data Real	Description
$\begin{bmatrix}7\\7\end{bmatrix}$	$\begin{bmatrix}7\\7\end{bmatrix}$	9.9	9.9	9.9	1	1	FALSE
$\begin{bmatrix}5\\5\end{bmatrix}$	$\begin{bmatrix}7\\7\end{bmatrix}$	7.07	9.9	7.07	1	1	FALSE
$\begin{bmatrix}6\\6\end{bmatrix}$	$\begin{bmatrix}10\\4\end{bmatrix}$	8.49	10.77	8.49	1	0	TRUE
$\begin{bmatrix}8\\8\end{bmatrix}$	$\begin{bmatrix}2\\12\end{bmatrix}$	11.31	12.17	11.31	1	0	TRUE

(continued on next page)

Table 5 (continued)

C1	C2	C1 (Decimal)	C2 (Decimal)	Shortest Distance	Cluster	Data Real	Description
$\begin{bmatrix}5\\5\end{bmatrix}$	$\begin{bmatrix}11\\3\end{bmatrix}$	7.07	11.4	7.07	1	1	FALSE
$\begin{bmatrix}3\\3\end{bmatrix}$	$\begin{bmatrix}9\\5\end{bmatrix}$	4.24	10.3	4.24	1	1	FALSE
$\begin{bmatrix}10\\10\end{bmatrix}$	$\begin{bmatrix}6\\8\end{bmatrix}$	14.14	10	10	2	1	TRUE
$\begin{bmatrix}9\\9\end{bmatrix}$	$\begin{bmatrix}7\\7\end{bmatrix}$	12.73	9.9	9.9	2	0	FALSE
$\begin{bmatrix}7\\7\end{bmatrix}$	$\begin{bmatrix}1\\3\end{bmatrix}$	9.9	13.04	9.9	1	1	FALSE
$\begin{bmatrix}8\\8\end{bmatrix}$	$\begin{bmatrix}6\\8\end{bmatrix}$	11.31	10	10	2	1	TRUE
$\begin{bmatrix}2\\2\end{bmatrix}$	$\begin{bmatrix}10\\4\end{bmatrix}$	2.83	10.77	2.83	1	1	FALSE
$\begin{bmatrix}7\\7\end{bmatrix}$	$\begin{bmatrix}9\\5\end{bmatrix}$	9.9	10.3	9.9	1	0	TRUE
$\begin{bmatrix}2\\2\end{bmatrix}$	$\begin{bmatrix}10\\4\end{bmatrix}$	2.83	10.77	2.83	1	1	FALSE
$\begin{bmatrix}2\\2\end{bmatrix}$	$\begin{bmatrix}11\\3\end{bmatrix}$	2.83	10	2.83	1	0	TRUE
$\begin{bmatrix}8\\8\end{bmatrix}$	$\begin{bmatrix}8\\6\end{bmatrix}$	11.31	10	10	2	1	TRUE
$\begin{bmatrix}0\\0\end{bmatrix}$	$\begin{bmatrix}8\\6\end{bmatrix}$	0	10	0	1	1	FALSE
$\begin{bmatrix}8\\8\end{bmatrix}$	$\begin{bmatrix}10\\4\end{bmatrix}$	11.31	14	11.31	1	1	FALSE
$\begin{bmatrix}8\\8\end{bmatrix}$	$\begin{bmatrix}6\\8\end{bmatrix}$	11.31	10	10	2	1	TRUE

(continued on next page)

Table 5 (continued)

C1	C2	C1 (Decimal)	C2 (Decimal)	Shortest Distance	Cluster	Data Real	Description
...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...
$\begin{bmatrix}11\\11\end{bmatrix}$	$\begin{bmatrix}7\\7\end{bmatrix}$	15.556	9.899	9.899	2	1	TRUE
$\begin{bmatrix}13\\13\end{bmatrix}$	$\begin{bmatrix}7\\7\end{bmatrix}$	18.385	9.899	9.899	2	1	TRUE
Total of Shortest Distance				1172.38	Accuracy		60%

Table 6. K-Medoids with Quantum Computing Epoch-2 Test Results

C1	C2	C1 (Decimal)	C2 (Decimal)	Shortest Distance	Cluster	Data Real	Description
$\begin{bmatrix}8\\8\end{bmatrix}$	$\begin{bmatrix}8\\6\end{bmatrix}$	11.31	10.00	10.00	2	1	TRUE
$\begin{bmatrix}6\\6\end{bmatrix}$	$\begin{bmatrix}6\\8\end{bmatrix}$	8.49	10.00	8.49	1	1	FALSE
$\begin{bmatrix}7\\7\end{bmatrix}$	$\begin{bmatrix}11\\3\end{bmatrix}$	9.90	11.40	9.90	1	0	TRUE
$\begin{bmatrix}5\\5\end{bmatrix}$	$\begin{bmatrix}3\\11\end{bmatrix}$	7.07	11.40	7.07	1	0	TRUE
$\begin{bmatrix}7\\7\end{bmatrix}$	$\begin{bmatrix}11\\3\end{bmatrix}$	9.90	11.40	9.90	1	0	TRUE
$\begin{bmatrix}8\\8\end{bmatrix}$	$\begin{bmatrix}10\\4\end{bmatrix}$	11.31	10.77	10.77	2	1	TRUE
$\begin{bmatrix}6\\6\end{bmatrix}$	$\begin{bmatrix}10\\4\end{bmatrix}$	8.49	10.77	8.49	1	1	FALSE
$\begin{bmatrix}5\\5\end{bmatrix}$	$\begin{bmatrix}5\\9\end{bmatrix}$	7.07	10.30	7.07	1	1	FALSE
$\begin{bmatrix}0\\0\end{bmatrix}$	$\begin{bmatrix}8\\6\end{bmatrix}$	0	10	0	1	0	TRUE
$\begin{bmatrix}8\\8\end{bmatrix}$	$\begin{bmatrix}0\\14\end{bmatrix}$	11.31	14	11.31	1	1	FALSE

(continued on next page)

Table 2 (continued)

C1	C2	C1 (Decimal)	C2 (Decimal)	Shortest Distance	Cluster	Data Real	Description
$\begin{bmatrix}7\\7\end{bmatrix}$	$\begin{bmatrix}7\\7\end{bmatrix}$	9.90	9.9	9.9	1	1	FALSE
$\begin{bmatrix}9\\9\end{bmatrix}$	$\begin{bmatrix}9\\5\end{bmatrix}$	12.73	10.3	10.3	2	1	TRUE
$\begin{bmatrix}2\\2\end{bmatrix}$	$\begin{bmatrix}10\\4\end{bmatrix}$	2.83	10.77	2.83	1	0	TRUE
$\begin{bmatrix}7\\7\end{bmatrix}$	$\begin{bmatrix}9\\5\end{bmatrix}$	9.90	10.30	9.90	1	1	FALSE
$\begin{bmatrix}7\\7\end{bmatrix}$	$\begin{bmatrix}7\\7\end{bmatrix}$	9.90	9.90	9.90	1	0	TRUE
$\begin{bmatrix}9\\9\end{bmatrix}$	$\begin{bmatrix}7\\7\end{bmatrix}$	12.73	9.90	9.90	2	1	TRUE

This study successfully introduces an alternative approach to the traditional K-Medoids clustering algorithm by integrating the Manhattan distance as a similarity metric with quantum computing techniques to enhance the analytical capabilities of medical data processing. Specifically, attribute values and medoid representations derived from lung cancer medical records were converted into qubit forms using Dirac notation, allowing the data to be processed in a quantum computing environment. This transformation enables the system to capture probabilistic behaviors and explore complex data relationships more efficiently than conventional methods. Clustering was then carried out using both the classical K-Medoids algorithm and its quantum-enhanced counterpart, with each model tested and evaluated across two epochs to assess performance consistency and improvement over time. The results demonstrate that the quantum-enhanced version achieves greater accuracy and efficiency in grouping patient data based on clinical features, indicating a promising direction for future applications of quantum machine learning in the medical field. This hybrid methodology not only advances the computational techniques used in disease prediction but also provides a framework for integrating quantum logic into real-world health diagnostics.

In Epoch-1, the classical K-Medoids method achieved an accuracy of 71% with a total shortest distance of 884 (Table 3). In Epoch-2, its performance improved to an accuracy of 88% with a total shortest distance of 961 (Table 4). On the other hand, the quantum-based K-Medoids method initially achieved a lower accuracy of 60% in Epoch 1, with a total distance of 1172.38 (Table 5). However, in Epoch 2, it achieved the same accuracy as the classical method—88%—albeit with a higher total distance of 1,359.06 (Table 6). This suggests that quantum-based clustering can match the accuracy of classical methods after adequate training, although consistency and parameter stability remain areas for further refinement.

While both approaches achieved identical accuracy in the final epoch, performance variability across epochs indicates that the quantum method may require more precise parameter tuning or stabilization mechanisms. Moreover, although accuracy serves as the primary performance metric, further analysis is necessary to examine influencing factors such as dataset size, class distribution, and initial medoid selection. The absence of statistical significance testing (e.g., p-values or confidence intervals) also limits the interpretability of the performance comparison, raising questions about whether observed differences are meaningful or due to random variations.

Additionally, no direct comparisons have been made with other clustering algorithms such as K-Means, DBSCAN, or quantum alternatives like Quantum K-Means or Quantum Annealing. As a result, the relative performance and novelty of the proposed method in the broader context of clustering algorithms remain unclear. Furthermore, the study does not yet discuss algorithmic complexity or computational efficiency in quantum environments, which are crucial for assessing practical feasibility.

In summary, the quantum-enhanced K-Medoids model demonstrates encouraging accuracy in clustering medical data, particularly for the classification of early-stage lung cancer. However, to validate its broader applicability, future research should place greater emphasis on assessing the model's stability across multiple runs and its scalability when applied to larger and more diverse datasets. Evaluating its performance under varying data conditions, such as imbalanced classes or noisy attributes, will be essential to understand its robustness. Furthermore, benchmarking the model against other quantum and classical clustering algorithms—such as quantum k-means, spectral clustering, or DBSCAN—can provide deeper insights into its relative strengths and limitations. Integration with hybrid quantum-classical computing systems also presents a valuable direction for enhancing model accuracy, reducing processing time, and enabling real-time decision support, particularly in critical domains such as medical diagnostics. This exploration will help bridge the gap between experimental simulation and practical deployment in healthcare settings.

4. CONCLUSION

This study successfully integrates quantum computing into the K-Medoids clustering algorithm by incorporating Manhattan distance calculations, demonstrating a comparable level of accuracy to its classical counterpart. The findings suggest that quantum-enhanced clustering holds promise for improving data grouping performance in data mining tasks. However, several limitations must be acknowledged. First, the scalability of the proposed method to larger and more complex datasets has not been evaluated, which may impact its applicability in real-world scenarios. Second, the study does not address the technical challenges associated with implementing quantum algorithms on current quantum hardware, including noise, qubit decoherence, and limited qubit availability. Third, the practical deployment of this model, particularly in healthcare applications such as lung cancer prediction, remains unexplored and requires validation in real clinical environments. To enhance the robustness and practical relevance of the approach, future research should focus on testing the method with high-dimensional, imbalanced, and larger-scale datasets. Additionally, exploring hybrid quantum-classical algorithms and benchmarking against other clustering techniques may further optimize performance. Investigating the feasibility of integrating this model into decision support systems within clinical settings could also open new pathways for real-world adoption. These directions will help ensure that the proposed quantum K-Medoids model transitions from theoretical contributions to practical impact and broader usability.

5. ACKNOWLEDGEMENTS

The authors would like to express their sincere gratitude to STIKOM Tunas Bangsa, Universitas BSI and STMIK Antar Bangsa for the support. This support has been instrumental in facilitating the successful completion of this research.

6. DECLARATIONS

AUTHOR CONTIBUTION

Conceptualization: Solikhun, Mochamad Wahyudi. Methodology: Solikhun, Mochamad Wahyudi. Discussion of results: Solikhun, Mochamad Wahyudi, Lise Pujiastuti. Original Draft: Solikhun, Mochamad Wahyudi, Lise Pujiastuti. Writing Review and Editing: Lise Pujiastuti. Resources: Solikhun, Mochamad Wahyudi, Lise Pujiastuti. Supervision: Solikhun, Mochamad Wahyudi. Approval of the final text: Solikhun, Mochamad Wahyudi.

FUNDING STATEMENT

This research was supported and funded by STIKOM Tunas Bangsa, Universitas BSI, and STMIK Antar Bangsa. The authors gratefully acknowledge this support, which has contributed to the successful completion of this study.

COMPETING INTEREST

The authors declare that there are no competing interests related to this research, its findings, or its publication.

REFERENCES

[1] Y.-M. Li, H.-L. Liu, S.-J. Pan, S.-J. Qin, F. Gao, D.-X. Sun, and Q.-Y. Wen, “Quantum k -medoids algorithm using parallel amplitude estimation,” *Physical Review A*, vol. 107, no. 2, p. 022421, Feb. 2023, https://doi.org/10.1103/PhysRevA.107.022421.

[2] N. Gao, D. Li, A. Mishra, J. Yan, K. Simonov, and G. Chiribella, “Measuring Incompatibility and Clustering Quantum Ob-

servables with a Quantum Switch," *Physical Review Letters*, vol. 130, no. 17, p. 170201, Apr. 2023, https://doi.org/10.1103/PhysRevLett.130.170201.

[3] K. Hulliyah and S. Solikhun, "Q-Madaline: Madaline Based On Qubit," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 7, no. 5, pp. 1003–1008, Aug. 2023, https://doi.org/10.29207/resti.v7i5.5080.

[4] J. L. Pereira, L. Banchi, and S. Pirandola, "Quantum-Enhanced Cluster Detection in Physical Images," *Physical Review Applied*, vol. 19, no. 5, p. 054031, May 2023, https://doi.org/10.1103/PhysRevApplied.19.054031.

[5] N. Piatkowski, T. Gerlach, R. Hugues, R. Sifa, C. Bauckhage, and F. Barbaresco, "Towards Bundle Adjustment for Satellite Imaging via Quantum Machine Learning," in *2022 25th International Conference on Information Fusion (FUSION)*. Linköping, Sweden: IEEE, Jul. 2022, pp. 1–8, https://doi.org/10.23919/FUSION49751.2022.9841388.

[6] L. Zahrotun, U. Linarti, B. H. T. Suandi As, H. Kurnia, and L. Y. Sabila, "Comparison of K-Medoids Method and Analytical Hierarchy Clustering on Students' Data Grouping," *JOIV : International Journal on Informatics Visualization*, vol. 7, no. 2, p. 446, May 2023, https://doi.org/10.30630/joiv.7.2.1204.

[7] S. Al-Otaibi, V. Cherappa, T. Thangarajan, R. Shanmugam, P. Ananth, and S. Arulswamy, "Hybrid K-Medoids with Energy-Efficient Sunflower Optimization Algorithm for Wireless Sensor Networks," *Sustainability*, vol. 15, no. 7, p. 5759, Mar. 2023, https://doi.org/10.3390/su15075759.

[8] F. Faisal, L. A. G. Giopani, M. F. Fitriah, Z. C. D. Dwynne, S. S. H. Helma, and M. Mustakim, "Perbandingan Algoritma K-Means dan K-Medoids Untuk Pengelompokan Suhu di Provinsi Riau: Comparison of K-Means and K-Medoids Algorithms for Temperature Grouping in Riau Province," *Indonesian Journal of Informatic Research and Software Engineering (IJIRSE)*, vol. 2, no. 2, pp. 128–134, Sep. 2022, https://doi.org/10.57152/ijirse.v2i2.434.

[9] S. Samudi, S. Widodo, and H. Brawijaya, "The K-Medoids Clustering Method for Learning Applications during the COVID-19 Pandemic," *SinkrOn*, vol. 5, no. 1, p. 116, Oct. 2020, https://doi.org/10.33395/sinkron.v5i1.10649.

[10] Z. Wu, L. Jin, J. Zhao, L. Jing, and L. Chen, "Research on Segmenting E-Commerce Customer through an Improved K-Medoids Clustering Algorithm," *Computational Intelligence and Neuroscience*, vol. 2022, pp. 1–10, Jun. 2022, https://doi.org/10.1155/2022/9930613.

[11] Mustakim, M. Z. Fauzi, Mustafa, A. Abdullah, and Rohayati, "Clustering of Public Opinion on Natural Disasters in Indonesia Using DBSCAN and K-Medoids Algorithms," *Journal of Physics: Conference Series*, vol. 1783, no. 1, p. 012016, Feb. 2021, https://doi.org/10.1088/1742-6596/1783/1/012016.

[This page intentionally left blank.]