IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 10 October 2025, accepted 31 October 2025, date of publication 4 November 2025, date of current version 13 November 2025.

Digital Object Identifier 10.1109/ACCESS.2025.3628830

RESEARCH ARTICLE

# Comparative Analysis of Quantum Encoding Techniques for Biomarker Classification

**AICHA EUTAMENE[1], ISLAM DJEMMAL[2], HACENE BELHADEF[3], AND A. M. MUTAWA[4], (Senior Member, IEEE)**

[1]University Fréres Mentouri-Constantine 1, Constantine 25000, Algeria
[2]LIRE Laboratory, Department of Mathematics and Computer Science, Institute of Sciences and Technology, University Center Abdelhafid Boussouf, Mila 43000, Algeria
[3]NTIC Faculty, University of Constantine 2-Abdelhamid Mehri, Constantine 25016, Algeria
[4]Department of Computer Engineering, College of Engineering and Petroleum, Kuwait University, Kuwait City 13060, Kuwait

Corresponding authors: Aicha Eutamene (aicha.eutamene@umc.edu.dz) and A. M. Mutawa (dr.mutawa@ku.edu.kw)

**ABSTRACT** Biomarker classification represents a significant task in biological, medical diagnostics, and bioinformatics, where high-dimensional datasets represent challenges to classical machine learning models. The study presented in this paper focuses on the impact of data encoding and explores quantum machine learning algorithms for biomarker classification. Three quantum embedding methods; Angle Embedding, Amplitude Embedding, and IQP (Instantaneous Quantum Polynomial) Embedding, were combined with Quantum Support Vector Machines (QSVM) and Quantum k-Nearest Neighbors (QkNN) and compared against classical SVM and KNN baselines. Principal Component Analysis (PCA) was applied as a preprocessing step to reduce dimensionality. Experiments were conducted on the CuMiDa breast cancer dataset using quantum simulators. Results show that IQP embedding and Angle embedding significantly outperform classical models with QSVM, achieving the best performance, where IQP embedding reaching 93.1% accuracy and an F1-score of 0.93 in higher-dimensional settings, outperforming both the other quantum embeddings and the classical SVM. For QkNN, performance depended on dimensionality, with IQP achieving the strongest results in lower dimensions, including a precision of 87.1%, surpassing both the other quantum embeddings and classical KNN. In higher dimensions, Angle embedding yielded the best quantum results, though classical KNN remained superior. Amplitude embedding consistently underperformed in QSVM but was more competitive in QkNN. These findings demonstrate that advanced quantum data encoding, such as IQP and Angle, can enhance biomarker classification performance and highlight the promise of hybrid quantum-classical approaches in biomedical applications.

**INDEX TERMS** Biomarker classification, quantum data encoding, quantum machine learning, breast cancer, hybrid quantum-classical models.

## I. INTRODUCTION

The rapid growth of omics and biomedical imaging technologies has greatly expanded the field of biomarker discovery. This progress is fundamental for early diagnosis, disease stratification, and precision medicine. However, these datasets are typically high-dimensional, noisy, and imbalanced, posing major challenges for traditional machine learning models that depend on linear separability and large training samples. [1], [2]. Classical machine learning algorithms such as Support Vector Machines (SVM), k-Nearest Neighbors (kNN), and Random Forests have been widely applied to biomarker classification [3]. However, they face limitations when dealing with the curse of dimensionality and complex feature interactions found in biomedical data. Feature selection and dimensionality reduction methods such as Principal Component Analysis (PCA) partially mitigate these issues, but at the cost of potentially losing discriminative biological information. These limitations motivate the exploration of new paradigms that can better exploit the

The associate editor coordinating the review of this manuscript and approving it for publication was Abdullah ILIYASU.

complex structure of biomedical data. Quantum Machine Learning (QML) has recently emerged as a hybrid approach that uses quantum mechanics to map classical data into exponentially larger feature spaces [4]. The central element in these models is the quantum data encoding (embedding), that defines how classical vectors are transformed into quantum states. Despite growing interest in QML for biomedical tasks, there remains a notable gap: most existing studies focus on a single encoding method or specific application, without a unified comparison across encoding techniques. This lack of systematic evaluation makes it difficult to identify which encoding strategies are most effective for biomarker classification problems [5], [6]. Although Quantum Machine Learning (QML) shows great potential, its practical application in biomedical data analysis still faces significant technical and theoretical challenges. Biomedical datasets, particularly those obtained from omics and imaging technologies, often suffer from limited sample sizes, missing data, and strong inter-feature correlations, which hinder the ability of models to generalize beyond the training set [7], [8]. In addition, quantum models are highly sensitive to noise and decoherence, both of which can substantially reduce circuit fidelity on current Noisy Intermediate-Scale Quantum (NISQ) devices [9]. These issues underscore the necessity of developing resource-efficient quantum models that can maintain accuracy under realistic hardware constraints. Consequently, achieving an optimal balance between circuit depth, qubit count, and expressivity is a key consideration in designing scalable quantum architectures for biomedical applications [10]. While accuracy remains a key metric, the computational and physical cost of implementing quantum circuits cannot be overlooked. Quantum algorithms must be assessed not only for predictive performance but also for their scalability on near-term devices. Circuit depth, gate complexity, and qubit requirements directly influence both runtime and susceptibility to errors. To address this gap, the present study investigates how different quantum data encoding schemes, specifically Angle, Amplitude, and Instantaneous Quantum Polynomial (IQP) embeddings, influence the classification performance of biomarker datasets. It further examines how quantum classifiers such as QSVM and QkNN perform relative to their classical counterparts in terms of accuracy and robustness. Finally, the study explores the impact of dimensionality reduction through Principal Component Analysis (PCA) on the stability and computational efficiency of these quantum embedding techniques. Based on these research questions, our main objectives are as follows: the first is to implement and benchmark three quantum encoding techniques; Angle, Amplitude, and IQP on real biomarker dataset; and second aims to evaluate hybrid quantum-classical models (QSVM and QkNN) against their classical counterparts (SVM and kNN) across multiple PCA configurations; and the final one focuses on analyzing the computational resource requirements of these models to assess their scalability on near-term quantum devices. This study makes four key contributions to the existing literature, summarized as follows. First, it provides a systematic comparison of three commonly used quantum embedding techniques for biomarker classification. Second, it introduces a hybrid benchmarking framework that evaluates quantum and classical models under identical experimental conditions. Third, it offers a resource-aware assessment that quantifies circuit cost and feasibility within the constraints of near-term NISQ devices. Finally, it presents a conceptual framework linking encoding strategies, data dimensionality, and classification performance, offering practical guidelines for selecting optimal encodings in biomedical quantum machine learning applications.

The rest of the paper is organized as follows: Section II reviews classical and quantum methods for biomarker classification algorithms, outlines key quantum encoding approaches and identifies gaps motivating this comparative study. Section III describes the methodology, including the dataset, preprocessing, quantum encoding and classifier models. Section IV presents the experimental results and comparative analysis of the encoding techniques. Section V discusses key findings, limitations, and perspectives. Finally, Section VI concludes the paper and outlines future research directions.

## II. LITERATURE REVIEW

### A. OVERVIEW OF CLASSICAL BIOMARKER CLASSIFICATION TECHNIQUES

Classical approaches for biomarker classification have relied extensively on statistical learning and machine learning techniques. Their pipeline typically combines the following steps: preprocessing, feature selection or dimensionality reduction (e.g., PCA), and supervised classifiers. Linear models such as logistic regression and SVM have demonstrated effectiveness for high-dimensional datasets by identifying optimal separating hyperplanes. Other methods include KNN, while Ensemble methods such as Random Forests and Gradient-Boosting exploit ensemble learning to enhance predictive performance. In recent years, particularly for large biomedical labeled datasets, deep learning models have been applied successfully. They gained popularity for their ability to model complex non-linear relationships in large biomedical datasets. Practical biomedical studies often rely on PCA or other projection methods to reduce noise and avoid overfitting, since biomarker datasets are frequently high-dimensional and sample-scarce. These classical methods remain strong baselines and are still widely used in biomarker studies, including breast cancer diagnostics. However, they often encounter limitations when implemented on high-dimensional, sparse, or noisy biomarker data, which motivates the necessity for the investigation of quantum-enhanced approaches [11], [12].

### B. RECENT QML WORK APPLIED TO BIOMARKERS AND BIOMEDICAL DATA

Several recent studies have explored the application of quantum neural networks (QNNs) and Quantum machine learning techniques in the field of biomarker classification and biomarker discovery. Some notable studies include:

1) **Biomarker Discovery with Quantum Neural Networks:** This study proposes a QNN architecture to identify genetic biomarkers associated with activation pathways using maximum relevance and minimum redundancy criteria to evaluate candidate biomarker sets [13].
2) **Integration of Multi-omics Data and Quantum Machine Learning for Lung Cancer:** This research focuses on the integration of multi-omics data and the use of feature selection techniques to classify Lung cancer subtypes, highlighting the potential of quantum machine learning in this context [14].
3) **Potential of Quantum Machine Learning to Solve Real-World Cancer Classification Problems:** This study explores the capabilities of quantum machine learning for classifying large biological datasets, including different cancer types, and highlights its potential in biomarker discovery [15].
4) **QuanAnts Machine: A Quantum Algorithm for Biomarker Discovery:** This study proposes a quantum algorithm, a quantum analog of ant colony optimization, for biomarker discovery. The algorithm is applied to genetic networks to identify biomarkers associated with specific activation pathways [16].

These studies highlight the increasing interest in applying quantum approaches, such as quantum neural networks and machine learning methods, to biomarker classification and discovery.

### C. QUANTUM DATA-ENCODING APPROACHES

How classical vectors are mapped to quantum states is the crucial question that the quantum data encoding step must address. Several studies have conducted comparative analyses of different quantum encoding techniques, including Amplitude, Angle, and Basis encoding, by evaluating their impact on the performance of quantum machine learning algorithms. For example, a recent study has compared these three techniques, focusing on their impact on machine learning accuracy. However, this work did not specifically address biomarker classification within its empirical study [17]. Another study compared encoding methods specifically in the context of hybrid quantum-classical machine learning, using the QuClassi quantum neural network architecture to perform binary classification of the digits '3' and '6' from the MNIST dataset [18]. Furthermore, a study extensively analyzed Amplitude, Angle, and Basis encoding, demonstrating how various strategies affect improvements in quantum algorithms, while also identifying key challenges related to encoding in quantum machine learning, including scalability, computational load, and noise [19]. To develop a comprehensive overview of quantum encoding techniques and the quantum classification of biomarker biological data in general, it is essential to refer to recent work in these fields. **Table** 1 presents a selection of relevant studies.

### D. LIMITATIONS OF CURRENT APPROACHES AND MOTIVATION FOR OUR COMPARATIVE STUDY

The main limitations of this study include the use of quantum simulators rather than actual quantum hardware and dataset size constraints due to current simulator capabilities. Moreover, the lack of benchmarks and comprehensive evaluations makes it difficult to identify the most suitable quantum encoding strategies for biomedical data. Early quantum machine learning approaches have shown promise but have generally been limited to experiments with a single encoding technique. These limitations justify the present comparative study to benchmark the three embedding methods (Angle, Amplitude, IQP) with consistent preprocessing (PCA) and multiple classifier algorithms (QSVM, QkNN, classical SVM, classical KNN), and to evaluate both predictive and robustness metrics (accuracy, recall, precision, F1, Cohen's kappa, MCC) to establish clear guidelines on their strengths, weaknesses, and suitability for biomarker classification.

## III. METHODOLOGY

We conducted a series of experiments on a biomarker-based cancer classification dataset to evaluate the effect of different quantum embedding strategies on quantum machine learning performance. All simulations were performed on Google Colab, which provides a consistent and reproducible cloud-based environment. The experiments were implemented using the PennyLane quantum machine learning framework with the lightning.qubit simulator backend, enabling efficient and precise statevector simulation on classical hardware. Each session was executed on a virtual machine equipped with a 16-core CPU and 12 GB of RAM, operating in a CPU-only configuration. Random seeds were fixed across all runs to ensure reproducibility, and identical simulation settings were applied for all quantum embeddings and classifiers. This configuration ensured stable benchmarking and reliable comparison of the quantum encoding techniques without hardware-induced noise. Specifically, we investigated three widely used quantum embedding techniques in combination with two hybrid quantum–classical algorithms: the QSVM and QKNN. **Figure** 1 presents the overall pipeline of our experimental framework, which includes data preprocessing, quantum embedding, model training, and evaluation. To evaluate the robustness of the models under varying feature dimensions, we applied Principal Component Analysis (PCA) and systematically varied the number of retained components. To examine the feasibility of the proposed methods on near-term NISQ devices, we conducted a qualitative analysis of the computational resources required for each encoding strategy. The number of qubits scales directly with the dimensionality of the reduced input features. Specifically, PCA was applied with two components $d = 5$ and $d = 10$, corresponding to circuits that use 5 and 10 qubits, respectively, in all embeddings. This setup enabled

**TABLE 1.** Summary of some critical studies on quantum encoding.

| Study and Year | Quantum Encoding Techniques | PCA Application | Key Findings |
|---|---|---|---|
| LaRose et al. (2019) [20] | Robust Data Encodings | Not specified | Investigated the impact of data encodings on decision boundaries and robustness in the presence of noise. |
| Thumwanit et al. (2021) [21] | Discrete Feature Embeddings | Not specified | Proposed trainable discrete feature embeddings for quantum machine learning, enhancing classification performance. |
| Fauzi et al. (2022) [22] | Various Quantum Encoding Methods | Not specified | Analyzed several quantum encoding methods implemented on quantum circuits to improve classification accuracy. |
| Gujju et al. (2023) [23] | Various Quantum Techniques | PCA for dimensionality reduction | Reviewed supervised and unsupervised learning applications on quantum hardware, evaluating QML implementations against classical methods. |
| Balewski et al. (2024) [24] | Quantum-parallel Vectorized Encodings | Not specified | Explored quantum-parallel vectorized data encodings and computations, demonstrating efficiency on quantum processors. |
| Munikote et al. (2024) [18] | Various Quantum Encoding Techniques | Not specified | Compared different quantum encoding techniques, highlighting their effectiveness in various contexts. |
| Chen et al. (2024) [25] | Coupling and Splitting Data Encoding | PCA for dimensionality reduction | Developed methods for complex-valued data classification, showing improved performance with coupling amplitude encoding. |
| Ranga et al. (2024) [19] | Basis, Amplitude, Angle Encoding | PCA for dimensionality reduction | Analyzed various data-encoding techniques in Quantum Machine Learning (QML) and identified challenges such as scalability and noise. |
| Sharma et al. (2024) [26] | Hybrid Encoding Methods | Not specified | Suggested hybrid encoding methods integrating amplitude and angle encoding to emulate Euclidean distance. |
| Wang et al. (2024) [27] | Quantum Kernel PCA | Not specified | Proposed a quantum algorithm for kernel PCA, achieving near-optimal performance in dimensionality reduction. |
| Wang et al. (2024) [28] | Quantum Kernel PCA | Not specified | Introduced a resource-efficient quantum PCA method for high-dimensional data analysis, enhancing efficiency in quantum information processing. |
| Munikote et al. (2024) [18] | Various Quantum Encoding Techniques | PCA for dimensionality reduction | Compared encoding methods using QuClassi architecture, focusing on accuracy and computational complexity. |
| Flöther et al. (2025) [10] | Various Quantum Techniques | Not specified | Discussed how quantum computing can enhance biomarker discovery, emphasizing the role of encoding methods. |
| Johri et al. (2025) [29] | Bit-bit Encoding | Not specified | Focused on scalable quantum machine learning techniques, including bit-bit encoding and optimizer-free training. |
| Rath et al. (2025) [30] | Continuous-Variable Encoding | PCA for performance assessment | Explored CVQC data encoding techniques and their impact on classical ML models, showing improved classification accuracy. |
| Wang et al. (2025) [31] | Amplitude Encoding | PCA for feature enrichment | Investigated limitations of amplitude encoding in quantum classification, revealing concentration phenomena that hinder performance. |
| Jueco et al. (2025) [32] | Quantum Transforms | PCA for image reconstruction | Evaluated QFT, QWT, and QPCA in reconstructing images under noise, highlighting QWT's superior performance. |

us to analyze how the reduction in dimensionality affects classification performance in different embedding-algorithm configurations.

### *A. QUANTUM EMBEDDING TECHNIQUES*

Input data encoding in the QML model refers to the process of mapping classical data into a quantum state within a quantum circuit. This initial step is crucial as it determines how the classical information is represented and processed by the quantum operations. The efficiency of a QML model critically depends on how classical data are encoded in quantum-computable form [33]. Different encoding methods have varying resource requirements in terms of qubit numbers and gate depth, impacting the feasibility of implementation on near-term quantum hardware. Below, we summarize the three encoding techniques studied in this paper.

#### 1) ANGLE EMBEDDING

Angle embedding is a widely used quantum feature encoding technique in which classical data are mapped into the rotation angles of single-qubit gates, typically chosen from the set $\{R_X, R_Y, R_Z\}$. Given a real-valued feature vector $x = (x_1, x_2, \ldots, x_d) \in \mathbb{R}^d$,, the embedding is implemented by applying a tensor product of single-qubit rotations, as follows:

$$U(x) = \bigotimes_{j=1}^{d} R_{\alpha_j}(x_j) \tag{1}$$

where $R_{\alpha_j}(\cdot)$ is a rotation around the axis $\alpha_j \in \{X, Y, Z\}$ of the Bloch sphere, applied to the $j$-th qubit.

In our study, we adopt $R_Y$ **-based Angle embedding**, where each feature value $x_j$ is mapped directly to a rotation about the $Y$-axis:

$$U(x) = \bigotimes_{j=1}^{d} R_Y(x_j) \tag{2}$$

The single-qubit rotation about the $Y$-axis is defined as:

$$R_Y(x_j) = \begin{bmatrix} \cos\left(\frac{x_j}{2}\right) & -\sin\left(\frac{x_j}{2}\right) \\ \sin\left(\frac{x_j}{2}\right) & \cos\left(\frac{x_j}{2}\right) \end{bmatrix} \tag{3}$$

This encoding ensures that each input feature directly influences the quantum state through trigonometric functions of the data, while keeping the circuit shallow and hardware-efficient. For high-dimensional datasets, Angle embedding typically requires either many qubits (one per feature) or a prior dimension reduction step. Each feature is mapped using a single rotation gate, and the circuit depth grows linearly with the number of features, making it well-suited for NISQ devices. Despite this, empirical studies have shown that angle embeddings combined with expressive entangling layers can provide robust classification performance [34], [35].

#### 2) AMPLITUDE EMBEDDING

Amplitude embedding is a quantum encoding technique where a classical feature vector is directly represented in the amplitudes of a quantum state. Given a normalized feature vector $\mathbf{x} = (x_0, x_1, \ldots, x_{N-1}) \in \mathbb{R}^N$ and $\|\mathbf{x}\|_2 = 1$, the corresponding quantum state is prepared as:

$$U_\phi(x) = \sum_{i=0}^{N-1} x_i |i\rangle \tag{4}$$

where $|i\rangle$ denotes the computational basis states of $\log_2(N)$ qubits.

This means that an $N$-dimensional classical vector can be encoded into only $\log_2(N)$ qubits, leading to exponential compression in terms of qubit requirements. For example, a feature vector of dimension $N = 2^k$ can be represented using only $k$ qubits.

However, the main challenge lies in state preparation. Constructing a unitary circuit that transforms the initial state $|0\rangle^{\otimes \log_2(N)}$ into $|\mathbf{x}\rangle$ generally requires a deep quantum circuit with many controlled rotations and entanglement gates [36]. Such complexity makes the method highly resource-intensive and particularly vulnerable to noise in near-term Noisy Intermediate-Scale Quantum (NISQ) devices [37].

#### 3) IQP (INSTANTANEOUS QUANTUM POLYNOMIAL TIME) EMBEDDING

The IQP embedding is a quantum feature map introduced by [38] for supervised learning in quantum-enhanced feature spaces. This embedding leverages the structure of IQP embedding circuits to transform classical data into high-dimensional quantum states that can be exploited in kernel-based machine learning. The process begins by initializing all qubits in the computational ground state $|0\rangle^{\otimes n}$. A layer of Hadamard gates is then applied to each qubit, producing a uniform superposition of basis states. The core of the embedding is a data-dependent diagonal unitary operator $U_\phi(x)$, where $x = (x_1, x_2, \ldots, x_d)$ represents the classical input. This operator encodes the features through single and two-qubit phase rotations, typically of the form:

$$U_\phi(x) = \exp\Bigg(i \sum_{j=1}^{n} \phi_j(x) Z_j + i \sum_{1 \le j < k \le n} \phi_{j,k}(x) Z_j Z_k\Bigg) \tag{5}$$

where $Z_j$ denotes the Pauli-$Z$ operator on qubit $j$, and the functions $\phi_j(x)$, $\phi_{j,k}(x)$ define how data components are mapped to quantum phases.

To increase expressivity, the diagonal unitary is switched between layers of Hadamard gates. The feature state after two repetitions can be written as:

$$|\Phi(x)\rangle = U_\phi(x) H^{\otimes n} U_\phi(x) H^{\otimes n} |0\rangle^{\otimes n} \tag{6}$$

Here, $H^{\otimes n}$ denotes Hadamard operations on all qubits, while $U_\phi(x)$ injects the classical information.

This construction effectively embeds classical data into a non-linear quantum feature space [39], where even simple linear classifiers, such as SVM, can separate data that

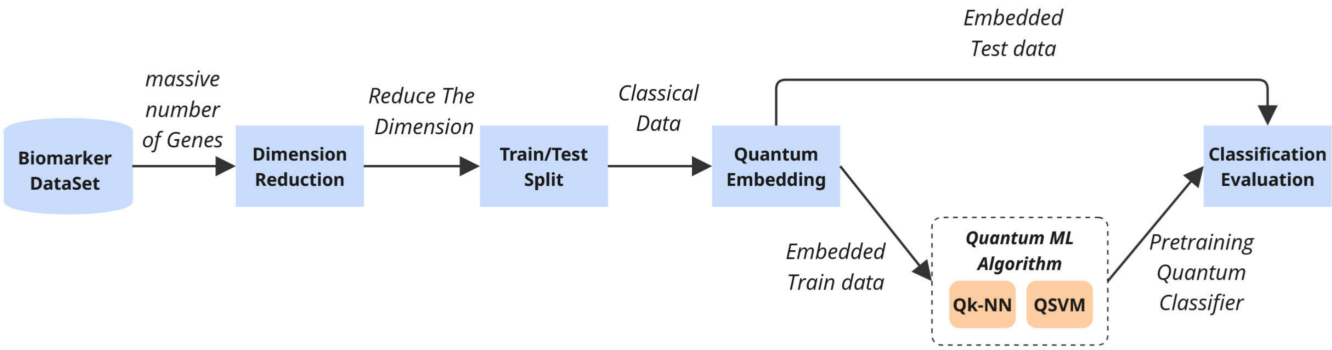


**FIGURE 1.** Pipeline diagram of our implementation.

would otherwise be inseparable in the original space. Importantly, IQP embedding are conjectured to be classically hard to simulate, owing to their connection with IQP embedding circuit complexity and the intractability of approximating partition functions of complex-valued Ising models. This provides a foundation for potential quantum advantage in machine learning, as the kernel values that underpin classification may be efficiently estimated on quantum hardware but remain inaccessible to the classical algorithm.

### B. QUANTUM SUPPORT VECTOR MACHINE

The Support Vector Machine (SVM) is a widely used classical algorithm for both classification and regression. Its quantum counterpart, the Quantum Support Vector Machine (QSVM), leverages the principles of quantum computing to provide potential benefits, particularly manipulation of high-dimensional feature spaces and efficiently evaluating complex kernel functions [40]. The key advancement of QSVM is its integration of quantum feature maps and the quantum kernel trick. The first step in QSVM is to map classical data points into a quantum state within a high-dimensional Hilbert space. For a classical data vector $\mathbf{x} \in \mathbb{R}^n$, this is achieved by a parameterized quantum circuit (PQC), often called a feature map circuit $U_\phi(\mathbf{x})$, which acts on an initial null state:

$$|\phi(\mathbf{x})\rangle = U_\phi(\mathbf{x})|0\rangle^{\otimes n} \tag{7}$$

This step is crucial for constructing quantum kernels in our implementation. In our study, we experiment with different quantum embedding techniques to build the kernel for the QSVM; Its primary advantage is the ability to evaluate kernel functions efficiently for certain feature maps, a task that remains classically intractable [41]. The kernel function $K(x_i, x_j)$ measures the similarity between two data points in the high-dimensional feature space. In QSVM the fidelity-based quantum kernel defined in was computed directly from the simulated statevectors using the simulator backend. This provides an exact evaluation of the kernel entries without requiring a Swap Test circuit. This is defined as the fidelity between their corresponding quantum states:

$$\begin{aligned} K(x_i, x_j) &= \left|\langle \phi(x_i) \mid \phi(x_j)\rangle\right|^2 \\ &= \left|\langle 0^{\otimes n} \mid U^\dagger_{\phi(x_i)} U_{\phi(x_j)} \mid 0^{\otimes n}\rangle\right|^2 \end{aligned} \tag{8}$$

The kernel is computed on a quantum processor by first preparing the quantum state $U_\phi(x_j)|0^{\otimes n}\rangle$, and then applying the inverse circuit $U_\phi(x_i)^\dagger$. This procedure constructs a quantum circuit that evaluates the kernel for a pair of inputs $(x_i, x_j)$. The probability of measuring the all-zero state directly yields the kernel value. This compute–uncompute strategy is highly resource-efficient, as it avoids the ancillary qubits usually required by alternative methods like the SWAP test [42]. Once the quantum kernel is defined, the Quantum Support Vector Machine (QSVM) follows the same dual optimization framework as its classical counterpart, with the distinction that kernel evaluations are performed on a quantum device.

Consider a training dataset $(x_i, y_i)_{i=1}^N$ with labels $y_i \in -1, +1$. The dual optimization problem is formulated as:

$$\sum_{i=1}^{n} \alpha_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \alpha_i\alpha_j y_i y_j K(x_i, x_j) \tag{9}$$

subject to the constraints $\sum_{i=1}^N \alpha_i y_i = 0, \ 0 \leq \alpha_i \leq C \quad \forall i$, where $C$ is a regularization hyperparameter controlling the trade-off between maximizing the margin and minimizing classification error.

The decision function for classifying a new data point $\tilde{x}$ is then given by:

$$f(\tilde{x}) = \text{sgn}!\left(\sum_{i=1}^{N} \alpha_i y_i K(x_i, \tilde{x}) + b\right) \tag{10}$$

where the bias term $b$ is determined using the support vectors, i.e., the training samples for which $\alpha_i > 0$. The role of the kernel $K(x_i, \tilde{x})$ is to implicitly map the input data into a higher-dimensional feature space, enabling the classifier to find a separating hyperplane even for nonlinearly separable data. Quantum Support Vector Machines (QSVMs) represent a promising near-term application, leveraging quantum feature maps to construct classically intractable kernels and enhance pattern recognition. Although limited by noise and scalability in the NISQ era, advances in error mitigation, parameter tuning, and data encoding continue to improve their practicality and performance [43].

### C. QUANTUM K-NEAREST NEIGHBORS

The k-NN algorithm is a non-parametric supervised learning method that classifies a sample based on the majority label of its nearest neighbors. Still, distance calculations in high-dimensional spaces are computationally costly, motivating quantum-assisted variants. In this study, we implement a Hybrid Quantum–Classical k-Nearest Neighbors (QkNN) algorithm that preserves the structure of classical k-NN while leveraging quantum mechanical properties to evaluate similarities in Hilbert space. The key idea of this approach is that classical feature vectors are encoded into quantum states, and their fidelity is computed via the SWAP test, which serves as the distance metric for neighbor selection, as illustrated in **Figure** 2. Unlike QkNN variants based on Hamming distance [44], which are limited to binary-valued features and discrete comparisons, the SWAP-test formulation offers a more general and flexible similarity measure applicable to continuous-valued feature vectors. The swap test adds an ancillary qubit and controlled-SWAP operations, increasing circuit depth and entangling gate count compared to QSVM. However, it avoids explicit training and instead classifies by evaluating fidelities between test and training states.

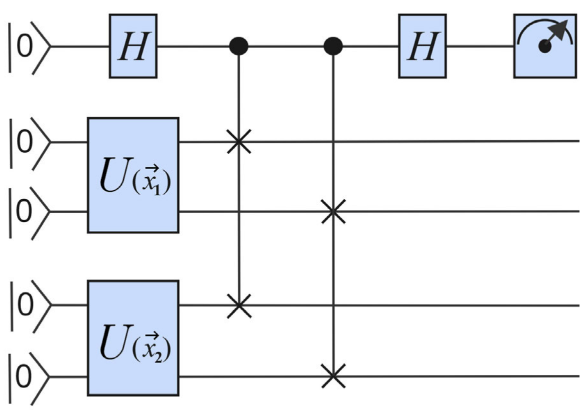


**FIGURE 2.** The QKNN circuit.

Formally, we define the labeled dataset as follows: $D = \{(x_i, y_i)\}_{i=1}^{M}$, where each $x_i \in \mathbb{R}^n$ is a classical feature vector and $y_i \in \mathcal{C}$ is its associated class label. For our case, we divide the dataset into training and test subsets, denoted as $x_{\text{train}}$ and $x_{\text{test}}$, respectively. Given an unlabeled test instance $x_{\text{test}}$, the quantum k-Nearest Neighbors (QkNN) classifier proceeds in three stages: encoded into a quantum state $|\psi(x_i)\rangle$

1) **Distance Computation**: For the test instance $x_{\text{test}}$, we first encoded it along with all training samples in $x_{\text{train}}$, then embedded those points into quantum states. This step requires encoding both the test state and each training state into two separate quantum registers, followed by a SWAP test using an ancillary qubit. Each measurement outcome provides a probability distribution from which the fidelity and corresponding distance are estimated. Importantly, unlike classical kNN, which must compute distances sequentially, quantum parallelism potentially enables simultaneous evaluation of multiple inner products, thereby reducing asymptotic runtime [45]. To quantify the similarity between two quantum states $|\psi(x_1)\rangle$ and $|\psi(x_2)\rangle$, the SWAP test was executed. An ancillary qubit initialized in $|0\rangle$ was first placed into superposition via a Hadamard gate. Subsequently, controlled-SWAP (CSWAP) operations were performed between the corresponding qubits of the two registers, conditioned on the ancillary qubit. A final Hadamard operation followed by measurement of the ancillary qubit yielded the probability distribution:
$$P(0) = \tfrac{1}{2}\left(1 + |\langle\psi(x_1) \mid \psi(x_2)\rangle|^2\right) \tag{11}$$
From this, the fidelity was extracted as
$$F(x_1, x_2) = 2P(0) - 1 \tag{12}$$
and the quantum distance metric is defined as
$$d(x_1, x_2) = 1 - F(x_1, x_2) \tag{13}$$
This fidelity-based distance captures the geometric closeness of states in Hilbert space.
2) **Neighbor Selection**: Once all distances $\{d(x_{\text{test}}, x_i)\}_{i=1}^{M}$ are obtained, the next step is to determine the indices of the $k$ training samples that are closest to $x_{\text{test}}$. Formally,
$$N_k = \arg \min_{S\subseteq\{1,\ldots,M\},\ |S|=k} \sum_{i\in S} d(x_{\text{test}}, x_i) \tag{14}$$
This selection yields the subset $N_k$ corresponding to the $k$ smallest distances. In practice, this is achieved by sorting the set of computed distances and retaining the $k$ nearest indices. While this step is executed classically in our implementation, quantum-enhanced minimum-search subroutines such as Grover's algorithm have been proposed to further accelerate neighbor selection.
3) **Majority Voting**: Once the $k$ nearest neighbors of the test sample $x_{\text{test}}$ have been identified, the classification decision is made by applying a majority voting scheme. The key idea is that the class most frequently occurring among the neighbors determines the label of the test sample.

## IV. EXPERIMENT AND RESULTS

### A. DATASET DESCRIPTION

The dataset used in this study originates from the CuMiDa (Curated Microarray Database) [46], a carefully compiled repository of cancer-related microarray experiments. We specifically focused on the Breast-cancer subset, which consists of 289 tissue samples, including 146 normal breast tissues and 143 breast adenocarcinoma, originally obtained from the Gene Expression Omnibus (GEO). To ensure consistency and biological validity, the data underwent a rigorous curation and preprocessing pipeline. Each sample represents a comprehensive gene expression profile, capturing transcriptomic activity through thousands of probes distributed across the genome. To maintain data integrity, several preprocessing steps were applied, including background correction and normalization to reduce technical variability, systematic removal of low-quality samples during quality control, and manual curation to eliminate probes lacking valid nucleic acid sequences. As a result, the final dataset provides a biologically meaningful and well-balanced representation of breast tissue, providing a robust foundation for biomarker classification using machine learning and quantum machine learning approaches.

### B. PERFORMANCE EVALUATION

In this section, we investigate how different quantum embeddings influence the performance of quantum machine learning algorithms on the Biomarker Classification task. Each model is evaluated in various configurations, including varying the number of features through PCA. The **Table** 2 shows the comparative classification performance of QSVM and classical SVM across two PCA dimensions (5 and 10 components) and different embedding strategies, Angle, Amplitude, IQP, and the classical feature map for SVM. These results illustrate the impact of embedding choice on accuracy, precision, recall, and other evaluation metrics. Similarly, **Table** 3 reports the performance of QKNN and classical KNN under the same experimental setup, showing how changes in the number of principal components and embedding methods affect classification outcomes on the Biomarker cancer dataset.

For QSVM, the performance of different embeddings varied significantly with the number of PCA components. With five components, Angle embedding performed best, achieving an accuracy of 0.873, Cohen's kappa 0.745, and MCC 0.747, slightly outperforming classical SVM (accuracy 0.862, Cohen's kappa 0.722 and MCC 0.733). The strong performance of Angle Embedding in this setting can be attributed to its efficient representation of low-dimensional feature spaces, where the limited number of parameters is sufficient to capture the discriminative structure of the data. In contrast, IQP embedding was less effective with five components (accuracy 0.851, kappa 0.699 and MCC 0.701), as its expressiveness is underutilized in such a small feature space. When increasing to 10 components, IQP embedding became the strongest performer, reaching an accuracy of 0.931, F1-score 0.931, MCC 0.862, and Cohen's kappa 0.862, outperforming Angle Embedding (accuracy 0.908) and classical SVM (accuracy 0.897). This indicates that IQP embedding is better suited for higher-dimensional spaces, where its richer parameterization can capture more complex feature interactions. In contrast, Amplitude Embedding consistently underperformed, stagnating at an accuracy of 0.529 with poor precision (0.264) and F1-score (0.345). A key limitation of this embedding is that it encodes data into a Hilbert space of size $2^n$, requiring padding when the number of features does not match a power of two. This padding introduces artificial components that act as noise, reducing the quality of the quantum state and, consequently, the classification performance. Overall, these results suggest a clear trade-off: Angle Embedding is advantageous in low-dimensional settings, IQP embedding excels as feature dimensionality increases, and Amplitude Embedding proves unsuitable for this dataset due to the padding-induced noise. Notably, in the 10-PC setting, QSVM with IQP embedding not only outperforms its quantum counterparts but also surpasses classical SVM.

For QKNN, the three quantum embeddings performed better overall compared to their counterparts in QSVM, although the trend in performance was reversed. With five components, the IQP embedding achieved the strongest results, reaching an accuracy of 0.839, precision 0.871, recall 0.831, Cohen's kappa 0.671, F1-score 0.832, and MCC of 0.701. These values were slightly higher than those of Angle embedding (accuracy 0.816, kappa 0.623) and Amplitude embedding (accuracy 0.804, kappa 0.599). Remarkably, IQP embedding even matched the performance of classical KNN (accuracy 0.839, kappa 0.674) and slightly exceeded it in precision. When the feature space was expanded to 10 components, the performance ranking shifted. Angle embedding emerged as the best configuration, with an accuracy of 0.862, precision 0.896, recall 0.853, Cohen's kappa 0.718, F1-score 0.856, and MCC of 0.718. IQP and Amplitude embeddings showed similar but lower performance levels in this setting. At the same time, classical KNN outperformed all quantum variants, obtaining an accuracy of 0.885, precision 0.901, recall 0.879, Cohen's kappa 0.766, F1-score 0.882, and MCC of 0.781. The opposite trend compared to QSVM arises from the different similarity mechanisms used. QSVM relies on a quantum kernel based on inner products, where IQP embedding becomes more expressive in larger feature spaces. In contrast, QKNN uses the swap test to measure state similarity, which directly compares quantum states without relying on kernel functions. This makes QKNN less sensitive to the dimensionality-driven effects observed in QSVM, allowing IQP embedding to perform better in lower dimensions and Angle to take the lead when more features are included. Another important distinction is the behavior of Amplitude Embedding. Unlike in QSVM, where padding to match $2^n$ features introduces noise into the kernel computation, the swap-test-based similarity in QKNN is less affected by such padding. This explains why Amplitude Embedding achieves relatively competitive results in QKNN compared to its poor performance in QSVM. Finally, the superior performance of classical KNN in higher dimensions highlights the strength of Euclidean distance for structured biomedical datasets. While quantum fidelity offers an alternative notion of similarity, it appears less effective

**TABLE 2.** Classification performance results with different embeddings and principal components dimensions.

| **Metrics** | **5 PCs** | | | | **10 PCs** | | | |
|---|---|---|---|---|---|---|---|---|
| Algorithm | **QSVM** | | | **SVM** | **QSVM** | | | **SVM** |
| Embedding | Angle | Amplitude | IQP | Classical | Angle | Amplitude | IQP | Classical |
| Accuracy | 0.873 | 0.528 | 0.851 | 0.862 | 0.908 | 0.528 | 0.931 | 0.897 |
| Precision | 0.875 | 0.264 | 0.852 | 0.877 | 0.912 | 0.264 | 0.931 | 0.903 |
| Recall | 0.871 | 0.500 | 0.848 | 0.856 | 0.905 | 0.500 | 0.931 | 0.892 |
| Cohen's Kappa | 0.745 | 0.0 | 0.698 | 0.722 | 0.814 | 0.0 | 0.861 | 0.791 |
| F1-Score | 0.872 | 0.345 | 0.849 | 0.858 | 0.907 | 0.345 | 0.931 | 0.895 |
| MCC | 0.747 | 0.0 | 0.701 | 0.733 | 0.818 | 0.0 | 0.862 | 0.796 |

**TABLE 3.** Classification performance results with different embeddings and principal components dimensions.

| **Metrics** | **5 PCs** | | | | **10 PCs** | | | |
|---|---|---|---|---|---|---|---|---|
| Algorithm | **QKNN** | | | **KNN** | **QKNN** | | | **KNN** |
| Embedding | Angle | Amplitude | IQP | Classical | Angle | Amplitude | IQP | Classical |
| Accuracy | 0.816 | 0.804 | 0.839 | 0.839 | 0.862 | 0.851 | 0.851 | 0.885 |
| Precision | 0.856 | 0.849 | 0.871 | 0.846 | 0.896 | 0.889 | 0.889 | 0.901 |
| Recall | 0.806 | 0.794 | 0.831 | 0.834 | 0.853 | 0.841 | 0.841 | 0.879 |
| Cohen's Kappa | 0.623 | 0.599 | 0.671 | 0.674 | 0.718 | 0.694 | 0.694 | 0.766 |
| F1-Score | 0.806 | 0.793 | 0.832 | 0.836 | 0.856 | 0.843 | 0.843 | 0.882 |
| MCC | 0.661 | 0.641 | 0.701 | 0.681 | 0.748 | 0.729 | 0.729 | 0.781 |

in capturing separability in high-dimensional feature spaces, allowing classical KNN to maintain an edge as more PCA components are included.

In summary, the results indicate that the selected quantum embedding plays a crucial role in determining both the model's performance and scalability. The IQP embedding delivers high expressiveness but requires significant computational resources, while the Angle embedding achieves a practical balance between accuracy and hardware efficiency. Although the Amplitude embedding is theoretically more compact, its effectiveness is hindered by circuit complexity and instability. These findings emphasize that carefully optimizing the embedding method is essential for attaining reliable biomarker classification on current quantum hardware.

## V. DISCUSSION

The obtained results demonstrate that quantum encodings significantly influence model performance and highlight several significant findings:

### A. IMPACT OF ENCODING TECHNIQUE

The choice of the appropriate encoding technique has a decisive effect on quantum model performance. IQP embedding consistently provided the highest accuracy and kappa values, especially when combined with QSVM, confirming its ability to capture non-linear relationships in biomarker data. In contrast, Amplitude Embedding performed poorly across all settings, likely due to its higher circuit complexity and sensitivity to dimensionality. Angle Embedding offered stable but slightly weaker results compared to IQP embedding.

### B. QUANTUM VERSUS CLASSICAL MODELS

- QSVM with IQP embedding outperformed both classical SVM and KNN in terms of accuracy (93.10% vs. 89.65% for classical SVM) and Cohen's kappa (0.8616 vs. 0.7910). This suggests that quantum kernels can extract richer feature representations from biomarker datasets.
- QKNN, while competitive, did not surpass classical KNN, reflecting its greater sensitivity to encoding and circuit depth.

### C. ROLE OF DIMENSIONALITY REDUCTION

Increasing PCA components from 5 to 10 significantly improved performance for both quantum and classical models. This indicates that quantum embeddings benefit from richer input features, provided that the dimensionality is controlled to remain compatible with the number of available qubits.

### D. EVALUATION BEYOND ACCURACY

Table 2 presents the comparative classification performance of QSVM and classical SVM in different PCA dimensions (5 and 10 components) and embedding strategies. The results highlight the effect of embedding choice on the model's accuracy, precision, recall, and other evaluation metrics.

Metrics such as Cohen's kappa and MCC provided more profound insight into the models' robustness, particularly in handling class imbalance. QSVM-IQP's kappa score of 0.8616 demonstrates substantial agreement beyond chance, validating the strength of this approach in biomarker classification.

### E. LIMITATIONS

Despite these encouraging results, several limitations must be acknowledged:

- All experiments were conducted on simulators via PennyLane, meaning that hardware noise and decoherence were not considered.
- Scalability to larger biomarker datasets remains constrained by qubit availability and circuit complexity.
- The performance gap between QSVM and QKNN indicates that not all classical algorithms benefit equally from quantum feature space mappings.

### F. POSITIONING WITHIN EXISTING WORK

While previous studies have applied individual quantum machine learning models to biomedical and biomarker datasets, direct comparisons between different quantum embedding strategies have not been extensively explored. The present work addresses this gap by systematically analyzing and contrasting Angle, Amplitude, and IQP embeddings under consistent experimental conditions. This comparative perspective highlights how encoding choices affect classification outcomes, thereby extending the existing literature beyond isolated demonstrations of specific algorithms. By positioning our results in this way, the contribution of the paper lies in providing the first comparative encoding-level study for biomarker classification, which complements and contextualizes prior quantum machine learning applications in biomedical domains.

### G. FUTURE PERSPECTIVES

Future work should focus on:

- Testing on real quantum hardware to evaluate resilience under noise.
- Exploring hybrid quantum-classical architectures with PennyLane for more scalable solutions.
- Investigating novel embeddings, such as Neural Quantum Embedding (NQE), to further enhance feature expressiveness.

## VI. CONCLUSION

This study presented a comparative analysis of three quantum data encoding techniques, Angle embedding, Amplitude embedding, and IQP embedding, applied to breast cancer biomarker classification using the CuMiDa dataset. Principal Component Analysis (PCA) was employed for dimensionality reduction, and the quantum classifiers (QSVM and QkNN) were systematically benchmarked against their classical counterparts (SVM and KNN). The results demonstrate that IQP Embedding combined with QSVM achieved the best performance, reaching 93.1% accuracy and an F1-score of 0.93 in higher-dimensional settings, followed by Angle Embedding, both surpassing the classical SVM baseline. For

QkNN, performance depended on the dimensionality of the features: IQP Embedding achieved the strongest results in lower dimensions, including a precision of 87.1%, while Angle Embedding outperformed other quantum approaches in higher dimensions. Amplitude Embedding, although less effective in QSVM, showed competitive behavior in QkNN. These findings underscore the novel contribution of this work as the first systematic comparison of multiple quantum embeddings for biomarker classification. They also highlight the importance of the choice of data encoding and the reduction in dimensionality in optimizing quantum machine learning performance. Furthermore, our analysis of computational resource requirements, covering qubit count, circuit depth, and gate complexity, demonstrates the practical feasibility of these approaches on near-term NISQ devices. We acknowledge that this evaluation focused on classification-based metrics rather than direct embedding-level analyses (e.g., visualization or entanglement measures), which represents an important direction for future research. Additionally, experiments were conducted on quantum simulators, and while this enabled exact kernel computation, it did not capture real hardware noise. The dataset size and scope were limited by simulation constraints, and the study focused on a single biomedical dataset and two classifier types. Future work will extend this framework to larger and more diverse biomedical datasets, integrate additional encoding schemes and hybrid architectures, and implement the models on real quantum hardware to assess robustness under realistic conditions. Moreover, coupling quantum feature maps with explainable AI methods presents a promising avenue toward interpretable and clinically relevant quantum-enhanced decision support systems.

## ACKNOWLEDGMENT

The authors acknowledge the use of AI-assisted tools to support and improve the language and clarity of this manuscript. while all conceptual contributions, results, and conclusions are solely those of them.

## REFERENCES

[1] Y. Hasin, M. Seldin, and A. Lusis, "Multi-omics approaches to disease," *Genome Biol.*, vol. 18, no. 1, p. 83, May 2017, doi: 10.1186/s13059-017-1215-1.

[2] L. Li, W.-K. Ching, and Z.-P. Liu, "Robust biomarker screening from gene expression data by stable machine learning-recursive feature elimination methods," *Comput. Biol. Chem.*, vol. 100, Oct. 2022, Art. no. 107747. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S147692712200127X

[3] M. W. Libbrecht and W. S. Noble, "Machine learning applications in genetics and genomics," *Nature Rev. Genet.*, vol. 16, no. 6, pp. 321–332, Jun. 2015, doi: 10.1038/nrg3920.

[4] T. Verma, B. K. Kumar, J. Rajendar, and B. K. N. Swamy, "A review on quantum machine learning," in *Proc. 6th Int. Conf. Commun. Cyber Phys. Eng.*, 2024, pp. 407–415, doi: 10.1007/978-981-99-7137-4_39.

[5] A. Abbas, D. Sutter, A. Figalli, and S. Woerner, "Effective dimension of machine learning models," 2021, *arXiv:2112.04807*.

[6] S. Jerbi, L. J. Fiderer, H. P. Nautrup, J. M. Kübler, H. J. Briegel, and V. Dunjko, "Quantum machine learning beyond kernel methods," *Nature Commun.*, vol. 14, no. 1, p. 517, Jan. 2023, doi: 10.1038/s41467-023-36159-y.

[7] X. Wang, Y. Du, Y. Luo, and D. Tao, "Towards understanding the power of quantum kernels in the NISQ era," *Quantum*, vol. 5, p. 531, Aug. 2021, doi: 10.22331/q-2021-08-30-531.

[8] B. Khanal, P. Rivas, A. Sanjel, K. Sooksatra, E. Quevedo, and A. Rodriguez, "Generalization error bound for quantum machine learning in NISQ era—A survey," *Quantum Mach. Intell.*, vol. 6, no. 2, p. 90, Dec. 2024, doi: 10.1007/s42484-024-00204-w.

[9] J. Yang, W. Xie, and X. Xu, "Stability and generalization of quantum neural networks," 2025, *arXiv:2501.12737*.

[10] F. F. Flöther, D. Blankenberg, M. Demidik, K. Jansen, R. Krishnakumar, R. Krishnakumar, N. Laanait, L. Parida, C. Y. Saab, and F. Utro, "How quantum computing can enhance biomarker discovery," *Patterns*, vol. 6, no. 6, Jun. 2025, Art. no. 101236. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2666389925000844

[11] P. Lamichhane and D. B. Rawat, "Quantum machine learning: Recent advances, challenges, and perspectives," *IEEE Access*, vol. 13, pp. 94057–94105, 2025.

[12] Y. Li, Z. Wang, R. Han, S. Shi, J. Li, R. Shang, H. Zheng, G. Zhong, and Y. Gu, "Quantum recurrent neural networks for sequential learning," *Neural Netw.*, vol. 166, pp. 148–161, Sep. 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S089360802300360X

[13] P.-N. Nguyen, "Biomarker discovery with quantum neural networks: A case-study in CTLA4-activation pathways," *BMC Bioinf.*, vol. 25, no. 1, p. 149, Apr. 2024, doi: 10.1186/s12859-024-05755-0.

[14] M. K. Saggi, A. S. Bhatia, I. K. Mensah, H. Gowher, and S. Kais, "Multi-omic and quantum machine learning integration for lung subtypes classification," *Future Gener. Comput. Syst.*, vol. 174, Jan. 2026, Art. no. 107905. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0167739X25002006

[15] M. Z. Ghobadi and E. Afsaneh, "Potential of quantum machine learning for solving the real-world problem of cancer classification," *Discover Appl. Sci.*, vol. 6, no. 10, p. 513, Sep. 2024, doi: 10.1007/s42452-024-06220-6.

[16] P.-N. Nguyen, "QuanAnts machine: A quantum algorithm for biomarker discovery," 2023, *arXiv:2309.00001*.

[17] M. Rath and H. Date, "Quantum data encoding: A comparative analysis of classical-to-quantum mapping techniques and their impact on machine learning accuracy," *EPJ Quantum Technol.*, vol. 11, no. 1, p. 72, Dec. 2024.

[18] N. Munikote, "Comparing quantum encoding techniques," 2024, *arXiv:2410.09121*.

[19] D. Ranga, A. Rana, S. Prajapat, P. Kumar, K. Kumar, and A. V. Vasilakos, "Quantum machine learning: Exploring the role of data encoding techniques, challenges, and future directions," *Mathematics*, vol. 12, no. 21, p. 3318, Oct. 2024. [Online]. Available: https://www.mdpi.com/2227-7390/12/21/3318

[20] R. LaRose and B. Coyle, "Robust data encodings for quantum classifiers," *Phys. Rev. A, Gen. Phys.*, vol. 102, no. 3, Sep. 2020, Art. no. 032420. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevA.102.032420

[21] N. Thumwanit, C. Lortaraprasert, and R. Raymond, "Invited: Trainable discrete feature embeddings for quantum machine learning," in *Proc. 58th ACM/IEEE Design Autom. Conf. (DAC)*, Dec. 2021, pp. 1352–1355. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/9586190

[22] R. Fauzi, M. Zarlis, H. Mawengkang, and P. Sihombing, "Analysis of several quantum encoding methods implemented on a quantum circuit architecture to improve classification accuracy," in *Proc. 6th Int. Conf. Electr., Telecommun. Comput. Eng. (ELTICOM)*, Nov. 2022, pp. 152–154, doi: 10.1109/ELTICOM57747.2022.10038017.

[23] Y. Gujju, A. Matsuo, and R. Raymond, "Quantum machine learning on near-term quantum devices: Current state of supervised and unsupervised techniques for real-world applications," *Phys. Rev. Appl.*, vol. 21, no. 6, Jun. 2024, Art. no. 067001. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevApplied.21.067001

[24] J. Balewski, M. G. Amankwah, R. Van Beeumen, E. W. Bethel, T. Perciano, and D. Camps, "Quantum-parallel vectorized data encodings and computations on trapped-ion and transmon QPUs," *Sci. Rep.*, vol. 14, no. 1, p. 3435, Feb. 2024, doi: 10.1038/s41598-024-53720-x.

[25] J. Chen and Y. Li, "Empowering complex-valued data classification with the variational quantum classifier," *Frontiers Quantum Sci. Technol.*, vol. 3, Feb. 2024, Art. no. 1282730.

[26] S. Sharma and N. Krishnaveni, "Survey of encoding techniques for quantum machine learning," *Cybern. Phys.*, vol. 13, no. 2, pp. 152–160, 2024.

[27] Y. Wang, "Near-optimal quantum kernel principal component analysis," *Quantum Sci. Technol.*, vol. 10, no. 1, Nov. 2024, Art. no. 015034, doi: 10.1088/2058-9565/ad9176.

[28] Y. Wang and Y. Luo, "Resource-efficient quantum principal component analysis," *Quantum Sci. Technol.*, vol. 9, no. 3, May 2024, Art. no. 035031, doi: 10.1088/2058-9565/ad466c.

[29] S. Johri, "Bit-bit encoding, optimizer-free training and sub-net initialization: Techniques for scalable quantum machine learning," 2025, *arXiv:2501.02148*.

[30] M. Rath and H. Date, "Continuous-variable quantum encoding techniques: A comparative study of embedding techniques and their impact on machine learning performance," 2025, *arXiv:2504.06497*.

[31] X. Wang, Y. Wang, B. Qi, and R. Wu, "Limitations of amplitude encoding on quantum classification," 2025, *arXiv:2503.01545*.

[32] J. C. Jueco and M. G. B. Palconit, "Assessing the fidelity and noise resilience of quantum Fourier, wavelet, and PCA transforms in amplitude-encoded image reconstruction," in *Proc. 4th Int. Conf. Comput. Model.*, 2025, pp. 413–422.

[33] M. Schuld, R. Sweke, and J. J. Meyer, "Effect of data encoding on the expressive power of variational quantum-machine-learning models," *Phys. Rev. A, Gen. Phys.*, vol. 103, no. 3, Mar. 2021, Art. no. 032430.

[34] M. Schuld, A. Bocharov, K. M. Svore, and N. Wiebe, "Circuit-centric quantum classifiers," *Phys. Rev. A, Gen. Phys.*, vol. 101, no. 3, Mar. 2020, Art. no. 032308.

[35] A. Pérez-Salinas, A. Cervera-Lierta, E. Gil-Fuster, and J. I. Latorre, "Data re-uploading for a universal quantum classifier," *Quantum*, vol. 4, p. 226, Feb. 2020.

[36] C.-T. Li and H.-C. Cheng, "Adaptive circuit learning of born machine: Towards realization of amplitude embedding and quantum data loading," *Quantum Sci. Technol.*, vol. 10, no. 2, Feb. 2025, Art. no. 025019, doi: 10.1088/2058-9565/adaede.

[37] K. Nakaji, S. Uno, Y. Suzuki, R. Raymond, T. Onodera, T. Tanaka, H. Tezuka, N. Mitsuda, and N. Yamamoto, "Approximate amplitude encoding in shallow parameterized quantum circuits and its application to financial market indicators," *Phys. Rev. Res.*, vol. 4, no. 2, May 2022, Art. no. 023136. [Online]. Available: https://link.aps.org/doi/10.1103/PhysRevResearch.4.023136

[38] V. Havlíček, A. D. Córcoles, K. Temme, A. W. Harrow, A. Kandala, J. M. Chow, and J. M. Gambetta, "Supervised learning with quantum-enhanced feature spaces," *Nature*, vol. 567, no. 7747, pp. 209–212, Mar. 2019, doi: 10.1038/s41586-019-0980-2.

[39] D. Hangleiter and J. Eisert, "Computational advantage of quantum random sampling," *Rev. Modern Phys.*, vol. 95, no. 3, Jul. 2023, Art. no. 035001. [Online]. Available: https://link.aps.org/doi/10.1103/RevModPhys.95.035001

[40] M. Slysz, K. Kurowski, G. Waligóra, and J. Wóglarz, "Exploring the capabilities of quantum support vector machines for image classification on the MNIST benchmark," in *Proc. ICCS*, 2023, pp. 193–200.

[41] Q.-L. Wang, Y. Jin, X.-H. Li, Y. Li, Y.-C. Li, K.-J. Zhang, H. Liu, and L. Cheng, "An advanced quantum support vector machine for power quality disturbance detection and identification," *EPJ Quantum Technol.*, vol. 11, no. 1, p. 70, Dec. 2024.

[42] A. Miroszewski, J. Mielczarek, F. Szczepanek, G. Czelusta, B. Grabowski, B. L. Saux, and J. Nalepa, "Cloud detection in multispectral satellite images using support vector machines with quantum kernels," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Mar. 2023, pp. 796–799.

[43] J. Preskill, "Quantum computing in the NISQ era and beyond," *Quantum*, vol. 2, p. 79, Aug. 2018, doi: 10.22331/q-2018-08-06-79.

[44] J. Li, S. Lin, K. Yu, and G. Guo, "Quantum K-nearest neighbor classification algorithm based on Hamming distance," *Quantum Inf. Process.*, vol. 21, no. 1, p. 18, Dec. 2021, doi: 10.1007/s11128-021-03361-0.

[45] A. Maldonado-Romo, J. Y. Montiel-Pérez, V. Onofre, J. Maldonado-Romo, and J. H. Sossa-Azuela, "Quantum K-nearest neighbors: Utilizing QRAM and SWAP-test techniques for enhanced performance," *Mathematics*, vol. 12, no. 12, p. 1872, Jun. 2024. [Online]. Available: https://www.mdpi.com/2227-7390/12/12/1872

[46] B. C. Feltes, E. B. Chandelier, B. I. Grisci, and M. Dorn, "CuMiDa: An extensively curated microarray database for benchmarking and testing of machine learning approaches in cancer research," *J. Comput. Biol.*, vol. 26, no. 4, pp. 376–386, Apr. 2019.

**AICHA EUTAMENE** received the M.S. degree in computer science from the University Fréres Mentouri Constantine, Algeria, in 2009, and the Ph.D. degree in computer science from the University of Constantine 2-Abdelhamid Mehri, Algeria, in 2016. She is currently a Lecturer with the Faculty of Natural and Life Sciences, University Fréres Mentouri-Constantine 1, Algeria. Her research interests include natural language processing, machine learning, bioinformatics, quantum computing, and applications of artificial intelligence in computational biology.

**ISLAM DJEMMAL** received the B.S. degree in computer science from Setif-1-University Ferhat Abbas, Sétif, Algeria, and the M.Sc. degree in artificial intelligence and data engineering. He is currently pursuing the Ph.D. degree in applications of artificial intelligence. His research interests include quantum computing, quantum machine learning, and their integration with classical machine learning models and hybrid quantum–classical architectures. He is also interested in natural language processing, large language models, and multilingual text processing, with a particular focus on Arabic and low-resource languages.

**HACENE BELHADEF** received the Ph.D. degree from the Faculty of Engineering, University Fréres Mentouri-Constantine 1, in 2010, and the HDR degree, in 2013. He was a Lecturer. In 2019, he was promoted to the rank of Full Professor. He is currently a Professor with the Faculty of New Technologies of Information and Communication, University of Constantine 2-Abdelhamid Mehri. He teaches several academic modules and supervises several student projects at both the Master's and Ph.D. levels, in addition to being involved in multiple research projects. He has held several administrative positions, the most recent of which was the Head of the Digital Transformation Department at the university. His research interests include artificial intelligence, natural language processing, and quantum machine learning. He is also an active member of various international conference program committees and serves as a reviewer for several international scientific journals.

**A. M. MUTAWA** (Senior Member, IEEE) received the M.S. and Ph.D. degrees in artificial intelligence from Syracuse University, New York, in 1999. He is currently a Faculty Member with the Department of Computer Engineering, Kuwait University, where he founded the robotics, artificial intelligence, and intelligent computing laboratories. He has served in various leadership roles, including the Assistant Vice President for IT and the Acting Vice President for Academic Support Services. He is an international expert on AI ethics with UNESCO, the President of MILSET Asia, and a Board Member of Arab Robotic Association. He is the author of the peer-reviewed book *Blended E-Learning and Its Impact on Critical Thinking*. His research interests include robotics, expert systems, AI, signal processing, NLP, pattern recognition, deep learning, e-learning, and quantum computing. He was a recipient of the Sheikh Salem Al-Ali Al-Sabah Award for Excellence in AI, in 2022.