Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

NQ-SVM: Scalable Quantum Kernel Learning via Nyström-Based Support Vector Machines

Duong The Do

duongdo@asu.edu

Arizona State University

Srikanth Thudumu

Institute of Applied Artificial Intelligence and Robotics

Prem Prakash Jayaraman

Swinburne University of Technology

Duong Tung Nguyen

Arizona State University

Research Article

Keywords: Quantum support vector machines, Nystr, "om approximation, hybrid quantum-classical learning

Posted Date: March 3rd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-8962794/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

NQ-SVM: Scalable Quantum Kernel Learning via Nyström-Based Support Vector Machines

Duong The Do[1*], Srikanth Thudumu[2],
Prem Prakash Jayaraman[3], Duong Tung Nguyen[1]

[1]School of ECEE, Arizona State University, Tempe, 85281, AZ, USA.
[2]Institute of Applied Artificial Intelligence and Robotics (IAAIR), Germantown, 38139, TN, USA.
[3]School of Science, Computing, and Emerging Technologies, Swinburne University of Technology, Hawthorn, 3122, VIC, Australia.

*Corresponding author(s). E-mail(s): duongdo@asu.edu;
Contributing authors: srikanth@iaair.ai; pjayaraman@swin.edu.au; duongnt@asu.edu;

Abstract

Support vector machines (SVMs) are widely used for classification, but become computationally expensive as datasets grow and data become high-dimensional or nonlinearly separable. Quantum support vector machines (QSVMs) enhance the traditional SVMs by replacing conventional feature mappings with quantum-implemented embeddings that project data into high-dimensional Hilbert spaces, potentially improving expressive power and efficiency. However, training QSVMs on current noisy intermediate-scale quantum (NISQ) devices remains challenging, as constructing the quantum kernel matrix requires repeated circuit executions and scales quadratically with dataset size. We propose *NQ-SVM*, a hybrid quantum-classical method that incorporates the Nyström approximation to enhance the scalability of QSVMs. By selecting a small set of landmark points, the Nyström method yields a low-rank kernel approximation, reducing complexity from $\boldsymbol{O(n^2)}$ to $\boldsymbol{O(nm)}$, while preserving the quantum advantages in handling high-dimensional, non-linear data. Extensive simulations demonstrate that *NQ-SVM* achieves comparable classification accuracy to classical SVMs and full QSVMs, while substantially reducing training time and kernel evaluation costs. These results underscore the potential of *NQ-SVM* as a scalable and practical hybrid framework for near-term quantum machine learning (QML).

Keywords: Quantum support vector machines, Nyström approximation, hybrid quantum-classical learning.

1 Introduction

Support Vector Machines (SVMs) are widely recognized as one of the most effective supervised learning methods for classification, owing to their ability to construct maximum-margin hyperplanes that separate data classes in high-dimensional spaces [1–4]. By incorporating kernel functions, SVMs can also model nonlinear relationships through implicit mappings into more tractable higher-dimensional feature spaces, leading to strong accuracy and generalization performance [5–8]. Despite their potential, classical SVMs face significant challenges when applied to large-scale or high-dimensional datasets, particularly when complex kernel functions are involved [6, 9, 10]. Training SVMs requires solving a quadratic optimization problem, and the kernel matrix grows quadratically with the number of samples, resulting in substantial computational and memory demands for large datasets [10–12].

To overcome these limitations, various enhanced formulations and parallel optimization methods have been proposed to improve SVM scalability. Approaches such as distributed and parallel SVMs, GPU-accelerated least-squares SVMs, low-rank kernel approximations, decomposition techniques, and communication-efficient optimization frameworks have achieved significant reductions in training time while maintaining accuracy comparable to traditional SVMs [4, 6, 9, 13–15]. Despite these advances, the scalability of kernel-based optimization on classical hardware remains limited when dealing with exponential data growth and high-dimensional feature mappings. These persistent challenges motivate the exploration of quantum computing (QC), which leverages intrinsic parallelism and high-dimensional state representations to potentially deliver superior performance.

QC departs from sequential classical processing by utilizing the principles of superposition, entanglement, and interference to perform computations in high-dimensional Hilbert spaces [16–19]. These properties enable quantum processors to execute complex linear algebra operations, such as matrix inversion, eigenvalue decomposition, and inner-product estimation, more efficiently than their classical methods [20, 21]. Hence, QC has become an attractive approach for data-intensive applications, with the potential to accelerate machine learning tasks. By exploiting quantum parallelism and high-dimensional state representations, quantum algorithms can overcome the computational and memory limitations of classical techniques [22–25].

Quantum machine learning (QML) integrates quantum computation with traditional machine learning to enhance efficiency and model expressiveness. QML algorithms exploit the exponentially large Hilbert space of quantum systems to encode classical data into quantum states and enable computations that may be infeasible on classical hardware [26–29]. By using quantum circuits for tasks such as kernel evaluations, feature mappings, and similarity calculations, QML aims to achieve computational acceleration and enhanced generalization across various learning problems

[30, 31]. Hybrid quantum-classical frameworks further enable practical QML implementations on NISQ devices by combining quantum feature extraction with classical optimization [32–35].

The quantum support vector machine (QSVM) has emerged as one of the most widely studied approaches for classification and pattern recognition. QSVMs leverage quantum feature maps for high-dimensional data embedding and perform quantum kernel evaluations to obtain inner products between the corresponding embedded states [1, 28]. This capability allows QSVMs to exploit the expressive power of quantum states and realize complex decision boundaries that may be difficult for classical SVMs to represent. QSVMs typically adopt a hybrid architecture in which the quantum processor performs state preparation and kernel estimation, while the classical processor solves the convex optimization problem defining the SVM decision boundary [32, 36–38]. This structure aligns well with NISQ hardware constraints by reducing quantum resource demands while retaining theoretical equivalence to classical SVMs. Moreover, by exploiting quantum parallelism for inner-product computations, QSVMs can achieve polynomial or even exponential speedups in certain scenarios, particularly when data access and feature encoding can be efficiently realized on quantum hardware [39, 40].

While QSVMs demonstrate the potential for quantum advantage in kernel-based learning, their practical deployment on NISQ devices remains constrained by the high computational cost of kernel estimation [39, 41]. For a training set of n samples, constructing the full quantum kernel matrix requires estimating all pairwise inner products between encoded quantum states, leading to $O(n^2)$ quantum kernel evaluations, including state preparations and measurements [41, 42]. This quadratic scaling poses a significant challenge for large datasets or applications requiring frequent kernel updates. Sampling noise, hardware errors, and limited qubit counts in NISQ devices further exacerbate these issues, making full kernel computation time-consuming and resource-intensive [28, 39, 43].

Low-rank kernel approximation techniques, most notably the Nyström method, have recently gained attention as promising approaches for reducing the computational burden of handling the full kernel matrix [44–46]. The Nyström approach approximates the full kernel using a small subset of m landmark points ($m \ll n$), capturing the essential structure of the data with substantially fewer kernel evaluations and reducing computational complexity from $O(n^2)$ to $O(nm)$ [47–49]. Compared with random feature methods, the Nyström approximation often achieves better generalization due to its data-dependent sampling strategy, which captures more informative spectral properties of the kernel matrix [46]. Recent studies have shown that Nyström-based approximations can scale SVMs to large datasets while maintaining competitive accuracy even with substantial reductions in model size [47, 48, 50, 51]. When extended to QC, the Nyström method can be applied to quantum kernel matrices, allowing a reduction in the number of required kernel evaluations and significantly lowering quantum resource consumption.

To the best of our knowledge, the integration of scalable kernel approximation techniques with QSVMs has been overlooked in the literature. This motivates the question

of whether classical kernel approximation methods, such as the Nyström approximation, can be effectively incorporated into quantum kernel learning to improve the scalability of QSVMs while maintaining classification performance. To this end, we propose *NQ-SVM*, a hybrid quantum-classical framework that integrates the Nyström approximation into quantum kernel learning to enhance the scalability of QSVMs. Instead of computing all pairwise quantum kernel entries, *NQ-SVM* selects a small subset of representative quantum states and estimates their overlaps to approximate the full quantum kernel matrix. This low-rank approximation significantly reduces the number of quantum circuit executions required for training, enabling the model to handle larger datasets and operate more efficiently on NISQ hardware. While the Nyström method offers a potential solution to the scalability limitations of quantum kernel evaluation, it introduces additional challenges arising from landmark selection, low-rank approximation error, and numerical stability. Inaccurate landmark selection may lead to poor kernel approximations that may reduce the effectiveness of the quantum kernel. Furthermore, errors introduced by the Nyström approximation can combine with measurement noise and sampling uncertainty in quantum kernel estimation, resulting in reduced kernel accuracy and stability within the learning process. To address this challenge, we propose to study several landmark selection strategies and evaluate their impact on the approximation quality and classification performance within the proposed *NQ-SVM* framework. By choosing an appropriate number of landmark points, *NQ-SVM* can maintain high classification performance while reducing computational demands. Fewer quantum kernel evaluations also mitigate the effects of sampling noise and device errors, improving the reliability of kernel estimation in NISQ settings. This integration offers a scalable hybrid quantum–classical learning framework that preserves the expressive power of quantum feature spaces while achieving the computational efficiency of low-rank kernel methods. The main contributions of this paper are summarized as follows.

- **Algorithm Design**: We develop *NQ-SVM*, a hybrid quantum-classical framework that incorporates the Nyström approximation with quantum kernel estimation, enabling efficient feature construction without evaluating the full $n \times n$ quantum Gram matrix. By reducing the number of state preparations and kernel evaluations, our approach improves the practical feasibility of QSVMs on near-term devices and allows training on large datasets.
- **Data Preprocessing**: We employ principal component analysis (PCA) to reduce the dimensionality of the input data prior to quantum feature encoding, which helps control circuit complexity and improves the efficiency of quantum kernel evaluation.
- **Landmark Selection**: We investigate multiple landmark selection strategies, including uniform sampling, k-means clustering, and farthest-point sampling, and analyze their impact on the performance of the Nyström approximation in quantum kernel estimation.
- **Numerical Validation**: We evaluate the proposed method on both the full and reduced PneumoniaMNIST datasets, explore multiple quantum feature maps, and compare different kernel construction strategies. Our simulations show that *NQ-SVM* achieves comparable classification accuracy to the full QSVM approach while significantly reducing computational requirements. We also analyze how

landmark size, regularization parameters, and qubit count influence performance. Results demonstrate that *NQ-SVM* maintains stable accuracy even under finite measurement shots and limited computational resources.

The rest of the paper is organized as follows. Section 2 review related works. Section 3 discusses quantum feature encoding, kernel estimation, and variational enhancements. Section 4 describes the algorithm design. Section 5 presents numerical results, followed by the conclusions with a discussion of open challenges and future directions in Section 6.

2 Related Works

The intersection of QC and machine learning has given rise to the field of QML. Early study has introduced a quantum algorithm for SVMs based on the Harrow–Hassidim–Lloyd (HHL) method for solving linear systems of equations [52]. Although this approach offers exponential speedup under strong assumptions, such as access to, qRAM, quantum random access memory and fault-tolerant quantum operations, these requirements make it impractical for near-term quantum hardware. To address these practical limitations of HHL-based methods, recent studies have introduced QSVM frameworks designed for NISQ devices, based on quantum feature maps and kernel estimation [53, 54]. In NISQ-based QSVMs, classical inputs are encoded into quantum states using parameterized circuits, and kernel entries are obtained by estimating inner products between quantum states. The resulting kernel matrix is then used in a classical SVM optimization. This approach enables implementation on near-term quantum devices and has been demonstrated on small-scale datasets [53]. A connection between quantum feature maps and classical kernel methods has been also studied by interpreting quantum kernels within the framework of reproducing kernel Hilbert spaces [54]. These studies have not only shown potential quantum advantages, introducing connection between quantum kernels and classical ML frameworks, but also enabled the development of near-term quantum kernel-based learning methods.

Despite these advances, several challenges limit the scalability and performance of QSVMs. A comprehensive overview of quantum learning algorithms has been presented, discussing key challenges such as noise resilience, data encoding, and hardware scalability, as well as potential application of QML [55]. Recent studies have explored variational and hybrid approaches to improve model expressiveness and robustness to noise. Variational quantum classifiers introduce trainable circuit parameters optimized using classical methods, enabling shallower circuits and improved tolerance to hardware noise [56]. This approach has been further extended through quantum variational kernel methods, in which both the quantum feature map and the classifier are jointly optimized to enhance classification performance [57]. In addition, experimental studies on existing quantum hardware have demonstrated the feasibility of implementing QSVMs using shallow circuits on small problem instances. Implementations on trapped-ion quantum processors have shown performance comparable to classical SVMs for limited-scale classification tasks, highlighting both the practical feasibility and scalability limitations of QSVMs on current quantum devices [58].

Unlike existing works, in this study, we focus on improving the scalability of QSVMs by integrating classical kernel approximation techniques into quantum kernel learning. Specifically, we study the incorporation of the Nyström approximation to reduce the cost of quantum kernel evaluations, investigate the impact of different landmark selection strategies on approximation quality and classification performance, and apply principal component analysis to reduce the dimensionality of input data prior to quantum feature encoding. Through this hybrid quantum–classical framework, the proposed approach aims to address practical limitations related to kernel construction cost and circuit complexity in QSVMs.

Table 1 List of abbreviations

Abbreviation	**Meaning**
SVM	Support Vector Machine
QSVM	Quantum Support Vector Machine
QML	Quantum Machine Learning
QC	Quantum Computing
NISQ	Noisy Intermediate-Scale Quantum
PCA	Principal Component Analysis
PSD	Positive Semi-Definite
RBF	Radial Basis Function
CNOT	Controlled-NOT Gate

Table 2 List of symbols

Symbol	**Meaning**
$\mathbf{x}_i \in \mathbf{R}^d$	Input feature vector of sample i
$y_i \in \{-1, +1\}$	Class label of sample i
$\mathbf{w}$	Weight vector defining separating hyperplane
b	Bias term of the hyperplane
ξ_i	Slack variable allowing margin violations
C	Regularization parameter controlling margin vs. error
α_i	Lagrange multiplier in SVM dual formulation
$K(\mathbf{x}_i, \mathbf{x}_j)$	Kernel function between samples i and j
$\mathbf{K} \in \mathbf{R}^{n \times n}$	Full Gram (kernel) matrix
$U_\phi(\mathbf{x})$	Quantum feature map (unitary encoding)
$\|\phi(\mathbf{x})\rangle$	Quantum state representing encoded data x
$K(\mathbf{x}, \mathbf{z})$	Quantum kernel (fidelity between feature states)
$Z = \{\mathbf{z}_1, \ldots, \mathbf{z}_m\}$	Set of m landmark points (Nyström)
$W = K(Z, Z)$	Kernel submatrix among landmarks
$C = K(\mathcal{X}, Z)$	Kernel matrix between dataset $\mathcal{X}$ and landmarks
U, Σ	Eigenvectors and eigenvalues of W
$\widetilde{\Phi}(\mathbf{x})$	Nyström feature embedding of $\mathbf{x}$
q	Number of qubits used in quantum circuits
S	Number of measurement shots for kernel estimation
L	Number of layers in quantum feature map (depth)

3 Quantum Computing Preliminaries

3.1 Fundamental of Quantum Computing

QC is an active research area that uses the rules of quantum physics to build new types of computers for storing and processing information. Unlike classical computers where information is stored in bits, which can be either 0 or 1. In a quantum computer, information is represented by quantum bits, referred to as qubits. A qubit also has two basic states, $|0\rangle$ and $|1\rangle$, which can be associated with the ground and excited states of a physical system, but unlike a classical bit, it can exist in a superposition of both states at the same time [16]. Superposition allows a qubit to represent a combination of $|0\rangle$ and $|1\rangle$, which gives quantum computers access to a larger space of possible states. A general single-qubit state can be written as follows:

$$|\psi_1\rangle = \alpha_0 |0\rangle + \alpha_1 |1\rangle \in \mathbf{C}^2, \tag{1}$$

where α_0 and α_1 are complex probability amplitudes satisfying the normalization $|\alpha_0|^2 + |\alpha_1|^2 = 1$. This means that when taking the measurement, the superposition collapses, and the qubit becomes either a state $|0\rangle$ with probability $\|\alpha_0\|^2$ or $|1\rangle$ with probability $|\alpha_1|^2$, as displayed in Fig. 1.

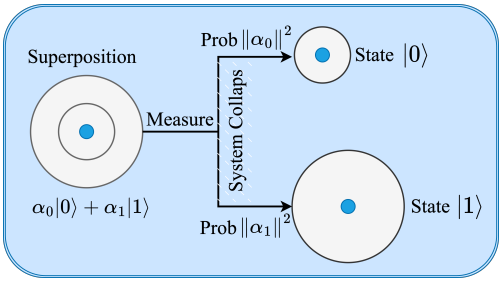


Fig. 1 Illustration of a qubit superposition

The idea of superposition extends directly from a single qubit to systems with multiple qubits. While a single qubit is a combination of the basis states $|0\rangle$ and $|1\rangle$, adding more qubits increases the size of the state space exponentially. For example, a system of two qubits has four computational basis states $|00\rangle, |01\rangle, |10\rangle$, and $|11\rangle$ which correspond to all possible combinations of the individual qubit states. A general two-qubit state is a superposition of these basis states and can be expressed as follows:

$$|\psi_2\rangle = \alpha_{00} |00\rangle + \alpha_{01} |01\rangle + \alpha_{10} |10\rangle + \alpha_{11} |11\rangle, \tag{2}$$

where $|\psi_2\rangle \in \mathbf{C}^4$ and the amplitudes satisfy $|\alpha_{00}|^2 + |\alpha_{01}|^2 + |\alpha_{10}|^2 + |\alpha_{11}|^2 = 1$. This representation generalizes to larger systems. An n-qubit state lies in a 2^n-dimensional

complex vector space and can be written as follows:

$$|\psi\rangle = \sum_{x\in\{0,1\}^n} \alpha_x |x\rangle , \tag{3}$$

with suitable weights normalization $\sum_{x\in\{0,1\}^n} |\alpha_x|^2 = 1$ [19]. The ability to work within such a large state space offers quantum algorithms a computational advantage for solving certain problems more efficiently than classical approaches.

Quantum entanglement is another key feature of multi-qubit systems. While superposition allows a single qubit to occupy multiple states simultaneously, entanglement produces strong correlations between qubits that have no classical counterpart [16, 19]. These non-classical correlations are essential for achieving computational advantages and underpin the performance of many quantum algorithms, communication protocols, and quantum learning methods.

3.2 Quantum Gates and Operations

Quantum computers are designed to run algorithms that can outperform certain classical approaches. To develop and execute these algorithms effectively, the states of qubits must be manipulated in a precise and controlled way. This is achieved through a quantum circuit which is constructed as an ordered sequence of quantum gates acting on qubits [16]. Each gate changes the qubit state according to a specific transformation. A quantum gate acting on n qubits is represented by a $(2^n \times 2^n)$ unitary matrix $\mathbf{U}$, which satisfies $\mathbf{U}\mathbf{U}^\dagger = \mathbf{U}^\dagger\mathbf{U} = \mathbf{I}$, where $\mathbf{U}^\dagger$ is the adjoint (conjugate transpose) of $\mathbf{U}$, and $\mathbf{I} \equiv \mathbb{1}$ denotes the identity operator. This unitarity ensures that quantum operations preserve probability and maintain the normalization of the quantum state.

Common single-qubit gates include the Pauli operators (e.g., Pauli-X), the Hadamard gate $\mathbf{H}$, and rotation gates such as $\mathbf{R}_z$. While the Pauli matrices $(\mathbf{X}, \mathbf{Y}, \mathbf{Z})$ form a fundamental set of operations that generate many single-qubit transformations, the Hadamard gate, which has no classical analogue, plays a unique role by creating equal-amplitude superposition from computational basis states and serving as a key component in many quantum algorithms. These gates are defined as:

$$\begin{aligned} \mathbf{X} \equiv \sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \mathbf{Y} \equiv \sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \\ \mathbf{Z} \equiv \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \ \mathbf{H} = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{pmatrix}. \end{aligned} \tag{4}$$

With the ability to generate superposition and control phase relationships between basis states, these operations play a central role in quantum computation. For example, applying the Hadamard operator $\mathbf{H}$ to the ground state $|0\rangle$, the plus superposition can be then derived as follows:

$$\mathbf{H}|0\rangle = \frac{1}{\sqrt{2}} (|0\rangle + |1\rangle) = |+\rangle . \tag{5}$$

Additionally, aside from the Pauli operators, there exist three additional Pauli-rotation gates that allow continuous rotations of the state vector on the Bloch sphere. They are generated by taking the exponential of the Pauli operators, given by:

$$\begin{aligned}
\mathbf{R}_X(\theta) &= e^{-i\frac{\theta}{2}\mathbf{X}} = \cos\frac{\theta}{2}\mathbf{I} - i\sin\frac{\theta}{2}\mathbf{X}, \\
\mathbf{R}_Y(\theta) &= e^{-i\frac{\theta}{2}\mathbf{Y}} = \cos\frac{\theta}{2}\mathbf{I} - i\sin\frac{\theta}{2}\mathbf{Y}, \\
\mathbf{R}_Z(\theta) &= e^{-i\frac{\theta}{2}\mathbf{Z}} = \cos\frac{\theta}{2}\mathbf{I} - i\sin\frac{\theta}{2}\mathbf{Z}.
\end{aligned} \tag{6}$$

It is apparent that rotations along two different axes can generate any arbitrary single-qubit unitary. In particular, any single-qubit gate $\mathbf{U}$ can be decomposed using Euler angles α, β, and γ as follows:

$$\mathbf{U} = \mathbf{R}_Z(\alpha)\,\mathbf{R}_Y(\beta)\,\mathbf{R}_Z(\gamma). \tag{7}$$

This decomposition motivates the parameterization used in many hardware platforms, including IBM's superconducting qubits, where the universal single-qubit gate $\mathbf{U}_3$ is defined as:

$$\mathbf{U}_3(\theta,\phi,\lambda) := \begin{pmatrix} \cos\frac{\theta}{2} & e^{-i\lambda}\sin\frac{\theta}{2} \\ e^{i\phi}\sin\frac{\theta}{2} & e^{i(\phi+\lambda)}\cos\frac{\theta}{2} \end{pmatrix}. \tag{8}$$

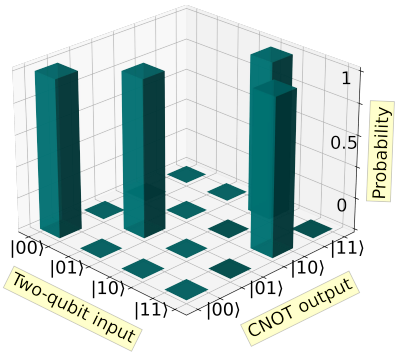


Fig. 2 Illustration of CNOT gate acting on the basis states.

While single-qubit gates provide basic state transformations for local operations, universal quantum computation requires multi-qubit gates. These gates enable correlated transformations that cannot be achieved with single-qubit operations alone. In particular, two-qubit gates are essential for generating entanglement, a key resource underpinning the performance of many quantum algorithms. A widely used two-qubit operator is the Controlled-NOT (CNOT) gate, which flips the 2^{nd} qubit conditioned on the 1^{st} qubit being in the $|1\rangle$ state, for instance, $\text{CNOT}\,|00\rangle = |00\rangle$, $\text{CNOT}\,|10\rangle = |11\rangle$.

Fig. 2 illustrates the CNOT circuit and its action on the computational basis states. Mathematically, the controlled-bit-flip gate or CNOT gate can be expressed as follows:

$$\text{CNOT} : |x_1\rangle |x_2\rangle \rightarrow |x_1\rangle |x_1 \oplus x_2\rangle , \tag{9}$$

where x_1, x_2 denotes the *control* and the *target* qubit, respectively and $\oplus$ represents a XOR function. The CNOT gate is a central component of many quantum algorithms and is widely used in state preparation, entanglement generation, and quantum error correction.

Multi-qubit operations extend the capabilities of two-qubit interactions to larger systems. For an n-qubit system, the state lie in 2^n-dimensional Hilbert space and any operation acting on this state corresponds to a $(2^n \times 2^n)$ unitary matrix. These multi-qubit gates can be built from smaller gates using the tensor product. For example, if $\mathbf{U}_1$ acts on n_1 qubits and $\mathbf{U}_2$ acts on n_2 qubits, then the tensor $\mathbf{U}_1 \otimes \mathbf{U}_2$ implements an $(n_1 + n_2)$-qubit transformation. This construction allows large quantum operations to be built from small, well-defined components. In practice, arbitrary single-qubit rotations combined with a two-qubit entangling gate such as CNOT are sufficient to approximate any unitary on an n-qubit system. This enables the basis for implementing quantum circuits on both simulators and real quantum hardware.

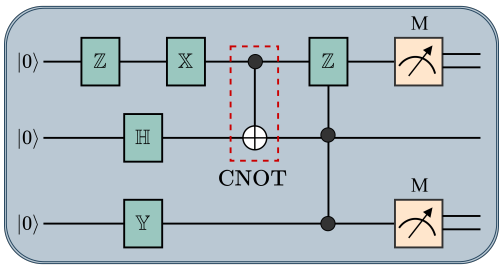


Fig. 3 Visualization of a simple quantum circuit.

By arranging these one-qubit and multi-qubit gates in a specified order, we obtain a quantum circuit that implements a larger computation. Each gate corresponds to a unitary transformation, and the entire circuit corresponds to the product of these unitaries in the order they are applied. The computation ends with measurement, which converts the final quantum state into classical outcomes according to the encoded probability distribution. Measurement collapses the quantum state into one of the computational basis states, with probabilities determined by the amplitudes generated by the circuit. Running the same circuit multiple times yields a distribution over measurement outcomes, which is used as the circuit's classical output. A simple quantum circuit is shown in Fig. 3. In this circuit, a $\mathbf{Z}$ gate is applied to the 1^{st} qubit, followed by $\mathbf{X}, \mathbf{H}$, and$\mathbf{Y}$ gates on the three qubits. A CNOT gate is then applied to the 1^{st} and 2^{nd}qubits, followed by a three-qubit gate acting on all qubits. The circuit concludes with measurements on the 1^{st} and 3^{rd} qubits, yielding classical results.

4 Nyström–Quantum SVM Algorithm Design

This section reviews classical SVMs and kernel methods, highlights the computational challenges of kernel evaluation, and introduces quantum feature maps and quantum kernels that embed data into high-dimensional Hilbert spaces. We then integrate the Nyström approximation into QSVM to enable scalable training via efficient low-rank kernel compression.

4.1 Classical Support Vector Machines

A classical SVM is a supervised learning model primarily designed for binary classification. It aims to construct an optimal separating hyperplane in a (possibly high-dimensional) feature space such that the margin between the two classes is maximized [2, 4]. The data points closest to this hyperplane, known as *support vectors*, determine the classifier. Given a labeled dataset $\{(\mathbf{x}_i, y_i)\}_{i=1}^n$ with input feature vectors $\mathbf{x}_i \in \mathbf{R}^d$ and labels $y_i \in \{-1, +1\}$, the soft-margin SVM seeks a hyperplane defined by $(\mathbf{w}, b)$ that balances margin maximization with tolerance for noisy or non-separable samples. This leads to the following convex quadratic program [??]:

$$\begin{aligned} \min_{\mathbf{w},b,\boldsymbol{\xi}} \quad & \frac{1}{2}\|\mathbf{w}\|_2^2 + C\sum_{i=1}^{n}\xi_i \\ \text{s.t.} \quad & y_i(\mathbf{w}^\top \mathbf{x}_i + b) \geq 1 - \xi_i,\ \forall i = \overline{1,n}, \\ & \xi_i \geq 0,\ \forall i = \overline{1,n}, \end{aligned} \tag{10}$$

where $\mathbf{w}$ and b denote the weight vector and bias that define the optimal separating hyperplane, ξ_i are slack variables capturing deviations from correct classification, and $C > 0$ is a regularization coefficient. A higher value of C imposes a stronger penalty on misclassifications, leading to a smaller and tighter margin that fits the training data strictly, while a lower C value allows for a larger margin and more tolerance for misclassified points, which can improve generalization in noisy datasets. The optimization problem (10), known as the primal problem, admits a global minimum and can be equivalently reformulated into the dual problem by introducing Lagrange multipliers α_i as follows [5]:

$$\begin{aligned} \max_{\boldsymbol{\alpha}} \quad & -\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\,\alpha_j\,y_i\,y_j\,\mathbf{x}_i^T\mathbf{x}_j + \sum_{i=1}^{n}\alpha_i \\ \text{s.t.} \quad & \sum_{i=1}^{n} y_i\,\alpha_i = 0, \\ & 0 \leq \alpha_i \leq C,\ \forall i = \overline{1,n}. \end{aligned} \tag{11}$$

It is observed that the dual problem depends only on the inner products of the training data, and it is easier to solve, as it involves only the Lagrange multipliers.

In practice, the linear formulation of SVMs is insufficient as real-world datasets often exhibit complex, nonlinear structures that cannot be captured by a single hyperplane in $\mathbf{R}^d$ [4, 6]. To address this limitation, SVMs are extended through kernel

methods, which implicitly map the input data into a high-dimensional feature space $\mathcal{H}$ where linear separability becomes achievable [4]. This is achieved using a kernel function that computes inner products in the feature space without explicitly performing the mapping, given by:

$$K(\mathbf{x}_i, \mathbf{x}_j) := \langle \phi(\mathbf{x}_i), \phi(\mathbf{x}_j) \rangle_{\mathcal{H}} = \phi(\mathbf{x}_i)^\top \phi(\mathbf{x}_j), \tag{12}$$

where $\phi(\cdot) : \mathbf{R}^d \to \mathcal{H}$ denotes the mapped feature vector. By substituting the kernel function (12) for the inner products in the dual formulation (11), the classifier can operate as if it were constructing a separating hyperplane in $\mathcal{H}$ without explicitly computing the mapping. This substitution reformulates the dual problem (11) entirely in terms of kernel evaluations, thereby enabling nonlinear separation in the original input space while preserving the convex structure of the SVM. Consequently, the SVMs equipped with appropriate kernels are capable of learning highly nonlinear decision boundaries, bridging the gap between the simplicity of linear models and the flexibility required for complex classification tasks.

Despite their effectiveness, classical SVMs suffer from scalability issues with large datasets. Although kernel methods provide powerful representational capabilities, they introduce significant computational and storage challenges. For a training set of size n, the kernel matrix, defined as follows:

$$\mathbf{K} = [K(\mathbf{x}_i, \mathbf{x}_j)]_{i,j=1}^n \in \mathbf{R}^{n \times n}, \tag{13}$$

contains all pairwise kernel evaluations. It is obvious that constructing the full $n \times n$ kernel matrix requires quadratic resources, that is, $O(n^2)$ evaluations and $O(n^2)$ memory space for storing. Also, solving the quadratic program with matrix operations (e.g., inversion, eigen-decomposition) can further scale up to $O(n^3)$. For large datasets with millions of samples, this becomes prohibitive, and training can be computationally intractable. This quadratic bottleneck has motivated the exploration of low-rank approximations that approximate $\mathbf{K}$ while ensuring prediction performance.

4.2 Quantum Support Vector Machines

Quantum kernel methods extend the classical SVM by leveraging quantum feature maps to embed classical data into exponentially large Hilbert spaces. Let $\mathbf{x} \in \mathbf{R}^d$ denote classical data. To encode this data into a quantum state space, the circuit associated with the feature map applies a unitary operator $U_\phi(\mathbf{x})$ to an initial register of q qubits prepared in the computational ground state. The resulting quantum feature representation is given by:

$$|\phi(\mathbf{x})\rangle = U_\phi(\mathbf{x})|0\rangle^{\otimes q}. \tag{14}$$

The corresponding quantum kernel is then defined as the fidelity between two feature states, given by:

$$\begin{aligned} K(\mathbf{x}, \mathbf{z}) &= \left| \langle \phi(\mathbf{x}) \,|\, \phi(\mathbf{z}) \rangle \right|^2 \\ &= \left| \langle 0|^{\otimes q} \, U_\phi^\dagger(\mathbf{z}) \, U_\phi(\mathbf{x}) \, |0\rangle^{\otimes q} \right|^2. \end{aligned} \tag{15}$$

This enables kernel evaluations in Hilbert spaces whose dimension may grow exponentially with the number of qubits, potentially offering representational advantages

over classical kernels. As the kernel (15) is defined as the fidelity between quantum states, $K(\mathbf{x}, \mathbf{z})$ satisfies the Mercer's theorem and its corresponding Gram matrices are guaranteed to be positive semi-definite.

In practice, quantum kernel $K(\mathbf{x}, \mathbf{z})$ can be evaluated on hardware using interference-based circuits such as the SWAP test, or through a more hardware-efficient overlap estimation circuit that applies the composed feature-map unitary $U_\phi^\dagger(\mathbf{z}), U_\phi(\mathbf{x})$ to the ground state and extracts the fidelity from the probability of measuring in the computational basis. Let $P_0 = |0\rangle\langle 0|^{\otimes q}$ denote the projector onto the all-zero basis state. The resulting probability of observing this outcome is then expressed as follows:

$$p_0(\mathbf{x}, \mathbf{z}) = \mathrm{Tr}\big(P_0\, U_\phi^\dagger(\mathbf{z}) U_\phi(\mathbf{x}) |0\rangle\langle 0|^{\otimes q} U_\phi^\dagger(\mathbf{x}) U_\phi(\mathbf{z})\big). \tag{16}$$

This probability corresponds directly to the fidelity between the two quantum feature states. Hence, quantum processors can compute the entries of the kernel matrix through state preparation, unitary composition, and measurement. This enables the application of classical SVM optimization to kernel matrices obtained directly from quantum circuit evaluations.

Encoding strategies play an important role in shaping model expressivity and trainability. A variety of data-encoding techniques have been explored, including angle (rotation) encoding, amplitude encoding, basis encoding, and Hamiltonian- or **ZZ**-style feature maps, often augmented with data re-uploading and structured entangling layers to enhance representational capacity.

4.2.1 Angle (Phase/Rotation) Encoding

Angle encoders implement a direct, hardware-efficient lift by mapping coordinates to single-qubit rotations. Given scaling parameters $\alpha, \beta \in \mathbf{R}$, the angle-embedding can be expressed as:

$$U_\phi(\mathbf{x}) = \prod_{i=1}^{q} R_Z(\alpha \mathbf{x}_i)\, R_Y(\beta \mathbf{x}_i), \tag{17}$$

where $R_\nu(\theta) = \exp\big(-i\,\frac{\theta}{2}\,\sigma_\nu\big)$. For a single qubit with $U_\phi(\mathbf{x}) = R_Y(\beta\mathbf{x})$ and no entanglement, the overlap value can be written as follows:

$$K(\mathbf{x}, \mathbf{z}) \;=\; |\langle\phi(\mathbf{x})|\phi(\mathbf{x})\rangle|^2 \;=\; \cos^2\Big(\tfrac{\beta}{2}\,(\mathbf{x} - \mathbf{z})\Big). \tag{18}$$

For a register of q qubits without entanglement, the overall feature state is a tensor product of single-qubit states, and the corresponding kernel factorizes into the product of the individual cosine-squared terms. This generates a stationary, periodic kernel whose structure is governed by the rotation parameters. Introducing entangling operations (e.g., CNOT gates) couples the qubits and generates more complex non-linear interactions between features, expanding the spectral support of the induced kernel $\mathbf{K}$ and enabling better representational capacity. The rotation scales α and β effectively serve as frequency-control parameters where small values lead to smoother,

low-frequency responses, whereas larger values introduce higher-frequency variations that increase model flexibility but may also lead to overfitting when data are limited.

4.2.2 Commuting Ising/ZZ Feature Maps

Commuting Ising or **ZZ** feature maps encode classical data using diagonal Hamiltonians generated by single-qubit **Z** terms and two-qubit **ZZ** interactions, defined as follows:

$$U_\phi(\mathbf{x}) \;=\; \exp\Big(i\sum_{i=1}^{q}\alpha_i(\mathbf{x})\,\mathbf{Z}_i \;+\; i\sum_{i<j}\beta_{ij}(\mathbf{x})\,\mathbf{Z}_i\mathbf{Z}_j\Big), \tag{19}$$

where α_i and β_{ij} denote scalar feature functions that determine how the input data influences the local phases and pairwise coupling interactions, respectively. Since all generators $\mathbf{Z}_i$ and $\mathbf{Z}_i\mathbf{Z}_j$ commute, the unitary $U_\phi(\mathbf{x})$ is diagonal in the computational basis ($U_\phi(\mathbf{x})$ shares an eigenbasis with the computational basis) and can be implemented using shallow circuits. This makes **ZZ**-style feature maps practical for near-term hardware.

4.2.3 Data Re-Uploading with Entanglement

Data re-uploading is a widely used method for increasing the expressive power of quantum feature maps. The main idea is to apply the data-dependent single-qubit rotations several times throughout the circuit, with each round of encoding followed by an entangling operation. Repeated encoding enables the circuit to build progressively richer nonlinear feature representations, while entangling gates combine information across qubits to capture complex relationships between the input dimensions. A depth-L data re-uploading feature map can be written as follows:

$$\begin{aligned} U_\phi(\mathbf{x}) = \; & \Big(\textstyle\prod_{i=1}^{q} R_Z(\alpha_L\mathbf{x}_i)\,R_Y(\beta_L\mathbf{x}_i)\Big)U_{\mathrm{ent},L} \\ & \cdots\Big(\textstyle\prod_{i=1}^{q} R_Z(\alpha_1\mathbf{x}_i)\,R_Y(\beta_1\mathbf{x}_i)\Big)U_{\mathrm{ent},1}, \end{aligned} \tag{20}$$

where each layer consists of single-qubit encodings followed by an entangling unitary operation $U_{\mathrm{ent},L}$. The parameters $\{\alpha_\ell,\beta_\ell\}_{\ell=1}^{L}$ control the strength of the data encoding at each layer, while the design of the entangling gates determines how information is shared across qubits. As the depth L increases, the circuit can compose more combinations of rotation and entangling operations, leading to greater expressive capacity in the resulting feature map. With appropriate parameter choices and a moderate number of layers, data re-uploading circuits can approximate a wide range of continuous functions, while keeping the circuit depth suitable for current devices.

4.2.4 Amplitude Encoding

Amplitude encoding represents classical data by embedding a normalized vector directly into the amplitudes of a quantum state. Given an input $x \in \mathbb{R}^d$, the encoding prepares the normalized quantum feature state $|\phi(\mathbf{x})\rangle = \mathbf{x}/\|\mathbf{x}\|$ such that the components of $\mathbf{x}$ serve as the amplitude coefficients of the computational basis. A minimum of q qubits is required to satisfy $2^q \geq d$, enabling the encoding of an entire d-dimensional

vector into the amplitudes of a 2^q-dimensional Hilbert space. The kernel induced by amplitude encoding can be then expressed as follows:

$$K(\mathbf{x}, \mathbf{z}) \;=\; \frac{|\langle \mathbf{x}, \mathbf{z} \rangle|^2}{\|\mathbf{x}\|^2 \|\mathbf{z}\|^2}. \tag{21}$$

Despite its representational advantages, amplitude encoding poses significant practical limitations on near-term quantum hardware. Preparing arbitrary amplitude states typically requires circuit depths that grow with the input dimension or access to specialized architectures such as QRAM, which are not available on current NISQ platforms. Consequently, although amplitude encoding allows compact embeddings in large Hilbert spaces, its use in practical quantum machine-learning experiments remains limited.

4.3 Nyström Approximation for Quantum Kernels

Quantum kernel methods, particularly QSVMs, offer a powerful framework for exploiting quantum feature spaces whose dimensionality can scale exponentially with the number of qubits. In practice, however, training QSVMs requires constructing the full kernel matrix, which requires $O(n^2)$ quantum overlap (or fidelity) evaluations. Each evaluation typically involves multiple circuit repetitions to suppress statistical noise. As the dataset size n increases, this quadratic scaling in quantum queries and classical memory becomes prohibitively expensive, posing a significant limitation to deploying QSVMs on current NISQ devices. To address this challenge, low-rank kernel approximation techniques, most notably the Nyström method, provide an effective strategy for reducing both quantum and classical resource demands while preserving the expressive power of quantum feature maps.

The Nyström approximation offers a principled approach to reducing the cost of quantum kernel evaluation by constructing a low-rank surrogate of the kernel matrix. Given a training set $\mathcal{X} = \{\mathbf{x}_1, \cdots, \mathbf{x}_n\}$, the method begins by selecting a subset of $m \ll n$ landmark points $Z = \{\mathbf{z}_1, \ldots, \mathbf{z}_m\} \subset \mathcal{X}$ that capture the structure of the dataset. Using these landmarks, two essential kernel sub-matrices then can be derived as follows:

$$W = K(Z, Z) \in \mathbf{R}^{m \times m},\; C = K(\mathcal{X}, Z) \in \mathbf{R}^{n \times m}, \tag{22}$$

where $W_{i,j} = K(\mathbf{x}_i, \mathbf{z}_j)$ encodes pairwise similarities within the landmark set and $C_{i,j} = K(\mathbf{z}_i, \mathbf{z}_j)$ contains the kernel evaluations between all samples and the landmark points. Since quantum kernels are positive semidefinite (PSD), the sub-matrix W is also PSD and admits an eigendecomposition $W = U\Sigma U^\top$, where U contains orthonormal eigenvectors and $\Sigma = \mathrm{diag}\,(\sigma_1, \cdots, \sigma_m)$ contains the non-negative eigenvalues. This decomposition enables the construction of an explicit low-dimensional feature representation. Hence, for any input $\mathbf{x} \in \mathcal{X}$, the Nyström feature map can be then defined as:

$$\tilde{\Phi}(\mathbf{x}) = k_Z(\mathbf{x})^\top U \Sigma^{-1/2} U^\top,\;\; k_Z(\mathbf{x}) = \begin{bmatrix} K(\mathbf{x}, \mathbf{z}_1) \\ \vdots \\ K(\mathbf{x}, \mathbf{z}_m) \end{bmatrix}, \tag{23}$$

where $k_Z(\mathbf{x}) \in \mathbb{R}^m$ denotes the landmark-kernel vector. It is obvious that the Nyström embedding $\tilde{\Phi}(\mathbf{x}) \in \mathbb{R}^m$ provides an explicit approximation of the implicit quantum feature representation $\phi(\mathbf{x})$, and by construction of the embedding (23), the inner product between two Nyström feature vectors can be expressed as follows:

$$\tilde{\Phi}(\mathbf{x})^\top \tilde{\Phi}(\mathbf{x}') = k_Z(\mathbf{x})^\top U \Sigma^{-1} U^\top k_Z(\mathbf{x}') \approx K(\mathbf{x}, \mathbf{x}'), \tag{24}$$

which provides a rank-m approximation to the quantum kernel value $K(\mathbf{x}, \mathbf{x}') \in \mathbb{R}^{n \times n}$. This implies that the Nyström embedding preserves the quantum kernel similarities up to the approximation error induced by projecting onto the rank-m subspace spanned by the landmark set $Z \in \mathcal{X}$. Applying the Nyström embedding for all samples we can obtain the corresponding low-rank approximation of the full quantum kernel matrix, as follows:

$$\tilde{K} = C W^\dagger C^\top = C U \Sigma^{-1} U^\top C^\top, \tag{25}$$

where $W^\dagger$ denote the Moore–Penrose pseudoinverse of the landmark Gram matrix W.

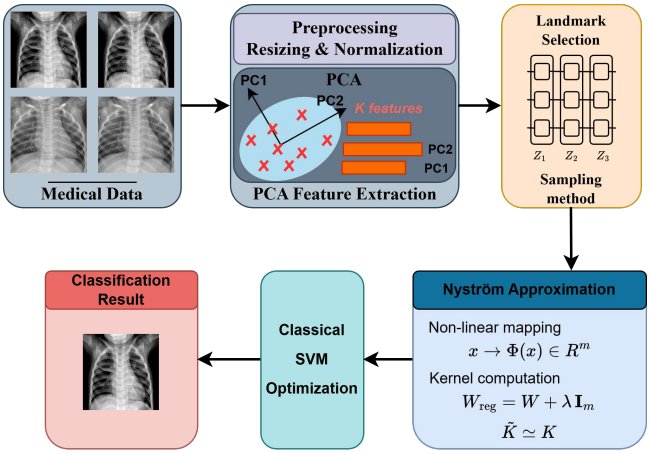


Fig. 4 Architecture of the classical Nyström SVM.

From (25), it is observed that $\tilde{K}$ is obtained by projecting the columns of C onto the spectral subspace of W and scaling by its corresponding eigenvalues Σ, which reduces the effective dimensionality of the feature space from n to m. Additionally, constructing the sub-matrices W and C requires only $O(nm)$ quantum kernel evaluations, each associated with a quantum state fidelity computation, representing a substantial reduction from the $O(n^2)$ evaluations needed to construct the full Gram matrix. Since each kernel evaluation corresponds to a quantum circuit execution, lowering the count from $O(n^2)$ to $O(nm)$ directly leads to a substantial decrease in runtime and quantum hardware usage. Thus, the Nyström method yields an efficient low-rank approximation that preserves kernel values to high accuracy while significantly reducing the quantum and classical computational resources, as shown in Table 3.

In practice, the landmark Gram matrix W often becomes ill-conditioned due to correlations among the selected landmarks, finite-precision effects, and noise arising

Table 3 Full vs *NQ-SVM* Complexity.

Method	Kernel Evaluations	Memory
Full QSVM	$O(n^2)$	$O(n^2)$
NQ-SVM	$O(nm)$	$O(nm)$

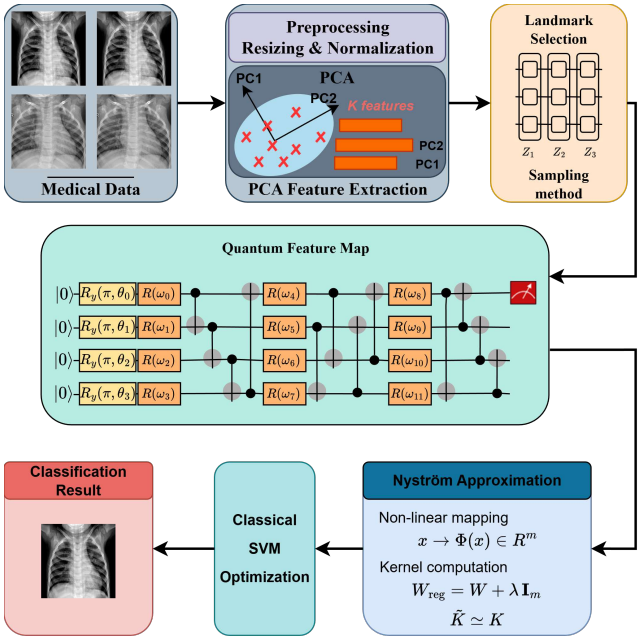


Fig. 5 Architecture of the proposed *NQ-SVM* model.

from quantum kernel evaluations. This can lead to unstable eigenvalues and unreliable Nyström feature embeddings. To ensure a stable low-rank approximation, we introduce a small regularization term and form the regularized matrix, defined as follows:

$$W_{\text{reg}} = W + \lambda \mathbf{I}_m, \tag{26}$$

where $\mathbf{I}_m$ denotes the $m \times m$ identity matrix, $\lambda > 0$ is a ridge regularization coefficient. The corresponding Nyström feature map is then constructed using the eigen-decomposition of the regularized matrix, given by:

$$\widetilde{\Phi}(\mathbf{x}) = k_Z(\mathbf{x})^\top W_{\text{reg}}^{-1/2} = k_Z(\mathbf{x})^\top U(\Sigma + \lambda \mathbf{I}_m)^{-1/2} U^\top, \tag{27}$$

Regularization ensures that all eigenvalues $\sigma_j + \lambda$ remain strictly positive, allowing a well-defined and numerically stable computation of $(\Sigma_{\text{reg}}^{-1/2} = \Sigma + \lambda I)^{-1/2}$ and leading to a robust Nyström approximation of the quantum kernel, which is essential for reliable QSVM training.

4.4 Landmark Selection

The performance of the Nyström approximation depends typically on the choice of landmark set. These points determine the subspace used to approximate the quantum kernel matrix, and different selection strategies provide different trade-offs between computational cost and approximation accuracy. A common strategy is uniform random sampling (URS), where landmark points are selected uniformly from the training set, given by:

$$\mathbf{z}_j \sim \text{Uniform}(\mathcal{X}),\ \forall j = 1, \cdots, m. \tag{28}$$

Despite its simplicity, this approach can yield strong approximations when the quantum kernel matrix possesses a rapidly decaying eigenvalue spectrum or when the data distribution is sufficiently diverse. Furthermore, in quantum implementations, uniform sampling is particularly advantageous as it requires no additional quantum kernel evaluations for the same encoded data points.

Another widely adopted technique is clustering-based landmark selection, where algorithms such as k-means are employed to identify centroids that act as representative samples of the input space. In this approach, the landmarks are selected as the cluster means, defined as follows:

$$z_j = \mu_j, \quad \mu_j = \frac{1}{|c_j|} \sum_{\mathbf{x} \in c_j} \mathbf{x}, \tag{29}$$

where $\{c_j\}_{j=1}^m$ are clusters obtained by minimizing the total within-cluster variance, given by:

$$\min_{c_1,\ldots,c_m} \sum_{j=1}^{m} \sum_{\mathbf{x} \in c_j} \|\mathbf{x} - \mu_j\|_2^2. \tag{30}$$

The resulting centroids provide a set of landmark points that reflect how the data are grouped and distributed in the input space, thereby offering an informative subset that supports the construction of a reliable low-rank approximation to the kernel matrix. This is particularly effective when the quantum feature map approximately preserves neighborhood relationships in the classical space. By ensuring that the selected landmarks provide balanced coverage over the data distribution, clustering methods frequently lead to improved accuracy in the resulting Nyström approximation.

The k-center method is an alternative approach for selecting informative landmarks, focusing on maximizing dataset coverage by iteratively choosing points that are farthest from the current landmark set, as summarized in Algorithm 1. Starting from an arbitrary initial point $\mathbf{z}_1$, subsequent landmarks are chosen according to the rule defined as follows:

$$\mathbf{z}_{t+1} = \arg\max_{\mathbf{x} \in \mathcal{X}} \min_{1 \le j \le t} \|\mathbf{x} - \mathbf{z}_j\|_2,\ \forall t = 1, \ldots, m-1. \tag{31}$$

This procedure ensures that each new landmark lies in a region of the dataset that has not been well represented by earlier selections, leading to all landmarkbeingre distributed across the dataset rather than concentrated in a single region. In practice,

Algorithm 1 k-Center (Farthest-First) Landmark Selection

Require: Dataset $\{\mathbf{x}_j\}_{j=1}^{n}$, number of landmarks m, seed s
Ensure: Index set $\mathcal{I}$ of m selected landmarks
Initialize RNG with s, pick i_1 uniformly from $\{1, \ldots, n\}$, $\mathcal{I} \leftarrow \{i_1\}$
$d_j \leftarrow \|\mathbf{x}_j - \mathbf{x}_{i_1}\|_2$ for all $j = 1, \ldots, n$
for $t = 2$ **to** m **do**
    $i_t \leftarrow \arg\max_{j \notin I} d_j$
    $\mathcal{I} \leftarrow \mathcal{I} \cup \{i_t\}$
    $d_j \leftarrow \min\big(d_j,\ \|\mathbf{x}_j - \mathbf{x}_{i_t}\|_2\big)$ for all $j = 1, \ldots, n$
end for
return $\mathcal{I}$

k-center often produces a more diverse and representative landmark set than uniform sampling, particularly when the dataset contains clusters or uneven density. This typically leads to better approximations of the kernel matrix and more effective Nyström embeddings.

Additionally, a more advanced approach is leverage-score sampling, where each point is selected with a probability proportional to its contribution to the approximate eigenspace of the kernel matrix. This strategy highlights data points that are especially informative for capturing the structure of the kernel-induced feature space, often leading to better low-rank approximations than uniform sampling. However, computing leverage scores requires partial information about the kernel matrix, and obtaining this information for quantum kernels demands additional quantum circuit evaluations. This added requirement makes leverage-score sampling less practical for quantum kernel methods, compared to other approaches like URS, k-center or k-mean clustering.

4.5 *NQ-SVM* Training

After computing the features $\tilde{\Phi}(\mathbf{x}_i)$, the remaining QSVM training process becomes a standard supervised learning task. The Nyström transformation compresses each data point into an m-dimensional feature vector that captures the structure of the quantum kernel using only the selected landmark set. This enables efficient training without computing or storing the full $n \times n$ kernel matrix, while still preserving the nonlinear relationships encoded by the quantum feature map. Before training the classifier, the Nyström features are standardized using statistics computed from the training split, ensuring that all feature dimensions operate on a comparable scale, $\widetilde{\Phi}(\mathbf{x}) \leftarrow \mathcal{S}_\Phi\big(\widetilde{\Phi}(\mathbf{x})\big)$ with $\mathcal{S}_\Phi(\cdot)$ applying feature-wise centering and scaling. The SVM is then trained on the transformed training features by solving the following optimization problem:

$$\min_{\mathbf{w},b} \ \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1} \max\Big(0,\ 1 - y_i(\mathbf{w}^\top \widetilde{\Phi}(\mathbf{x}_i) + b)\Big), \tag{32}$$

where $C > 0$ is the regularization parameter, controlling the margin–error trade-off. The overall training procedure of the proposed *NQ-SVM* approach is summarized

in Algorithm 2. Once trained, the final QSVM model consists of the learned classifier together with all pre-processing components, enabling prediction on new data by recomputing its Nyström features and applying the classifier. It is obvious that the computational cost is reduced to $O(nm)$ kernel evaluations, and the memory requirement scales as $O(nm)$, representing a dramatic improvement over the $O(n^2)$ requirements of full QSVM. At the same time, the expressivity of quantum kernels is preserved, enabling scalable quantum-enhanced learning.

5 Numerical Results

To evaluate the efficacy of *NQ-SVM*, we conduct numerical experiments on two medical datasets: Breast Cancer (WDBC) and PneumoniaMNIST. For both datasets, we assess classification accuracy, runtime efficiency, and kernel approximation quality under varying numbers of landmarks and landmark selection strategies. We also analyze the effect of quantum measurement shots and demonstrate the scalability advantages of *NQ-SVM* over the full QSVM.

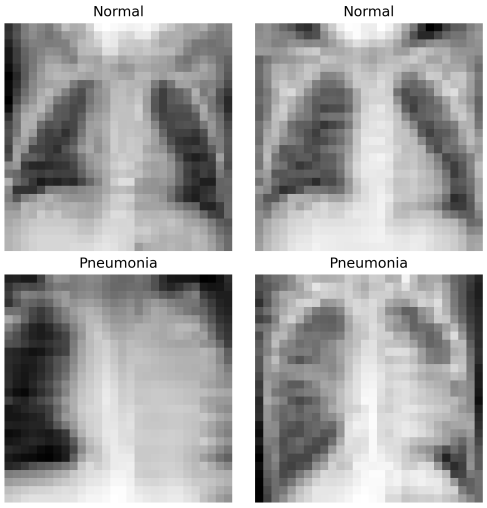
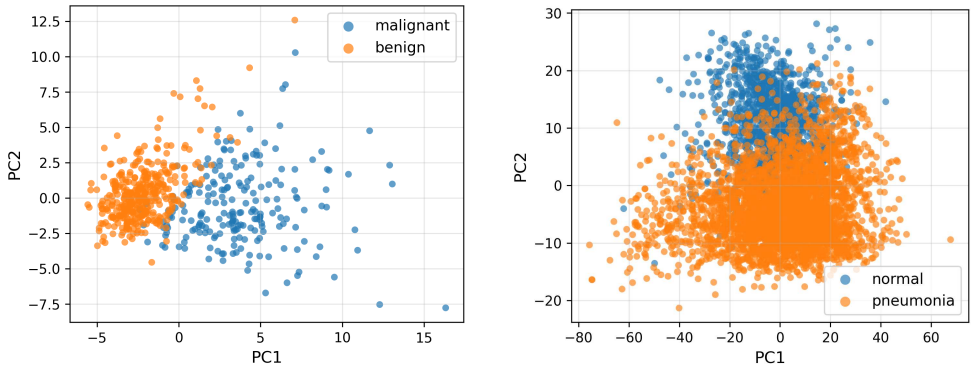


Fig. 6 Sample chest X-ray from PneumoniaMNIST

The WDBC dataset comprises 569 tabular samples obtained from fine-needle aspirate (FNA) images of breast tissue, each labeled as benign or malignant. Each sample is described by 30 continuous diagnostic features extracted from breast-cancer cell nuclei, such as radius, perimeter, symmetry, and texture, providing a relatively clean, low-noise representation of the two classes. PneumoniaMNIST is a large-scale medical imaging dataset designed for automated pneumonia detection. It consists of $5,856$ grayscale chest X-ray images resized to 28×28 pixels, forming a low-resolution medical imaging benchmark suitable for both classical and quantum machine-learning models. Each image is labeled as either normal or pneumonia. Fig. 6 shows example images,

with normal cases on the top row and pneumonia cases on the bottom, highlighting their visual differences.

Unlike the breast cancer dataset, PneumoniaMNIST exhibits higher variation and overlapping visual patterns within each label category arising from differences in patient anatomy, imaging equipment, and acquisition conditions, as shown in the PCA visualizations in Fig. **??**. The PCA projection of WDBC in Fig. **??** reveals relatively compact clusters and a clear separation between malignant and benign samples, with only small overlap near the boundary. In contrast, the PCA scatter for PneumoniaMNIST displays broad overlap between normal and pneumonia cases, with both classes across the first two principal components (PC1 and PC2), as shown in Fig. **??**. The wide spread reflects the diverse visual characteristics of medical X-ray images, including differences in contrast, lung structures, and disease patterns, which are difficult to capture in a small number of principal directions. This implies that PneumoniaMNIST has a more complex structure than the WDBC, and that simple low-dimensional projections may not provide reliable separation between classes. These characteristics imply that PneumoniaMNIST produces a noisier quantum kernel with smaller and less stable eigenvalues, making it a more challenging benchmark for kernel approximation than WDBC. The differences between the smaller, more structured WDBC dataset and the larger, highly heterogeneous PneumoniaMNIST dataset provide a meaningful evaluation of how the proposed *NQ-SVM* performs under different data configurations with varying levels of complexity and difficulty.



(a) WDBC (b) Pneumonia

Fig. 7 PCA 2D plots for datasets.

Before training, the data are standardized to zero mean and unit variance, and projected onto a lower-dimensional subspace using PCA. This preprocessing step reduces noise and redundancy by concentrating the most informative variance into a small number of orthogonal components, yielding a more stable representation that improves generalization while reducing computational cost. To ensure compatibility with quantum circuits, the number of principal components is chosen to match the number of encoding qubits in the quantum feature map. Since current quantum hardware and simulators support only a limited number of qubits at reasonable computational cost,

we consider encoding circuits with 2 to 6 qubits in our experiments. In all cases, the quantum kernels are constructed using a two-layer angle-encoding feature map, followed by controlled-Z (CZ) entangling gates arranged in a linear or ring topology, enabling nearest-neighbor interactions that introduce nonlinearity and inter-qubit correlations.

Table 4 Performance of *NQ-SVM* under varying landmark size m

Landmarks	**WDBC (%)**		**PneumoniaMNIST (%)**	
	Train Acc.	Test Acc.	Train Acc.	Test Acc.
5	87.39	85.09	90.92	83.50
10	91.20	89.47	91.92	84.75
20	95.89	92.11	92.92	86.75
40	97.07	94.74	93.67	86.50
60	97.07	94.74	93.42	86.75
80	97.07	94.74	93.50	86.75
100	97.07	94.74	93.58	86.75

Table 4 summarizes the training and test accuracies of *NQ-SVM* on the WDBC and PneumoniaMNIST datasets for different landmark sizes. For WDBC, which contains relatively low-dimensional and well-structured features, the quantum feature map is implemented using a 4-qubit circuit with a simple line entanglement pattern. The results show that *NQ-SVM* achieves strong and stable performance across a wide range of landmark sizes. As the number of landmarks increases, training accuracy improves from 87.39% to approximately 97%, indicating a more accurate approximation of the quantum kernel. For small landmark sets ($m \leq 10$), test accuracy exhibits moderate variability, ranging from 0.85 to 0.89, due to the limited representation capacity of the low-rank kernel approximation. However, once the number of landmarks reaches $m = 40$, performance stabilizes, with test accuracy converging to approximately 94.74% and training accuracy to 97.07%. This result suggests that the Nyström approximation has captured sufficient quantum kernel structure, and further increases in landmark size yield diminishing performance gains.

For the more complex and heterogeneous PneumoniaMNIST dataset, a 6-qubit circuit with a ring entanglement pattern is employed to provide greater expressive capacity. Similar to WDBC, increasing the number of landmarks generally improves performance, particularly when increasing m from 5 to 20. At $m = 40$, test accuracy drops slightly to 86.50%, which indicates that the selected landmarks may not provide a balanced representation of the data at this size, leading to a small reduction in generalization. For larger landmark sizes ($m \geq 60$), both training and test accuracy stabilize, with training accuracy exceeding 93% and test accuracy converging to approximately 86.75%. These results indicate that, with a sufficient

number of landmarks, the Nyström approximation yields a reliable quantum kernel representation.

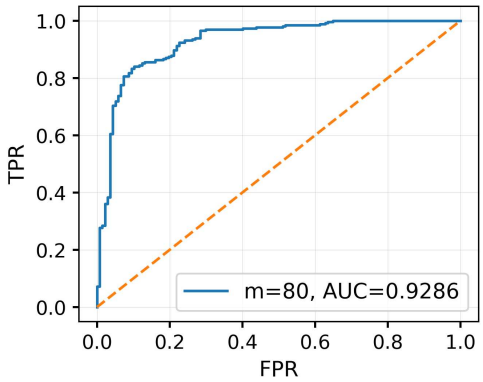


Fig. 8 ROC curve of *NQ-SVM* on PneumoniaMNIST

In medical classification, performance across varying decision thresholds is critical, particularly for imbalanced datasets such as PneumoniaMNIST, where correctly identifying positive cases is clinically important. While accuracy reflects performance at a fixed threshold, it does not fully capture a model's ability to separate classes as the threshold varies. Receiver operating characteristic (ROC) curves and the area under the curve (AUC) address this limitation by quantifying class separability across all thresholds.

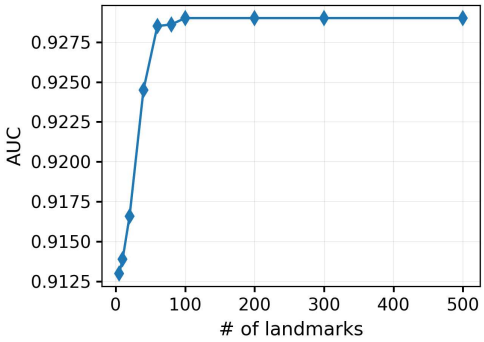


Fig. 9 AUC values for PneumoniaMNIST

Fig. 8 presents the ROC curve of *NQ-SVM* on the PneumoniaMNIST test set when using $m = 80$ landmarks. The curve consistently lies above the diagonal baseline, indicating that the classifier achieves high true positive rates (TPR) across a wide range of false positive rates (FPR). The corresponding AUC value of approximately

0.93 confirms that the classifier ranks positive samples correctly in most pairwise comparisons, showing reliable separation between pneumonia and non-pneumonia cases across varying decision thresholds.

Fig. 9 illustrates how the test AUC varies with the number of landmarks. Consistent with the accuracy trends in Table 4, the AUC improves steadily as the number of landmarks increases from small to moderately large ones, rising from approximately 0.913 at $m = 5$ to about 0.929 at $m = 80$. This trend reflects the increasing expressiveness of the Nyström kernel approximation as more landmarks are incorporated, enabling the classifier to better separate positive cases from negative ones across a wide range of thresholds. However, beyond $m = 80$, AUC values plateau around 0.929, indicating diminishing returns from additional landmarks. These results suggest that moderate landmark sizes are sufficient to capture the essential structure of the quantum kernel while maintaining strong and reliable classification performance.

These results underscore the strength of the Nyström approximation in quantum kernel learning. Despite the dimensionality reduction via PCA and the approximate nature of the kernel reconstruction, *NQ-SVM* maintains high ranking performance across clinically relevant operating points. Since AUC measures the probability that positive samples are assigned higher scores than negative samples, the consistently high AUC values for $m \geq 40$ indicate that the model can reliably distinguish pneumonia from non-pneumonia cases even as decision thresholds vary. This property is important in medical classification, where threshold selection may vary across diagnostic settings or clinical priorities.

To further evaluate the efficiency of the Nyström approximation, we examined the eigenvalue spectrum of the landmark Gram matrix. Fig. 10 presents the eigenvalues for $m = 20$ landmarks. The spectrum exhibits a clear decay, with the first few eigenvalues dominating the distribution. In particular, the largest eigenvalue is more than twice the magnitude of the second, and the top five eigenvalues account for the majority of the total spectral energy. This concentration indicates that the effective dimensionality of the quantum kernel space is substantially lower than the landmark size, justifying the use of low-rank approximations. This rapid decay implies that *NQ-SVM* does

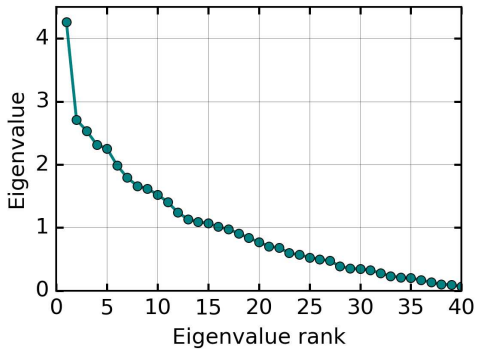


Fig. 10 Eigenvalue spectrum of the landmark Gram matrix

not require the full m-dimensional embedding to achieve strong performance. Instead, truncation to the leading spectral components reduces computational overhead while preserving classification accuracy, consistent with the observed performance plateau beyond $m = 20$ landmarks. Together, Fig. 10 and Table 4 highlights that the proposed method effectively balances scalability and performance by exploiting the spectral redundancy of quantum kernels.

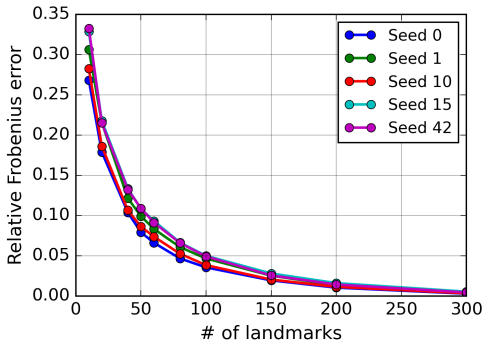


Fig. 11 Nyström approximation error

To quantify the Nyström approximation quality, we compute the relative Frobenius error between the full quantum kernel matrix and its Nyström approximation for varying landmark sizes. Fig. 11 reports the spectral error for m ranging from 10 to 300 across multiple random seeds. Across all trials, the error decreases monotonically with increasing m and exhibits rapid convergence: when $m \leq 50$, the error already falls below 0.05, and for $m \geq 100$, it approaches zero. The low variance across seeds indicates that the approximation is stable with respect to landmark selection. This result aligns with the eigenvalue spectrum in Fig. 10, demonstrating that the quantum kernel matrix has low effective rank. Thus, *NQ-SVM* achieves strong performance with relatively small landmark sets. The practical implication is that computational costs can be reduced substantially without sacrificing accuracy, which is evident in Table 4, where classification performance plateaued once $m \geq 20$. These observations confirm that *NQ-SVM* is both spectrally efficient and empirically robust, making it a scalable framework for quantum kernel methods.

To examine class-level performance, we present the validation confusion matrix for *NQ-SVM* with $m = 20$ landmarks in Fig. 12. The results show that the model reliably distinguishes between normal and pneumonia cases. Out of 400 validation samples, 283 pneumonia cases are correctly identified, with only 11 misclassified as normal. For the normal class, 89 out of 106 samples are correctly predicted, with 17 being incorrectly labeled as pneumonia. These results indicate that the model achieves both high sensitivity (true positive rate for pneumonia) and strong specificity (true negative rate for normal), with a slightly higher misclassification rate for the normal class. This asymmetry is consistent with the dataset imbalance, where pneumonia samples dominate, but it does not significantly affect the overall accuracy. Additionally, Fig. 13

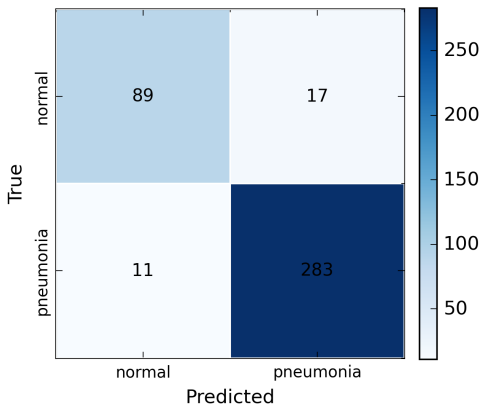


Fig. 12 Confusion matrix for *NQ-SVM* on PneumoniaMNIST ($m = 20$, seed $= 42$)

compares the confusion matrices of the proposed *NQ-SVM* with $m = 40$ landmarks and the full QSVM for binary classification of malignant and benign samples for the WDBC dataset. Despite relying on a low-rank Nyström approximation, the *NQ-*

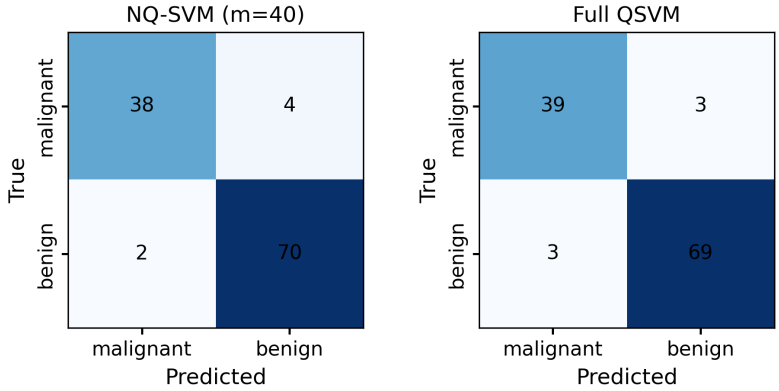


Fig. 13 Confusion matrix comparison between Nyström QSVM and full QSVM

SVM achieves classification performance that is highly comparable to that of the full QSVM, demonstrating the effectiveness of the approximation. In particular, our proposed model correctly classifies 70 benign samples, compared to 69 for the full QSVM, while maintaining a similar number of correct malignant predictions. The overall misclassification patterns of the two models are nearly identical, with only a one-sample difference in each off-diagonal entry. This indicates that the Nyström approximation preserves the essential decision boundaries learned by the full QSVM. These results demonstrate that the *NQ-SVM* offers an effective trade-off between computational efficiency and classification performance. With a moderate landmark

size, the *NQ-SVM* preserves the discriminative capability of the full QSVM while enabling more scalable and resource-efficient deployment.

Table 5 Comparison between *NQ-SVM* ($m = 20$) and a weak CNN model on PneumoniaMNIST (balanced split)

Model	**Accuracy**	**F1-score**	**AUROC**
NQ-SVM ($m = 20$)	0.8675	0.8921	0.9193
Simple CNN	0.6550	0.7915	0.6983

To further highlight the performance of *NQ-SVM*, we compare it against a simple convolutional neural network (CNN) trained for 10 epochs on the same balanced PneumoniaMNIST split (1200/400/400). The CNN achieved a final test accuracy of 65.5%, an F1-score of 0.7915, and an AUROC of 0.6983. In contrast, *NQ-SVM* with $m = 20$ landmarks reached a test accuracy of 86.8%, with substantially higher F1-score and AUROC values (see Table 5). Although the CNN learning curves indicate stable convergence, performance plateaus early and generalization remains weak, particularly on the test set, underscoring the limitations of shallow CNNs when trained on limited medical imaging data. By comparison, *NQ-SVM* effectively leverages the expressive power of quantum kernels combined with efficient low-rank approximation, yielding significantly improved discrimination between normal and pneumonia cases, as reflected in Fig. 12. Thus, while the CNN fails to surpass 70% AUROC, *NQ-SVM* achieves performance above 90%, showing superior generalization and class separability under limited training data.

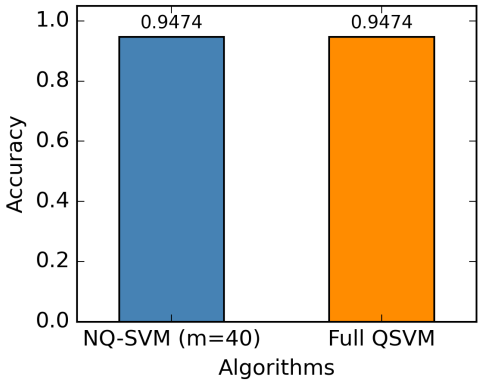


Fig. 14 *NQ-SVM* approximation vs. full QSVM models

Fig. 14 compares the performance of *NQ-SVM* with $m = 40$ landmarks against the full QSVM, which uses the complete kernel matrix without approximation. Both

methods achieve an accuracy of 94.74% on the WDBC dataset, indicating that the Nyström approximation with $m = 40$ captures the essential structure of the quantum kernel. Beyond this point, increasing the number of landmarks yields no measurable performance gains, as the decision boundary and generalization capability are already well approximated. This indicates that a moderately sized landmark set can match the performance of the full QSVM while significantly reducing computational cost, showing the efficiency and scalability advantages of *NQ-SVM*.

Table 6 Performance of *NQ-SVM* under different encoding strategies, circuit depths, and qubit counts.

Encoding	**Qubits**	**Layers**	**Train Acc. (%)**	**Test Acc. (%)**	**AUC**
Angle (RY)	2	1	89.3	87.0	0.927
Angle (RY)	2	2	91.4	88.6	0.934
Angle (RY)	2	3	91.0	88.2	0.932
Angle (RY)	3	1	92.5	90.3	0.941
Angle (RY)	3	2	94.8	92.8	0.952
Angle (RY)	3	3	94.1	92.1	0.949
Angle (RY)	4	1	94.2	92.1	0.947
Angle (RY)	**4**	**2**	**97.1**	**94.7**	**0.958**
Angle (RY)	4	3	96.8	94.3	0.955
Angle (RY)	6	1	93.0	91.0	0.943
Angle (RY)	6	2	95.2	93.0	0.950
Angle (RY)	6	3	94.8	92.8	0.948
Amplitude	2	1	85.2	82.7	0.914
Amplitude	2	2	87.0	84.2	0.920
Amplitude	2	3	87.5	84.5	0.921
Amplitude	4	1	90.3	87.5	0.931
Amplitude	4	2	92.8	89.4	0.938
Amplitude	4	3	93.5	90.0	0.941

Table 6 compares the performance of *NQ-SVM* for different quantum encoding strategies and different numbers of data-reuploading layers. The results show that angle encoding consistently outperforms amplitude encoding. For angle encoding, increasing the number of layers from one to two yields a noticeable improvement in both accuracy and AUC. The two-layer circuit achieves the best overall performance, with a test accuracy of 94.7% and an AUC of 0.958. In this setting, four qubits and a simple RY-based feature map provide sufficient expressive power to capture the structure of the WDBC dataset without overfitting. Although adding a third layer increases

circuit depth, it does not further improve performance and instead leads to a slight degradation, suggesting that deeper circuits introduce unnecessary complexity relative to the dataset size.

Amplitude encoding exhibits weaker performance across all tested depths. This outcome is expected in the present experimental setting, as amplitude encoding requires normalization and embedding of the full input vector into quantum state amplitudes, making the representation more sensitive to data variations. Moreover, preparing amplitude-encoded states is more demanding, and the limited number of available qubits restricts its ability to capture high-dimensional structure. Consequently, amplitude encoding remains inferior to angle encoding even with multiple reuploading layers. Overall, these results indicate that angle encoding with two data-reuploading layers and four qubits offers the best balance between expressiveness and stability. This configuration consistently achieves the strongest performance among the tested options and is therefore the most suitable choice for the *NQ-SVM* framework used in this study.

6 Conclusion

In this work, we proposed *NQ-SVM*, a scalable quantum kernel method that integrates the Nyström approximation with quantum feature maps to address the computational bottleneck of full QSVMs. By reducing kernel construction complexity from $O(n^2)$ to $O(nm)$, the method enables efficient training on larger datasets while preserving the expressive power of quantum kernels. Extensive simulation results confirm that *NQ-SVM* achieves classification accuracy comparable to full QSVMs, while substantially reducing the computational cost of kernel construction. Importantly, only a small fraction of landmark points ($m \ll n$) is sufficient to maintain strong performance, validating the scalability of the proposed approach. As a result, *NQ-SVM* offers a promising and scalable framework for quantum-enhanced classification. Future work will explore adaptive landmark selection strategies, robustness under realistic quantum noise, and deployment on near-term quantum hardware to further assess the practicality of this approach in real-world quantum machine learning applications.

Acknowledgements. This work was supported in part by the Fulton Schools of Engineering at Arizona State University, in part by the ARPA-H grant SP4701-23-C-0074, and in part by the Institute of Applied Artificial Intelligence and Robotics (IAAIR).

Declarations

Funding

Not applicable.

Conflict of interest

The authors declare that they have no competing interests.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Data availability

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Materials availability

Not applicable.

Code availability

The code used in this study is available from the corresponding author on reasonable request.

Clinical trial number

Not applicable.

References

[1] Rebentrost, P., Mohseni, M., Lloyd, S.: Quantum support vector machine for big data classification. Physical Review Letters **113**(13), 130503 (2014) https://doi.org/10.1103/PhysRevLett.113.130503

[2] Cervantes, J., Garcia-Lamont, F., Rodríguez-Mazahua, L., Lopez, A.: A comprehensive survey on support vector machine classification: Applications, challenges and trends. Neurocomputing **408**, 189–215 (2020) https://doi.org/10.1016/j.neucom.2019.10.118

[3] Grønlund, A., Kamma, L., Larsen, K.G.: Near-tight margin-based generalization bounds for support vector machines. In: Proceedings of the 37th International Conference on Machine Learning (ICML'20), p. 354. JMLR.org, ??? (2020)

[4] Akram-Ali-Hammouri, Z., Fernández-Delgado, M., Cernadas, E., Barro, S.: Fast support vector classification for large-scale problems. IEEE Transactions on Pattern Analysis and Machine Intelligence **44**(10, Part 1), 6184–6195 (2022) https://doi.org/10.1109/TPAMI.2021.3085969

[5] Djemai, S., Brahmi, B., Bibi, M.O.: A primal-dual method for svm training. Neurocomput. **211**(C), 34–40 (2016) https://doi.org/10.1016/j.neucom.2016.01.103

[6] Hsieh, C.-J., Si, S., Dhillon, I.S.: Communication-efficient distributed block minimization for nonlinear kernel machines. In: Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '17), pp. 245–254. Association for Computing Machinery, New York, NY, USA (2017). https://doi.org/10.1145/3097983.3098080 . https://doi.org/10.1145/3097983.3098080

[7] Kafai, M., Eshghi, K.: Croification: Accurate kernel classification with the efficiency of sparse linear svm. IEEE Transactions on Pattern Analysis and Machine Intelligence **41**(1), 34–48 (2019) https://doi.org/10.1109/TPAMI.2017.2785313

[8] Kaushik, C., McRae, A.D., Davenport, M., Muthukumar, V.: New equivalences between interpolation and svms: Kernels and structured features. SIAM Journal on Mathematics of Data Science **6**(3), 761–787 (2024) https://doi.org/10.1137/23M1568764

[9] Cipolla, S., Gondzio, J.: Training very large scale nonlinear SVMs using Alternating Direction Method of Multipliers coupled with the Hierarchically Semi-Separable kernel approximations (2021). https://arxiv.org/abs/2108.04167

[10] Du, K.-L., Jiang, B., Lu, J., Hua, J., Swamy, M.N.S.: Exploring kernel machines and support vector machines: Principles, techniques, and future directions. Mathematics **12**(24), 3935 (2024) https://doi.org/10.3390/math12243935

[11] Tanji, S., Vecchia, A.D., Glineur, F., Villa, S.: Snacks: a fast large-scale kernel svm solver. In: 2023 European Control Conference (ECC), pp. 1–6 (2023). https://doi.org/10.23919/ECC57647.2023.10178323

[12] Brahmi, B.: An efficient primal simplex method for solving large-scale support vector machines. Neurocomputing **599**, 128109 (2024) https://doi.org/10.1016/j.neucom.2024.128109

[13] Navia-Vázquez, A., Gutiérrez-González, D., Parrado-Hernández, E., Navarro-Abellán, J.J.: Distributed support vector machines. IEEE Transactions on Neural Networks **17**(4), 1091–1097 (2006) https://doi.org/10.1109/TNN.2006.875968

[14] Forero, P.A., Cano, A., Giannakis, G.B.: Consensus-based distributed linear support vector machines. In: Proceedings of the 9th ACM/IEEE International Conference on Information Processing in Sensor Networks. IPSN '10, pp. 35–46. Association for Computing Machinery, New York, NY, USA (2010). https://doi.org/10.1145/1791212.1791218 . https://doi.org/10.1145/1791212.1791218

[15] Craen, A.V., Breyer, M., Pflüger, D.: Plssvm: A (multi-)gpgpu-accelerated least squares support vector machine. arXiv preprint (2022) 2202.12674

[16] Nielsen, M.A., Chuang, I.L.: Quantum Computation and Quantum Information, 10th anniversary edition edn. Cambridge University Press, Cambridge, UK

(2010)

[17] Montanaro, A.: Quantum algorithms: an overview. npj Quantum Information **2**(15023) (2016) https://doi.org/10.1038/npjqi.2015.23

[18] Preskill, J.: Quantum computing in the nisq era and beyond. Quantum **2**, 79 (2018) https://doi.org/10.22331/q-2018-08-06-79

[19] Burgholzer, L., Raymond, R., Wille, R.: Verifying results of the ibm qiskit quantum circuit compilation flow. *2020 IEEE International Conference on Quantum Computing and Engineering (QCE)*, 356–365 (2020)

[20] Harrow, A.W., Hassidim, A., Lloyd, S.: Quantum algorithm for linear systems of equations. Phys. Rev. Lett. **103**, 150502 (2009) https://doi.org/10.1103/PhysRevLett.103.150502

[21] Low, G.H., Su, Y.: Quantum eigenvalue processing. In: 2024 IEEE 65th Annual Symposium on Foundations of Computer Science (FOCS), pp. 1051–1062. IEEE, ??? (2024). https://doi.org/10.1109/focs61266.2024.00070 . http://dx.doi.org/10.1109/FOCS61266.2024.00070

[22] Arute, F., Arya, K., Babbush, R., Bacon, D., Bardin, J.C., Barends, R., Biswas, R., Boixo, S., Brandao, F.G.S.L., Buell, D.A., al.: Quantum supremacy using a programmable superconducting processor. *Nature* **574**, 505–510 (2019)

[23] Zhong, H.-S., Wang, H., Deng, Y.-H., Chen, M.-C., Peng, L.-C., Luo, Y.-H., Qin, J., Wu, D., Ding, X., Hu, Y., Hu, P., Yang, X.-Y., Zhang, W.-J., Li, H., Li, Y., Jiang, X., Gan, L., Yang, G., You, L., Wang, Z., Li, L., Liu, N.-L., Lu, C.-Y., Pan, J.-W.: Quantum computational advantage using photons. *Science* **370**(6523), 1460–1463 (2020) https://doi.org/10.1126/science.abe8770

[24] Wu, Y., Bao, W.-S., Cao, S., Chen, F., Chen, M.-C., Chen, X., al.: Strong quantum computational advantage using a superconducting quantum processor. *Phys. Rev. Lett.* **127**, 180501 (2021) https://doi.org/10.1103/PhysRevLett.127.180501

[25] Li, X., Wang, Y., Huang, Z., Liu, H., *et al.*: Faster quantum subroutine for matrix chain multiplication. Scientific Reports **15**(1), 13092 (2025) https://doi.org/10.1038/s41598-025-13092-2

[26] Biamonte, J., Wittek, P., Pancotti, N., Rebentrost, P., Wiebe, N., Lloyd, S.: Quantum machine learning. Nature **549**(7671), 195–202 (2017) https://doi.org/10.1038/nature23474

[27] Schuld, M., Petruccione, F.: Supervised Learning with Quantum Computers. Quantum Science and Technology. Springer, ??? (2018). https://doi.org/10.1007/978-3-319-96424-9 . https://link.springer.com/book/10.1007/978-3-319-96424-9

[28] Schuld, M., Killoran, N.: Quantum machine learning in feature hilbert spaces. Phys. Rev. Lett. **122**, 040504 (2019) https://doi.org/10.1103/PhysRevLett.122.040504

[29] Goto, T., Tran, Q.H., Nakajima, K.: Universal approximation property of quantum machine learning models in quantum-enhanced feature spaces. Physical Review Letters **127**(9) (2021) https://doi.org/10.1103/physrevlett.127.090506

[30] Lloyd, S., Mohseni, M., Rebentrost, P.: Quantum principal component analysis. Nature Physics **10**(9), 631–633 (2014) https://doi.org/10.1038/nphys3029

[31] Li, L.H., Zhang, D.-B., Wang, Z.D.: Quantum kernels with gaussian state encoding for machine learning. Physics Letters A **436**, 128088 (2022) https://doi.org/10.1016/j.physleta.2022.128088

[32] Park, J.-E., Quanz, B., Wood, S., Higgins, H., Harishankar, R.: Practical application improvement to Quantum SVM: theory to practice (2020). https://arxiv.org/abs/2012.07725

[33] Chang, D.T.: Parameterized Quantum Circuits with Quantum Kernels for Machine Learning: A Hybrid Quantum-Classical Approach (2022). https://arxiv.org/abs/2209.14449

[34] K, N., Singh, D.P., Kavitha, V.R., Ravi, S., Giri, P., Ali Samhan, A.A.: Investigating hybrid algorithms for quantum computing in machine learning: Improving data processing speed and handling complex problems. In: 2025 International Conference on Networks and Cryptology (NETCRYPT), pp. 1438–1443 (2025). https://doi.org/10.1109/NETCRYPT65877.2025.11102526

[35] Long, C., Huang, M., Ye, X., Futamura, Y., Sakurai, T.: Hybrid quantum-classical-quantum convolutional neural networks. Scientific Reports **15**(1), 31780 (2025) https://doi.org/10.1038/s41598-025-13417-1

[36] Krunic, Z., Flöther, F.F., Seegan, G., Earnest-Noble, N.D., Shehab, O.: Quantum Kernels for Real-World Predictions Based on Electronic Health Records. IEEE Trans. Quantum Eng. **3**, 1–11 (2022) https://doi.org/10.1109/TQE.2022.3176806 arXiv:2112.06211 [quant-ph]

[37] Miroszewski, A., Mielczarek, J., Czelusta, G., Szczepanek, F., Grabowski, B., Le Saux, B., Nalepa, J.: Detecting clouds in multispectral satellite images using quantum-kernel support vector machines. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing **16**, 7601–7613 (2023) https://doi.org/10.1109/JSTARS.2023.3304122

[38] Alvarez-Estevez, D.: Benchmarking quantum machine learning kernel training for classification tasks. IEEE Transactions on Quantum Engineering **6**, 1–15 (2025) https://doi.org/10.1109/TQE.2025.3541882

[39] Havlíček, V., Córcoles, A.D., Temme, K., Harrow, A.W., Kandala, A., Chow, J.M., Gambetta, J.M.: Supervised learning with quantum-enhanced feature spaces. Nature **567**(7747), 209–212 (2019) https://doi.org/10.1038/s41586-019-0980-2

[40] Yin, Z., Agresti, I., Felice, G., Brown, D., Toumi, A., Pentangelo, C., Piacentini, S., Crespi, A., Ceccarelli, F., Osellame, R., Coecke, B., Walther, P.: Experimental quantum-enhanced kernel-based machine learning on a photonic processor. Nature Photonics **19**(9), 1020–1027 (2025) https://doi.org/10.1038/s41566-025-01682-5

[41] Schuld, M., Sweke, R., Meyer, F.: Effect of data encoding on the expressive power of quantum machine learning models. Physical Review A **103**(3), 032430 (2021)

[42] Huang, H.-Y., Broughton, M., Mohseni, M., Babbush, R., Boixo, S., Neven, H., McClean, J.R.: Power of data in quantum machine learning. Nature Communications **12**, 2631 (2021)

[43] Heyraud, V., Li, Z., Denis, Z., Le Boité, A., Ciuti, C.: Noisy quantum kernel machines. Phys. Rev. A **106**, 052421 (2022) https://doi.org/10.1103/PhysRevA.106.052421

[44] Williams, C.K.I., Seeger, M.: Using the nyström method to speed up kernel machines. In: Proceedings of the 14th International Conference on Neural Information Processing Systems. NIPS'00, pp. 661–667. MIT Press, Cambridge, MA, USA (2000)

[45] Kumar, S., Mohri, M., Talwalkar, A.: Sampling methods for the nyström method. J. Mach. Learn. Res. **13**(null), 981–1006 (2012)

[46] Yang, T., Li, Y.-F., Mahdavi, M., Jin, R., Zhou, Z.-H.: Nyström method vs random fourier features: a theoretical and empirical comparison. In: Proceedings of the 26th International Conference on Neural Information Processing Systems - Volume 1. NIPS'12, pp. 476–484. Curran Associates Inc., Red Hook, NY, USA (2012)

[47] Li, Z., Yang, T., Zhang, L., Jin, R.: Fast and accurate refined nyström based kernel svm. In: Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence. AAAI'16, pp. 1830–1836. AAAI Press, ??? (2016)

[48] Si, S., Hsieh, C.-J., Dhillon, I.S.: Memory efficient kernel approximation. J. Mach. Learn. Res. **18**(1), 682–713 (2017)

[49] Xia, J.: Making the Nyström method highly accurate for low-rank approximations (2023). https://arxiv.org/abs/2307.05785

[50] Münch, M., Bohnsack, K.S., Schleif, F., Villmann, T.: Data-distribution-informed

nyström approximation for structured data using vector quantization-based landmark determination. Neurocomputing **596**, 128100 (2024) https://doi.org/10.1016/j.neucom.2024.128100

[51] Martín-Baos, J.A., García-Ródenas, R., Rodriguez-Benitez, L., Bierlaire, M.: Scalable kernel logistic regression with nyström approximation: Theoretical analysis and application to discrete choice modelling. Neurocomput. **617**(C) (2025) https://doi.org/10.1016/j.neucom.2024.128975

[52] Rebentrost, P., Mohseni, M., Lloyd, S.: Quantum support vector machine for big data classification. Physical Review Letters **113**(13) (2014)

[53] Havlíček, V., Córcoles, A.D., Temme, K., Harrow, A.W., Kandala, A., Chow, J.M., Gambetta, J.M.: Supervised learning with quantum-enhanced feature spaces. Nature **567**(7747), 209–212 (2019)

[54] Schuld, M., Killoran, N.: Quantum machine learning in feature hilbert spaces. Physical Review Letters **122**(4), 040504 (2019)

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

[56] Blank, C., Park, D.K., Ryu, J., Petruccione, F.: Quantum classifier with tailored quantum kernel. npj Quantum Information **6**(1), 41 (2020)

[57] Innan, N., Khan, M.A.Z., Panda, B., Bennai, M.: Enhancing quantum support vector machines through variational kernel training. Quantum Information Processing **22**(10) (2023) https://doi.org/10.1007/s11128-023-04138-3

[58] Suzuki, T., Hasebe, T., Miyazaki, T.: Quantum support vector machines for classification and regression on a trapped-ion quantum computer. Quantum Machine Intelligence **6**(1) (2024) https://doi.org/10.1007/s42484-024-00165-0

Algorithm 2 *NQ-SVM* Training

Require: Training, testing data $(\mathcal{X}_{\text{tr}}, y_{\text{tr}}), (\mathcal{X}_{\text{te}}, y_{\text{te}})$, number of qubits q, layers L, scale α, shots S, number of landmarks m, ridge parameter λ, SVM regularization C, landmark strategy.

Ensure: Landmark set Z, Nyström eigenpairs (Σ, U) of W_{reg}, trained SVM classifier $f(\cdot)$.

1: Construct a standardization transform $\mathcal{S}$ from $\mathcal{X}_{\text{tr}}$, set $\mathcal{X}_k \leftarrow \mathcal{S}(\mathcal{X}_k)$ for $k \in \{\text{tr}, \text{te}\}$.

2: Compute a PCA projection $\mathcal{P}$ with q components using $\mathcal{X}_{\text{tr}}$, set $\mathcal{X}_k \leftarrow \mathcal{P}(\mathcal{X}_k)$ for $k \in \{\text{tr}, \text{te}\}$.

3: Compute a feature-wise min–max map $\mathcal{M}_{[-\pi,\pi]}$ on $\mathcal{X}_{\text{tr}}$, set $\mathcal{X}_k^{\text{ang}} \leftarrow \mathcal{M}_{[-\pi,\pi]}(\mathcal{X}_k)$ for $k \in \{\text{tr}, \text{te}\}$.

4: Select index set $\mathcal{I}$ from $\mathcal{X}_{\text{tr}}$ using landmark selection strategy (e.g., Alg. 1):

$$Z \leftarrow \textsc{SelectLandmarks}\big(\mathcal{X}_{\text{tr}}^{\text{ang}}, m, \mathsf{strategy}, \text{seed}\big)$$

5: Construct a q-qubit state-preparation $\phi(\cdot)$ with parameters $(L, \alpha, \text{entangling}, S)$.

6: Create quantum kernel matrices:
Compute landmark Gram matrix $W \in \mathbb{R}^{m \times m}$:

$$W_{i,j} = \big|\langle \phi(\mathbf{z}_i) \mid \phi(\mathbf{z}_j)\rangle\big|^2.$$

Compute cross-kernel matrices C_k for each $k \in \{\text{tr}, \text{te}\}$:

$$C_k(i,j) = \big|\langle \phi(\mathbf{x}_i^{(k)}) \mid \phi(\mathbf{z}_j)\rangle\big|^2.$$

7: Create the regularized landmark kernel W_{reg}:

$$W_{\text{reg}} \leftarrow \tfrac{1}{2}(W + W^\top) + \lambda I_m.$$

Compute the eigendecomposition: $W_{\text{reg}} = U \Sigma U^\top$ with $\Sigma = \text{diag}(\sigma_1, \ldots, \sigma_m)$.
Construct the inverse of the landmark Gram matrix:

$$W_{\text{reg}}^{-1/2} = U\, \Sigma^{-1/2}\, U^\top,$$

where $\Sigma^{-1/2} = \text{diag}(\sigma_1^{-1/2}, \ldots, \sigma_m^{-1/2})$.

8: Create Nyström features: $\widetilde{\Phi}_k = C_k W_{\text{reg}}^{-1/2}$, $k \in \{\text{tr}, \text{te}\}$.
Standardize in Nyström space: $\widetilde{\Phi}(\mathbf{x}) \leftarrow \mathcal{S}_\Phi(\widetilde{\Phi}(\mathbf{x}))$.

9: Training SVM: fit LinearSVC on $(\widetilde{\Phi}_{\text{tr}}, y_{\text{tr}})$ using the regularization C.

10: **return** $\{Z, U, \Sigma, \text{scaler}, \text{PCA}, f(\cdot)\}$.