Sciences Methods and Technologies International Journal (SciMeTech)

ISSN: 3085-5284, journal homepage: www.scimetech.com

Diabetes Prediction using Quantum Artificial Intelligence

Btissam Elaziz[1], Nadia Hachoumi[2], Charaf Eddine AIT ZAOUIAT[3], Yassin LAAZIZ[1]

1 LabTIC, ENSA of Tangier, Abdelmalek Essaadi University, Morocco.

2 Cadi Ayyad University, UCA, Higher School of Technology of Essaouira, LaRTID Laboratory, Marrakech, Morocco.

3Polydisciplinary Faculty of Sidi Bennour, Chouaîb Doukkali University, Morocco.

How to cite:

Btissam Elaziz, Nadia Hachoumi, Charaf Eddine AIT ZAOUIAT, Yassin LAAZIZ, "Diabetes Prediction using Quantum Artificial Intelligence", Sciences Methods and Technologies International Journal (SciMeTech), Vol 2, Issue 1, p 91-98

Keywords:

Diabetes Prediction
Quantum Artificial Intelligence
Variational Quantum Classifier
Quantum KMeans
Quantum computing

Abstract

The global epidemic of diabetes, driven by lifestyle and genetic factors and leading to severe complications, necessitates the use of Artificial Intelligence to predict its progression for enabling early, life-saving interventions. Quantum Artificial Intelligence combines principles of Quantum Computing and Artificial Intelligence, offering potential improvements in efficiency and accuracy for complex data-driven tasks. In this paper, we explore the use of Quantum Artificial Intelligence algorithms for binary classification, focusing on clustering methods for diabetes prediction. Specifically, we compare the performance of classical KMeans clustering with a quantum-enhanced version based on a Variational Quantum Classifier model. The Quantum KMeans algorithm includes quantum properties such as superposition and entanglement to enhance cluster formation and classification performance. Using the Pima Indian diabetes dataset, preprocessed with standard scaling, we evaluated and compared the results of classical KMeans and the quantum special model. Our experiments demonstrate that the quantum-enhanced KMeans achieves competitive performance compared to the classical approach, suggesting that Quantum Artificial Intelligence techniques offres more effective and scalable clustering in diabetes binary classification tasks. These findings highlight the complementary role of quantum computing in advancing classical machine learning methods for medical diagnosis.

I. Introduction

Diabetes has become a widespread global health crisis, largely driven by factors such as poor diet, sedentary lifestyles, and genetic predisposition. This chronic condition poses severe dangers, including heart disease, kidney failure, nerve damage, and vision loss, significantly impacting quality of life and increasing mortality risk. Given its progressive nature, early intervention is crucial, highlighting the importance of utilizing artificial intelligence (AI), it can analyse complex patient data to predict the progression of diabetes with remarkable accuracy, enabling personalized treatment plans and timely preventive measures to mitigate complications and improve long-term health outcomes.

AI has profoundly augmented human capabilities and contributed to societal advancement across multiple disciplines by automating complex operations, enhancing decision-making frameworks, and delivering customized experiences [2]. AI utilizes sophisticated algorithms and statistical methodologies to analyse and interpret intricate datasets, supporting strategic planning, decision support systems, predictive analytics, intelligent automation, and various other applications spanning sectors such as healthcare, finance, education, transportation, and defence [3]. The training and deployment of AI models can be computationally intensive and time-consuming, contingent upon data volume and methodological complexity. To address these challenges, research efforts focus on enhancing the efficiency and scalability of AI algorithms and architectures, alongside leveraging advancements in hardware technologies such as Graphics Processing Units (GPUs) and Tensor Processing Units (TPUs). Additionally, distributed machine learning techniques are employed to expedite both model training and inference processes [5]. Quantum computing offers the opportunity to enhances the performances of AI algorithms using parallel processing capabilities.

Quantum Computing (QC), is a multidisciplinary domain grounded in quantum theory with diverse applications in software engineering, offers an alternative computational paradigm aimed at reducing the time required to address complex problems while enhancing performance [6], QC employs the fundamental principles of quantum mechanics to manipulate information and resolve computational challenges that exceed the capabilities of classical computing systems. Unlike classical bits, which exist strictly in binary states of 0 or 1, quantum bits (qubits) can occupy a superposition of multiple states simultaneously. This characteristic allows quantum devices to execute parallel processing, thereby markedly increasing computational efficiency compared to classical machines. Although quantum computing is still in an early phase of advancement, it demonstrates significant potential for solving complex problems in classical computational AI techniques and algorithms [1]. In this paper we explore the combination of QC an AI in Quantum Artificial Intelligence (QAI) to integrate the strengths of both AI and QC and tackle challenges across multiple fields. By leveraging quantum mechanical principles, QAI seeks to augment computational efficiency and improve the performance metrics of existing ML algorithms. As QC technologies advance, the interplay between quantum algorithms and machine learning is anticipated to catalyse novel solutions to computational problems that were previously intractable [7].

This study focuses on binary classification using the Pima Indian Diabetes dataset, applying the KMeans clustering algorithm. The Pima Indian Diabetes dataset is widely used for diabetes research due to its comprehensive clinical data. Diabetes remains a major global health issue, with Type 2 diabetes mellitus being particularly prevalent and a serious public health concern. According to recent statistics [8], millions worldwide are affected by diabetes, with many cases remaining undiagnosed and leading to significant mortality. Effective detection and classification of diabetes are crucial for timely intervention, and clustering techniques like KMeans offer promising approaches to identify patterns within this dataset.

This section introduced the problem domain and outlined the scope of the study. The remainder of this manuscript is organized as follows: Section 2 provides essential background and reviews related work on QC and QAI. Section 3 describes the methodology used to develop and train the models. Section 4 presents the results. Lastly, Section 5 summarizes the key findings, offers conclusions, and proposes directions for future research.

II. Related Works

According to the authors in [9], classical computing encodes information in bits that take the value 0 or 1, whereas quantum computing uses quantum bits, or qubits, to represent information.

Quantum computers are not merely improved or accelerated versions of traditional classical machines; rather, they introduce an entirely new framework for processing information. Whereas classical bits can assume only the value 0 or 1, qubits have the remarkable capability to encode and maintain multiple combinations of these values simultaneously [10].

This capacity to exist in multiple states at once is known as superposition. This property allows information to be processed in a parallel and significantly expanded manner, offering computational possibilities far beyond those of classical systems.

A quantum state [11] refers to any allowable state of a quantum mechanical system or quantum hardware. Many naturally occurring quantum two-level systems can function as qubits, such as the electronic states of ions or the electron spin of atoms embedded in silicon. When a quantum system uses a computational unit with more than two levels, the unit is referred to as a qudit. Compared with qubits, qudits offer a larger state space, which can decrease circuit complexity and improve the overall efficiency of quantum algorithms.

The authors state [12] that in quantum computing, pairs of qubits can be prepared in an entangled state, meaning that any change applied to one qubit is instantly reflected in the state of the other. Entanglement has two remarkable properties: it is intrinsically private and allows perfect coordination between the qubits. Importantly, this correlation persists even when the qubits are separated by large distances.

Quantum computers [13] are developed using several hardware platforms, with the most common being superconducting circuits, photonic architectures, trapped ions, and quantum dots. Each platform comes with its own strengths and limitations, such as the ability to generate stronger entanglement or maintain coherence for longer periods. [14] A key objective in this field is achieving quantum supremacy, which describes the point at which a quantum device can carry out a computation that no classical computer—even the most advanced supercomputer—can perform. It is also expected that future quantum systems will not depend on a single hardware technology; instead, they may combine multiple approaches to improve efficiency and overall performance. Table 1 provides a comparison between classical and quantum computing based on several criteria.

Dirac notation is commonly used to represent quantum states in quantum mechanics. A quantum state is expressed as a column vector whose components are complex probability amplitudes corresponding to different measurable states. These amplitudes satisfy the normalization condition, meaning the sum of the squared magnitudes equals one [16]. A qubit, the fundamental unit of quantum information, can be either a physical qubit or a logical qubit composed of multiple physical qubits. Qubits can exist in the basic's states $|0\rangle$|and $|1\rangle$|, or in any superposition of these states [17]. A quantum state can also be described by a wave function ψ(x), as shown in the following expression:

$$|\Psi\rangle = \alpha|a\rangle + \beta|b\rangle \quad (1)$$

$$\text{Where } \alpha , \beta \in \mathbb{C} \quad |\alpha|^2+|\beta|^2 = 1 \quad (2)$$

The controlled-NOT or CNOT gate is the most prevalent two-qubit operation, which inverts the state of the second qubit (target) only when the first qubit (control) is in the $|1\rangle$ state. Any unitary quantum circuit can be expressed as a composition of single-qubit gates and CNOT gates. Since the two-qubit CNOT gate generally requires significantly more execution time on physical quantum devices compared to single-qubit gates, the complexity or cost of a quantum circuit is often evaluated based on the number of CNOT gates it contains [18]. Additionally, there exist three-qubit gates, such as the Toffoli gate, also referred to as the controlled-controlled-NOT (CCNOT) gate.

Table 1. Classical vs. Quantum computing

Feature	Quantum	Classical
Theory	Quantum mechanics	Classical physics
Computation	Probabilistic	Deterministic
Operations	Linear algebra operations	Boolean algebra operations
Information Storage	Qubits, qudits	Bits
System state	Continuos possibles states in superposition	Discrete number of possible states
Technology	Superconducting loops, trapped ions, quantum dots, etc.	Transistors
Applications	Complex problems, optimization, simulation	General purpose
Error rate	High	Low
Environment	Ultracold	Room temperature
Computing power	Exponential growth	Linear growth
Processing	QPU	CPU

III. Methodology of Quantum Encoding Approaches Applied to Diabetes Prediction

In this work, the objective is to enhance binary classification performance by integrating K-means clustering with a hybrid classical–quantum model. This model is designed to map input data into a quantum feature space, optimize variational parameters, and perform accurate binary classification. Our experiments are conducted using the Pima Indians Diabetes Dataset. Each input feature is encoded through an RY rotation applied to a dedicated qubit, while the trainable model parameters consist of RY and RZ rotations learned across multiple variational layers. Quantum entanglement is introduced through a chain of CNOT gates connecting adjacent qubits. The output is obtained by measuring the expectation value of the Pauli-Z operator on the first qubit, which is then converted into a classification probability. Overall, the circuit implements a Variational Quantum Classifier (VQC), as shown in Figure 1.

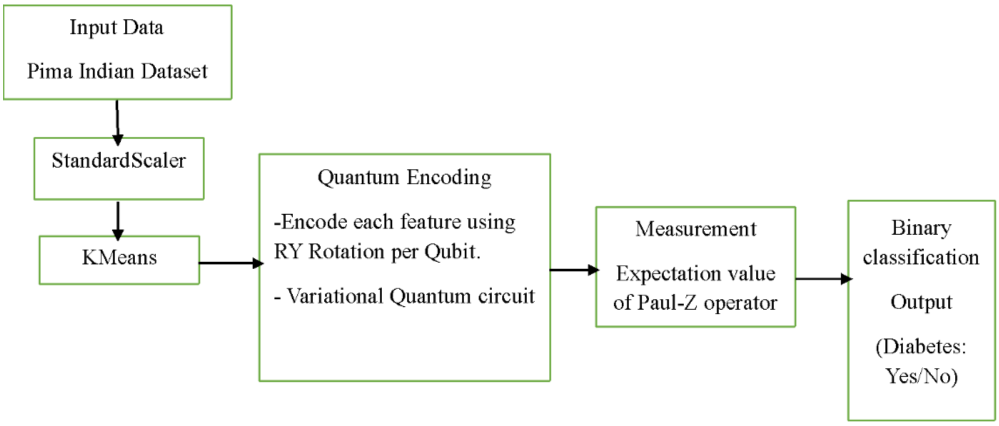

Figure 1: Hybrid Classical-Quantum Flowchart

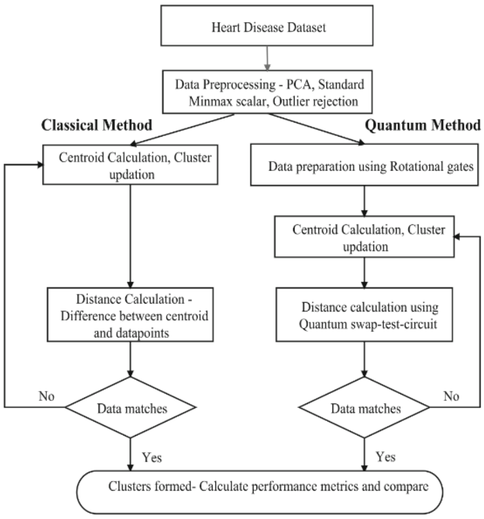

Figure 2: Classical vs Quantum K-Means Workflow

This pseudocode, presented in Figure 2, describes a quantum model employed for binary classification. The objective is to transform a classical data vector (normalized features) into quantum states, manipulate them using quantum rotations and gates, and subsequently measure the outcome to predict a class label.

Variational Quantum Classifier (VQC), is a binary quantum classification algorithm based on:

- Angle encoding of classical data
- A parameterized variational circuit (ansatz)
- Measurement of a Pauli-Z observable
- A simple decision rule

It is commonly used in quantum machine learning (QML).

Inputs

- n: number of qubits (also the number of features)
- L: number of variational layers
- x = (x_0, x_1, …, x_{n-1}): normalized classical feature vector
- θ: trainable parameters (rotation angles)

Data Encoding (Feature Map)

The classical data is encoded into a quantum state using Y-rotations.

- Initial state:

$$| 0\rangle^{\otimes n}$$

- Encoding:

$$|\Psi_0\rangle = \otimes_{i=0}^{n-1} \; R_y(x_i) \; |0\rangle$$

- Y-rotation gate:

$$R_y(\alpha) = e^{i\frac{\alpha}{2}Y} = cos(\alpha/2) \; - sin(\alpha/2) \; sin(\alpha/2) \; cos(\alpha/2)$$

Purpose: Map classical features into quantum amplitudes.

Variational Ansatz (Trainable Circuit)
For each layer l=0,…,L-1:

a) Trainable Y-rotations

$$R_y(\theta_{y,i}^{l})\ on\ each\ qubit$$

b) Entanglement

A chain of CNOT gates:

$$CNOT_{i,i+1}, i = 0 \dots n-2$$

c) Trainable Z-rotations

$$R_z(\theta_{z,i}^{l})\ on\ each\ qubit$$

Purpose:

- Learn complex decision boundaries
- Introduce entanglement and nonlinearity

Measurement
Measure the Pauli-Z operator on the first qubit:

$$\langle \mathrm{Z} \rangle = \langle \Psi_{out} | Z_0 | \Psi_{out} \rangle$$

Where :

$$Z = (100 \quad -1)$$

The result is in the range **[-1, 1]**

Probability of Class 1
The expectation value is converted into a probability:

- $p_1 \in [0,1]$
- Interpreted as the probability of class 1

Prediction Rule

$$If\ p_1 > 0.5 \Rightarrow class = 1 \quad Else\ class = 0$$

In this parameterized quantum circuit (PQC) with four qubits and two layers (figure 3), commonly used in variational quantum algorithms. Each qubit applies an Ry rotation to encode input data, followed by trainable Ry and Rz rotations that allow the circuit to learn. Controlled-NOT (CNOT) gates create connections between qubits to enable information sharing. The circuit's output is measured using the Pauli-Z operator, and a classical optimizer updates the parameters to minimize the error. In real applications, such as with the Pima Indians Diabetes dataset, the number of qubits matches the number of features—in this case, seven qubits—while the figure uses four for simplicity. The data is encoded using angle encoding, where each feature controls a rotation on a qubit. The quantum states are then simulated, and their real parts are extracted as new feature vectors. These vectors are clustered using classical KMeans, combining quantum feature mapping with classical methods to improve classification.

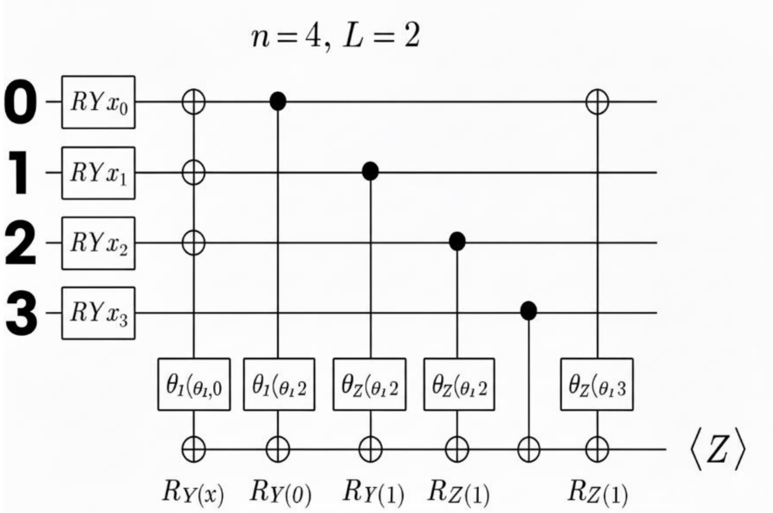

Figure 3: Quantum Circuit Encoding

To evaluate the performance of our hybrid classical–quantum binary classifier, which leverages K-means clustering for preprocessing, we employ three established metrics: ROC-AUC, F1-Score, and Accuracy.

In our model, each input feature from the Pima Indians Diabetes Dataset is encoded into the quantum feature space via an RY rotation on a dedicated qubit. A variational quantum circuit, composed of learnable layers of RY and RZ rotations entangled by a chain of CNOT gates, processes this state. The output is obtained by measuring the expectation value of the Pauli-Z operator on the first qubit, which is then interpreted as a classification probability. Crucially, model assessment is based on these final predictions.

The ROC-AUC (Area Under the Receiver Operating Characteristic Curve) quantifies the model's discriminative capacity across all classification thresholds. It is derived from the True Positive Rate (TPR/Recall) and False Positive Rate (FPR):

$$TPR = \frac{TP}{TP + FN} \quad and \ FPR = \frac{FP}{FP + TN}$$

where TP (True Positive) and TN (True Negative) are correct classifications, and FP (False Positive) and FN (False Negative) are errors.

The F1-Score, the harmonic mean of precision and recall, provides a balanced measure for the positive class, which is vital for imbalanced medical datasets:

$$F_1 + \ Score = \frac{2\,TP}{2TP + FP + FN}$$

Finally, **Accuracy** measures the overall fraction of correct predictions:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

These metrics offer a comprehensive, model-agnostic evaluation of our hybrid quantum-classical pipeline, focusing solely on its final predictive efficacy for diabetes classification.

IV. Result And Discussion

The results in Table 2 highlight the impact of preprocessing and quantum approaches on clustering performance when applied to the Pima Indians Diabetes dataset, which contains 7 input features related to health indicators for diabetes prediction and is known for its imbalanced class distribution. Classical KMeans without scaling shows relatively low precision (0.46) and F1-score (0.48), indicating poor cluster quality due to feature scale imbalance. Applying StandardScaler significantly improves performance (precision 0.57, F1-score 0.56), confirming that normalization is essential for distance-based algorithms like KMeans. However, the Quantum KMeans with StandardScaler achieves the highest scores (precision 0.72, recall 0.68, F1-score 0.68), demonstrating the advantage of quantum clustering. This improvement can be attributed to the ability of quantum algorithms to leverage superposition and entanglement, enabling richer feature representation and better separation in high-dimensional Hilbert space. Moreover, quantum methods can handle imbalanced data more effectively because they represent data points as quantum states in a continuous feature space, allowing subtle differences between minority and majority classes to be amplified through entanglement and interference. These findings suggest that quantum-enhanced clustering can outperform classical methods for complex and imbalanced datasets like Pima Indians Diabetes, provided that proper preprocessing is applied.

Table 2: Performance Comparison: Classical

	Precision	Recall	F1-score
KMeans	0.46	0.51	0.48
KMeans+Standardscaler	0.57	0.63	0.56
QKMeans+Standardscaler	0.72	0.68	0.68

V. Conclusion

This study demonstrates that quantum-enhanced clustering can significantly outperform classical methods when applied to complex and imbalanced datasets such as the Pima Indians Diabetes dataset. While classical KMeans improved with preprocessing (StandardScaler), Quantum KMeans achieved the highest precision (0.72), recall (0.68), and F1-score (0.68), highlighting the benefits of quantum principles like superposition and entanglement for richer feature representation and better separation in high-dimensional spaces. These results suggest that quantum machine learning offers promising advantages for healthcare analytics and other domains where data complexity and imbalance are common challenges.
Future work could explore more advanced quantum feature maps, deeper quantum circuits, and different encoding strategies such as amplitude encoding. Testing the method on larger datasets and evaluating performance on real quantum hardware would also help assess scalability and robustness. Additionally, studying noise mitigation techniques may improve the stability of quantum-based feature extraction.

References

[1] Lund, B. D., & Shahriar, S. (2025). Quantum Computing: A Concise Introduction. Encyclopedia, 5(4), 173.
[2] Siam, M. K. H., Bhattacharjee, M., Mahmud, S., Sarkar, M. S., & Rana, M. M. (2024). The impact of machine learning on society: An analysis of current trends and future implications. arXiv preprint arXiv:2404.10204.
[3] Khan, I. U., Ouaissa, M., Ouaissa, M., Fayaz, M., & Ullah, R. (Eds.). (2024). Artificial intelligence for intelligent systems: Fundamentals, challenges, and applications.
[4] Shalev-Shwartz, S., & Ben-David, S. (2014). Understanding machine learning: From theory to algorithms. Cambridge university press.
[5] Jouppi, N. P., Young, C., Patil, N., Patterson, D., Agrawal, G., Bajwa, R., ... & Yoon, D. H. (2017, June). In-datacenter performance analysis of a tensor processing unit. In Proceedings of the 44th annual international symposium on computer architecture (pp. 1-12).
[6] Mitarai, K., Negoro, M., Kitagawa, M., & Fujii, K. (2018). Quantum circuit learning. Physical Review A, 98(3), 032309.
[7] McArdle, S., Jones, T., Endo, S., Li, Y., Benjamin, S. C., & Yuan, X. (2019). Variational ansatz-based quantum simulation of imaginary time evolution. npj Quantum Information, 5(1), 75.
[8] Maheshwari, D., Sierra-Sosa, D., & Garcia-Zapirain, B. (2021). Variational quantum classifier for binary [10] classification: Real vs synthetic dataset. IEEE access, 10, 3705-3715.
Chae, E., Choi, J., & Kim, J. (2024). An elementary review on basic principles and developments of qubits for quantum computing. Nano Convergence, 11(1), 11.
[11] Wang, Y., Hu, Z., Sanders, B. C., & Kais, S. (2020). Qudits and high-dimensional quantum computing. Frontiers in Physics, 8, 589504.
[12] Yin, J., Cao, Y., Li, Y. H., Liao, S. K., Zhang, L., Ren, J. G., ... & Pan, J. W. (2017). Satellite-based entanglement distribution over 1200 kilometers. Science, 356(6343), 1140-1144.
[13] Sood, S. K. (2023). Quantum computing review: A decade of research. IEEE Transactions on Engineering Management, 71, 6662-6676.
[14] Arute, F., Arya, K., Babbush, R., Bacon, D., Bardin, J. C., Barends, R., ... & Martinis, J. M. (2019). Quantum supremacy using a programmable superconducting processor. Nature, 574(7779), 505-510.
[15] Ajibosin, S. S., & Cetinkaya, D. (2024). Implementation and performance evaluation of quantum machine learning algorithms for binary classification. Software, 3(4), 498-513.

[16] Ahmed, I., Zafar, S., & Saghir, A. (2022). QUANTUM MECHANICS: A COMPREHENSIVE STUDY OF FUNDAMENTAL PRINCIPLES AND APPLICATIONS. WORLDWIDE JOURNAL OF PHYSICS, 1(1), 53-46.
[17] Sakurai, J. J., & Napolitano, J. (2020). Modern quantum mechanics. Cambridge University Press.
[18] Bharti, K., Cervera-Lierta, A., Kyaw, T. H., Haug, T., Alperin-Lea, S., Anand, A., ... & Aspuru-Guzik, A. (2022). Noisy intermediate-scale quantum algorithms. Reviews of Modern Physics, 94(1), 015004.