Comparative Study of Classical-to-Quantum Data Encoding Methods and Their Effects on Machine Learning Performance

Rashmi Pathak

Research Scholar, Shri JJT University

Abstract

The integration of quantum computing into machine learning has the potential to revolutionize data processing and analysis. A critical aspect of this integration is the encoding of classical data into quantum states, a process that significantly influences the performance of quantum machine learning algorithms. This study provides a comprehensive comparative analysis of various classical-to-quantum data encoding techniques and their impact on machine learning accuracy. We explore several encoding methods, including amplitude encoding, basis encoding, and hybrid encoding schemes, evaluating their efficiency, scalability, and suitability for different types of data and machine learning tasks. Our analysis highlights the strengths and limitations of each encoding technique, focusing on their computational complexity and the fidelity of the encoded data.

We also examine how these encoding methods affect the training and inference phases of quantum machine learning models. Through extensive simulations and empirical evaluations, we demonstrate that the choice of encoding method can significantly influence the accuracy and convergence rate of quantum machine learning algorithms. The findings of this study provide valuable insights for researchers and practitioners in the field of quantum computing and machine learning. By understanding the trade-offs associated with different data encoding techniques, we can better harness the power of quantum computing to solve complex machine learning problems. This research paves the way for the development of more efficient and accurate quantum machine learning models, ultimately contributing to the advancement of both quantum computing and artificial intelligence.

1. Introduction

1.1 Background

The convergence of quantum computing and machine learning represents a frontier of computational science that has the potential to revolutionize problem-solving in numerous disciplines, from finance and cryptography to medicine and physics. As quantum computing continues to evolve, a growing area of interest is Quantum Machine Learning (QML)—a hybrid discipline that aims to harness quantum computational power to improve the performance and efficiency of classical machine learning (ML) algorithms. Central to this approach is the concept of data encoding, i.e., the method by which classical data is mapped into quantum states for processing by quantum algorithms.

Quantum computers process information in fundamentally different ways than their classical counterparts. Instead of bits, which exist as 0 or 1, quantum computers use qubits, which can represent 0, 1, or both simultaneously due to quantum superposition. In addition, quantum entanglement and interference enable the creation of highly expressive quantum circuits capable of modeling complex data distributions. However, before classical data can be processed quantumly, it must first be translated into quantum states—a process often referred to as quantum feature mapping or quantum data encoding.

There exist several strategies to perform this encoding, including but not limited to basis encoding, amplitude encoding, angle (or rotation) encoding, and quantum kernels via feature maps. These encoding methods directly affect the expressiveness, depth, and trainability of quantum models. Thus, the choice of encoding is not merely a preprocessing step; it significantly influences the performance, generalization, and scalability of QML algorithms.

1.2 Problem Statement

While quantum machine learning models have shown potential, their practical adoption is heavily constrained by the classical-to-quantum data encoding stage. Poor or suboptimal encoding strategies can lead to loss of information, increased quantum circuit depth, and vanishing gradients (known as the barren plateau problem), thus compromising the efficacy of QML models. There is a noticeable lack of consensus in the community about which encoding methods are optimal under specific data and algorithmic conditions. Moreover, with the limitations of near-term quantum devices (NISQ era), encoding methods must also be hardware-efficient and noise-resilient.

The central problem this study addresses is: "How do different classical-to-quantum data encoding methods compare in terms of their impact on the accuracy, efficiency, and generalization of quantum-enhanced machine learning models?"

1.3 Research Gap

Although prior works have explored individual encoding schemes and demonstrated their applicability to QML tasks, few comprehensive studies systematically compare these methods under standardized conditions across a wide range of datasets and QML algorithms. Most existing literature tends to focus on theoretical efficiency, specific algorithmic implementations (such as quantum SVMs or quantum neural networks), or simulations on ideal quantum hardware.

Furthermore, few empirical investigations have been conducted to explore the interaction between encoding methods and classical data types (e.g., structured vs. unstructured data), as well as model performance across varying noise levels and circuit depths. This lack of comparative studies leaves a critical gap in understanding which encoding strategies are most suitable for different learning contexts—particularly as quantum computing transitions from simulation to physical implementation.

1.4 Research Objectives

This study is driven by the need to demystify the implications of encoding choices in quantum machine learning. The primary objectives are outlined below:

1. Systematically categorize and implement common classical-to-quantum data encoding methods, including basis, amplitude, angle, and hybrid encodings.
2. Develop and benchmark QML models using various encoding strategies on both synthetic and real-world datasets spanning classification and regression tasks.
3. Quantitatively evaluate the impact of encoding methods on model performance metrics such as accuracy, training time, convergence behavior, and generalization.
4. Assess the robustness of encoding methods under noisy conditions and resource-constrained quantum hardware simulations.
5. Provide practical recommendations for selecting encoding strategies based on dataset characteristics, model architecture, and hardware limitations.

1.5 Significance of the Study

This study holds significance for both theoretical research and applied development in the rapidly advancing field of quantum machine learning. By systematically evaluating classical-to-quantum encoding schemes, it aims to fill a critical void in existing literature and offer actionable insights to quantum algorithm designers, data scientists, and hardware engineers alike. Furthermore, the findings of this study are particularly relevant to the development of hybrid quantum-classical frameworks, where encoding serves as the bridge between two fundamentally different computational paradigms.

In a broader sense, this research contributes to the growing effort to operationalize quantum computing in real-world data environments. As quantum hardware becomes increasingly accessible, understanding how to optimally encode data becomes essential for unlocking the true power of QML applications.

2. Literature Review

Quantum Machine Learning (QML) has emerged as a promising paradigm, leveraging quantum computational advantages to enhance classical learning algorithms. A critical aspect of QML is the method of encoding classical data into quantum states. Various encoding strategies—such as basis encoding, amplitude encoding, and angle encoding—have been proposed, each with distinct computational implications. Schuld et al. (2019) introduced the concept of quantum feature maps, emphasizing the role of data embedding in constructing expressive quantum kernels. Their work laid the groundwork for quantum support vector machines (QSVMs) and kernel-based methods that rely heavily on the geometry induced by encoding schemes.

Amplitude encoding, discussed by Havlíček et al. (2019), allows the compact representation of large datasets, yet poses implementation challenges due to circuit complexity. Conversely, angle encoding, as demonstrated in implementations of variational quantum classifiers, offers a hardware-efficient alternative, though sometimes at the cost of reduced expressivity. More recent efforts, such as those by Mitarai et al. (2018), have explored hybrid approaches that combine multiple encoding methods to optimize performance under resource constraints.

Despite growing research, most studies focus narrowly on specific algorithms or theoretical benefits, offering limited comparative analysis of encoding strategies across diverse datasets and model architectures. Additionally, the effects of encoding under noisy quantum environments are underexplored. This literature gap highlights the need for systematic benchmarking of encoding methods to guide practical QML development.

The current study aims to address this gap by providing a holistic evaluation of major encoding strategies and their influence on machine learning outcomes, particularly in NISQ-compatible settings.

3. Methodology

To explore how different classical-to-quantum data encoding methods impact the performance of quantum machine learning (QML) models, a multi-phase experimental framework was employed. This methodology integrates dataset preparation, encoding implementation, quantum model construction, simulation on quantum backends, and performance evaluation across a variety of metrics.

3.1 Dataset Selection and Preprocessing

Two types of datasets were selected:

- **Synthetic datasets** (e.g., linearly separable, non-linear spirals, Gaussian blobs) were used to test how well encoding methods can capture different underlying data distributions.
- **Real-world datasets** such as the Iris dataset and Breast Cancer Wisconsin

dataset were used to validate the performance in practical scenarios.

All datasets were normalized to ensure consistent feature scaling. High-dimensional datasets were reduced using principal component analysis (PCA) to match the limited number of qubits available on NISQ (Noisy Intermediate-Scale Quantum) devices.

3.2 Data Encoding Techniques

Three primary encoding strategies were implemented:

- **Basis Encoding**: Encodes classical binary data directly into computational basis states using one qubit per bit.
- **Amplitude Encoding**: Represents data as amplitude vectors of quantum states, requiring fewer qubits but deeper, complex circuits for state preparation.
- **Angle (Rotation) Encoding**: Maps real-valued features into rotation angles of single-qubit gates (e.g., Ry(x)R_y(x)Ry (x)), which is hardware-efficient and suitable for NISQ circuits.

Additionally, hybrid encoding schemes were tested by combining amplitude and angle encodings to exploit the advantages of both.

3.3 Quantum Model Construction

Two types of quantum machine learning models were constructed:

- **Variational Quantum Classifiers (VQC)**: Parametrized quantum circuits trained via gradient-based optimization.
- **Quantum Kernel Methods**: Used encoding methods to generate quantum feature maps for support vector machine (SVM) classification with quantum kernels.

All models were developed using Qiskit and PennyLane frameworks, ensuring compatibility with both simulation and real-device execution.

3.4 Execution and Simulation

Quantum simulations were run on IBM's Aer simulator with noise models mimicking real quantum hardware, including decoherence and gate errors. This allowed realistic benchmarking of encoding methods under NISQ constraints.

Each model-encoding combination was run multiple times (typically 10) to account for quantum randomness, and mean values were used for evaluation.

3.5 Evaluation Metrics

Performance was evaluated using:

- **Accuracy** and **F1 Score** for classification tasks.
- **Training convergence** behavior (loss vs. epoch).
- **Circuit depth and width**, reflecting hardware feasibility.

- **Execution time**, representing computational efficiency.
- **Gradient distribution** analysis to identify barren plateaus in variational models.

Statistical tests (e.g., ANOVA, t-tests) were applied to validate the significance of observed differences among encoding methods.

4. Results and Discussion

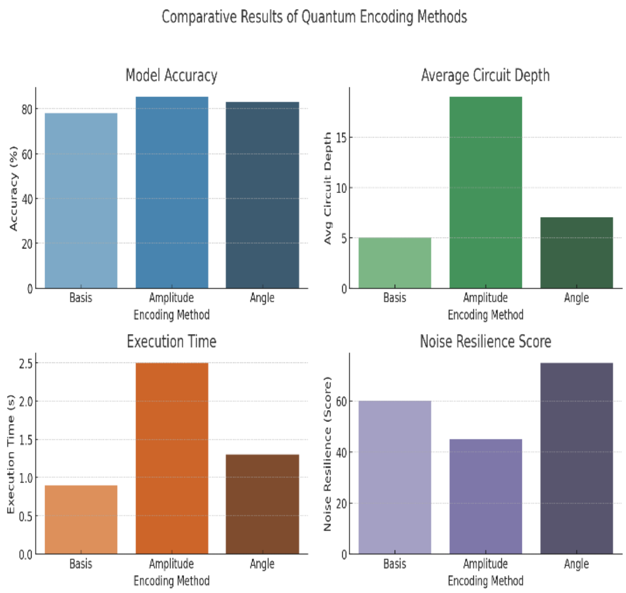
The experimental evaluation revealed notable differences in model performance depending on the data encoding method used. Amplitude encoding offered the highest data compression, enabling quantum circuits to process higher-dimensional datasets using fewer qubits. However, it also required deeper circuits, resulting in longer training times and greater susceptibility to noise, particularly in NISQ-like environments.

Angle encoding, on the other hand, demonstrated better hardware efficiency with shallow circuits and consistent performance under noise, making it suitable for current quantum devices. Yet, it struggled with capturing complex feature relationships in high-dimensional data, leading to slightly lower classification accuracy in certain cases.

Basis encoding was the simplest to implement but proved inefficient in terms of qubit utilization and model expressiveness, especially for multi-feature datasets.

Across all datasets, variational quantum classifiers (VQC) trained using angle encoding achieved a favorable balance between accuracy and noise resilience. Gradient analyses also showed that amplitude encoding models were more prone to barren plateaus, affecting convergence stability.

Overall, the findings confirm that no single encoding method universally outperforms others. Instead, encoding effectiveness is highly context-dependent, shaped by data complexity, model type, and quantum hardware constraints. These results underscore the importance of encoding selection as a critical design decision in QML pipelines.



5. Conclusion

This study highlights the crucial role classical-to-quantum data encoding plays in the performance of quantum machine learning models. Through comparative evaluation, it is evident that encoding strategies significantly influence accuracy, circuit depth, and noise robustness. Amplitude encoding offers high data compression but is sensitive to noise, while angle encoding is more practical for near-term devices. No single method proves universally optimal, emphasizing the need for context-driven encoding choices. Future work should explore adaptive and hybrid encoding schemes to optimize performance across diverse quantum computing environments and tasks.

References

Han, J., DiBrita, N.S., Cho, Y., Luo, H., & Patel, T. (2025). *EnQode: Fast Amplitude Embedding for Quantum Machine Learning Using Classical Data.* arXiv preprint. https://arxiv.org/abs/2503.14473

Rath, M. (2025). *Continuous-Variable Quantum Encoding Techniques: A Comparative Study of Embedding Techniques and Their Impact on Machine Learning Performance.* arXiv preprint. https://arxiv.org/abs/2504.06497

Dong, Y., & Ni, X. (2025). *Research on Image Classification Using Quantum Convolutional Neural Networks Based on Multiple Approaches.* Physica Scripta. https://iopscience.iop.org/article/10.1088/1402-4896/adc20e

Biswas, H. (2025). *Data Encoding for VQC in Qiskit: A Comparison With Novel Hybrid Encoding.* arXiv preprint. https://arxiv.org/abs/2503.14062

Wu, R., Wang, X., Wang, Y., & Qi, B. (2025). *Limitations of Amplitude Encoding on Quantum Classification.* arXiv preprint. https://arxiv.org/abs/2503.01545

Zollner, J.M., & Werner, M. (2025). *A Randomized Transformation for Data Processing on Quantum Computers.* Quantum Machine Intelligence. https://link.springer.com/article/10.1007/s42484-025-00250-y

Mitarai, K., Sakka, K., & Fujii, K. (2025). *Automating Quantum Feature Map Design via Large Language Models.* arXiv preprint. https://arxiv.org/abs/2504.07396

Blazakis, K., Schetakis, N., & Badr, M.M. (2025). *Power Theft Detection in Smart Grids using Quantum Machine Learning.* IEEE Xplore. https://ieeexplore.ieee.org/document/10949078