Quantum Kernel Methods for Breast Cancer Classification using QSVM and Qiskit

Kajal Choudhary

Department of Physics, IIT Kharagpur

kajal.choudhary@kgpian.iitkgp.ac.in

FTP 8.0 – Advanced Quantum Computing Track

Abstract

This project explores quantum kernel methods for binary classification using Qiskit Machine Learning. The Breast Cancer dataset is used as a benchmark to demonstrate the performance of quantum kernel-based support vector machines (QSVMs) compared to classical models. Due to the Noisy Intermediate-Scale Quantum (NISQ) nature of current quantum hardware, we restrict our study to only 4 normalized features. We also benchmark the quantum circuits to evaluate depth, width, and gate complexity using Qiskit transpilation tools. The PauliFeatureMap-based QSVM achieved a 96.4% test accuracy, comparable to the best classical model, while requiring significantly fewer features, highlighting the potential of quantum-enhanced learning.

1 Introduction

Machine learning has become a powerful tool in healthcare, enabling early diagnosis and classification of diseases based on patient data. Among various models, Support Vector Machines (SVMs) are known for their robustness in binary classification. However, SVM performance depends heavily on the kernel used to map data into a higher-dimensional feature space where it becomes linearly separable.

In recent years, quantum computing has emerged as a promising approach to enhance machine learning techniques by utilising principles of quantum mechanics, such as superposition and entanglement. Quantum-enhanced kernels offer the possibility of generating richer feature maps, potentially capturing complex data patterns that classical kernels may overlook.

This work focuses on Quantum Support Vector Machines (QSVMs) using Qiskit, an open-source quantum SDK. We apply QSVMs to a biomedical dataset—the Breast Cancer dataset from Scikit-learn—where the goal is to classify tumors as malignant or benign. Our analysis includes model performance, circuit complexity, and comparison with traditional models such as SVMs with RBF kernels, logistic regression, and decision trees.

2 Dataset and Preprocessing

The dataset used in this study is the Breast Cancer Wisconsin (Diagnostic) dataset, which is publicly available through Scikit-learn. It contains 569 instances, each with 30 numerical features derived from digitized images of breast mass. The target variable is binary: malignant (label 0) or benign (label 1).

Given the limitations of current quantum devices (specifically, their limited number of qubits and high noise), it is not feasible to process all 30 features. Therefore, we select the first 4 features after applying standard normalization. This dimensionality reduction ensures that the quantum circuits remain shallow and implementable on near-term quantum simulators or hardware.

The data is split into training and testing sets with a 70:30 ratio. We standardize the features using Scikit-learn's 'StandardScaler' to ensure that the input to quantum circuits lies within the range suitable for parameterized rotation gates.

3 Quantum Kernel Methods

3.1 Theoretical Background

In classical SVMs, the kernel function defines the similarity between two data points in feature space. Common kernels include linear, polynomial, and radial basis function (RBF). In QSVMs, we replace the classical kernel with a quantum kernel computed from the fidelity between quantum states that encode the data.

Let $\phi(x)$ be a feature map that encodes classical data x into a quantum state $|\phi(x)\rangle$. Then the quantum kernel is defined as:

$$K(x, x') = |\langle\phi(x)|\phi(x')\rangle|^2$$

This kernel captures the similarity between data points using quantum mechanics, allowing access to high-dimensional feature spaces that are hard to simulate classically.

3.2 Feature Maps Used

We experimented with two types of parameterized quantum circuits for feature mapping:

- **ZZFeatureMap:** Applies parameterized rotations around the Z-axis followed by entangling ZZ interactions between qubits. This feature map is expressive and captures interactions among features.
- **PauliFeatureMap:** Uses a mix of Pauli operators (X, Y, Z) applied to each feature, with entanglement between qubits. The inclusion of all three Pauli matrices increases expressivity.

Each feature map was repeated twice to enhance expressiveness while keeping circuit depth manageable.

3.3 Quantum Kernel Computation

We use Qiskit's `FidelityQuantumKernel`, which leverages statevector simulation to compute the fidelity between encoded quantum states. Due to limitations in accessing Aer backends, transpilation was used to benchmark circuits on a simulated basis.

4 Classical Baselines

To evaluate the relative performance of QSVMs, we compare them with traditional machine learning models trained on the same 4-feature dataset:

- **Support Vector Machine (RBF kernel):** A powerful baseline model with non-linear capabilities.
- **Logistic Regression:** A simple linear classifier used in many medical applications.
- **Decision Tree:** Interpretable, rule-based classifier. We limit tree depth to 5 to prevent overfitting.
- **Random Forest:** An ensemble of 100 decision trees with maximum depth 5.

All models were trained using Scikit-learn with default hyperparameters, except for depth constraints.

5 Results and Evaluation

5.1 Accuracy Comparison

Model accuracies on the test set are presented in Table 1. We observe that the PauliFeatureMap-based QSVM achieves 96.4% accuracy, on par with the classical RBF SVM. This is significant given the QSVM uses only 4 features.

Table 1: Test Accuracy Comparison

Model	**Accuracy**
QSVM (ZZFeatureMap)	0.928
QSVM (PauliFeatureMap)	0.964
SVM (RBF)	0.964
Logistic Regression	0.953
Decision Tree	0.912
Random Forest	0.953

5.2 Quantum Circuit Benchmarking

We evaluate circuit depth, width (number of qubits), and total gate count using Qiskit transpiler. Results are summarized below:

Table 2: Quantum Circuit Metrics

Feature Map	**Depth**	**Width**	**Gate Count**
ZZFeatureMap	13	4	36
PauliFeatureMap	15	4	48

These values indicate that QSVMs with 4-qubit circuits are feasible for simulation and shallow enough for potential NISQ execution.

6 Discussion

The results of this study demonstrate that quantum kernel methods are competitive with classical models even when using a very limited number of features. The QSVM achieved similar performance to SVM with RBF kernel, which is known for its powerful non-linear mapping.

While this is a small-scale experiment constrained by the current state of quantum hardware, it points to the future of quantum-enhanced machine learning. Once quantum devices scale to more qubits and lower error rates, we can envision using more features and deeper circuits.

Another benefit is that quantum kernels are data-driven: they don't require hand-engineered kernels or extensive hyperparameter tuning, making them potentially more robust across domains.

7 Conclusion

This work shows that QSVMs, when applied with expressive quantum feature maps, can perform competitively with classical machine learning models. Using only 4 normalized features, we reached up to 96.4% classification accuracy. Quantum circuit benchmarking confirms that such models can be realistically implemented on simulators and potentially on small NISQ devices.

In the future, exploring hybrid quantum-classical models and variational approaches may yield even better results. With the growing maturity of Qiskit and access to real quantum hardware, quantum machine learning has the potential to tackle problems once thought out of reach.

Acknowledgements

This work was done under the guidance of Prof. Shiva Pokhrel as part of Deakin University's FTP 8.0 - Advanced Quantum Computing Track. Special thanks to the Qiskit team for open-source tools and documentation.

References

- Qiskit Machine Learning Tutorials: `https://qiskit.org/learn`
- IBM Quantum Research: `https://www.ibm.com/quantum/research`
- Qiskit BenchPress: `https://github.com/qiskit/benchpress`
- Schuld, M. et al. "Supervised Learning with Quantum Computers." Springer, 2018.
- arXiv:2409.08844 - Quantum Kernel Learning Review