Quantum Models vs Classical Models on IEEE-CIS, SECOM, and NSL-KDD Datasets

Overview

This study explores the use of **Quantum Support Vector Machines (QSVM)** for classification tasks on three real-world datasets: **IEEE-CIS Fraud Detection**, **SECOM Manufacturing Process Data**, and **NSL-KDD Intrusion Detection**. The performance of QSVMs is compared to that of traditional **Classical SVMs** using a Radial Basis Function (RBF) kernel.

---

1. IEEE-CIS Fraud Detection Dataset

Dataset Description

- Total Samples Used: 1,000
- Features Selected: `TransactionAmt`, `card1`, `C1`, `C2`
- Target: `isFraud` (binary classification)

Preprocessing

- Features were selected based on correlation analysis.
- Data was standardized and split into training and validation sets (80/20).

Quantum Model

- **QSVM** using a custom **ZZ-feature map** encoded into 6 qubits.
- The quantum kernel was computed using state fidelity via Pennylane.
- Model trained using `SVC(kernel=kernel_matrix)` from `sklearn`.

Classical Model

- **SVM with RBF kernel** from `sklearn`.
- Balanced class weights and stratified splitting for handling class imbalance.

Results

Model	Accuracy (Validation)	Accuracy (Test)
QSVM	96.0%	96.0%
Classical SVM	73.5%	75.0%

> **Insight**: QSVM significantly outperformed the classical model, indicating the potential of quantum kernels to capture complex fraud-related patterns.

---

2. SECOM Manufacturing Dataset

Dataset Description

- Total Samples Used: 1,567
- Features Selected: Columns 199–210 (based on correlation and completeness)
- Target: `Pass/Fail` (-1 or 1)

Preprocessing

- Missing values were removed.
- Selected features were normalized between 0 and 1.
- Final dataset had 13 features with no missing data.

Quantum Model

- QSVM trained on a 12-feature subset using a quantum kernel with 12 qubits.
- Training and evaluation done on reduced sets due to computational cost.

Classical Model

- RBF-based SVM trained on the same data with class balancing.

Results (on subset of 50 samples)

Model	Accuracy (Train)	Accuracy (Test)
QSVM	96.0%	94.0%
Classical SVM	78.0%	84.0%

> **Insight**: QSVM showed better generalization and captured subtle patterns in noisy manufacturing data, although the classical model also performed reasonably well.

---

3. NSL-KDD Intrusion Detection Dataset

Dataset Description

- Total Samples Used: 1,000
- Features Selected:
  - `serror_rate`, `srv_serror_rate`, `rerror_rate`, `srv_rerror_rate`
  - `dst_host_serror_rate`, `dst_host_srv_serror_rate`
- Target: `class` (binary or multi-class intrusion labels)

Preprocessing

- Selected continuous features with high correlation to the target.
- Standardized and split using stratified sampling.

Quantum Model

- 6-feature quantum embedding with 6 qubits.

- QSVM trained and evaluated on full 80/20 split.

Classical Model

- RBF-based SVM with class balancing.

Results

Model	Accuracy (Validation)
QSVM	87.5%
Classical SVM	87.5%

> **Insight**: Both models achieved identical performance. This suggests that for this particular feature subset, classical models are competitive, though QSVM maintains parity even with small qubit representations.

Conclusion

Dataset	QSVM Accuracy	Classical SVM Accuracy	Winner
IEEE-CIS	96.0%	73.5–75.0%	**QSVM**
SECOM	94.0%	84.0%	**QSVM**
NSL-KDD	87.5%	87.5%	**Tie**

Key Takeaways

- **QSVM excels** in high-noise or nonlinear datasets like IEEE-CIS and SECOM.
- For more structured datasets like NSL-KDD, **quantum and classical models perform similarly**.
- **QSVM training is computationally more intensive**, limiting scalability unless hybrid quantum-classical methods or approximation techniques are adopted.

Future Work

- Scale QSVM to larger datasets using approximation methods or variational quantum classifiers.
- Explore **multi-class QSVM** and integrate **quantum feature selection** strategies.
- Benchmark against more advanced classical models (e.g., XGBoost, Random Forest) for deeper insights.
