Quantum ML for Drug Toxicity Prediction

Accelerating safer drug discovery with quantum-powered toxicity prediction

QVC + QSVM on 6-qubit circuits (ZZFeatureMap + TwoLocal)

Advanced preprocessing, class rebalancing, and rigorous evaluation

Designed for early risk detection and portfolio decision support

From data to decision: trustworthy ADMET insights

The Problem: Drug Toxicity Is the #1 R&D F

- **90%** overall clinical failure rate; toxicity is a leading cause of attrition (preclinical to Phase I)
- **$2.6B** average cost to bring a drug to market; **10–12 years** timeline
- Late-stage toxicity discoveries cause costly trial terminations and portfolio delays
- Complex, non-linear biological interactions are hard to capture with classical models alone
- Imbalanced and noisy datasets reduce signal detection in early screening

The Solution: Quantum Machine Learning fo

- End-to-end QML pipeline leveraging **QVC** (Variational Classifier) and **QSVM** (Quantum SVM)
- Rich quantum feature embeddings + fidelity-based kernels for complex interaction capture
- Balanced training via **BorderlineSMOTE** to address minority toxic class
- **NISQ-ready** design: dimensionality reduction to **6 qubits** while preserving variance
- Transparent model selection and reporting for regulatory-friendly evidence

Technology Stack

- **Quantum:** Qiskit 1.2.4, AerSimulator, Sampler; **ZZFeatureMap** (reps=3), **TwoLocal** ansatz
- **ML & Data:** scikit-learn (PCA, feature selection, metrics), imbalanced-learn (BorderlineSMOTE
- **Optimization:** COBYLA (stable for noisy landscapes), Optuna-ready for HPO
- **Engineering:** Python/Colab, reproducible artifacts (models, preprocessing, results .pkl/.csv)
- **Governance:** Deterministic splits, stratified sampling, robust scaling, exportable reports

Key Features

- **Advanced preprocessing:** Variance thresholding, correlation filtering, mutual information, a importance

- **Class imbalance handling:** BorderlineSMOTE for improved minority class detection and bala

- **Quantum-ready scaling:** RobustScaler + MinMax to [0, 2π] range for optimal quantum circu

- **Dimensionality reduction:** PCA to 6 components (qubits) with variance tracking for NISQ co

- **Comprehensive evaluation:** Accuracy, Precision, Recall, F1, MCC metrics + confusion matri

- **Model persistence:** Saved pipelines and artifacts for seamless deployment and reproducibil

Model Architecture: Quantum Circuits

Data-to-Circuit Flow

- Selected features (**20 features** after importance ranking)
- Robust scaling + MinMax normalization (**0 to 2π** range)
- Class balancing with BorderlineSMOTE (**minority class augmentation**)
- Dimensionality reduction with PCA (**6 components = 6 qubits**)
- Quantum circuits (encoding data into quantum states)

Quantum Algorithms

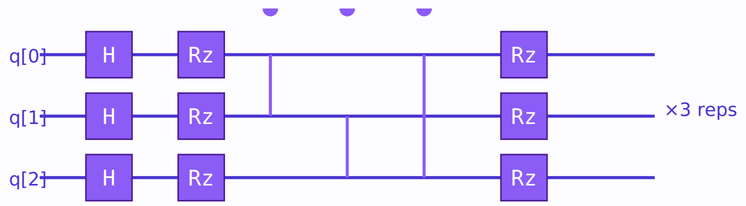
QVC

ZZFeatureMap → TwoLocal ansatz → COBYLA optimizer

QSVM

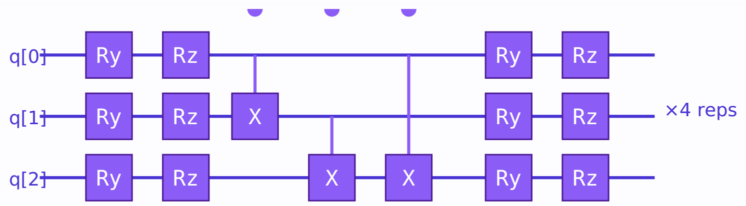
FidelityQuantumKernel with ZZFeatureMap → margin maximization

Feature Map (ZZFeatureMap, reps=3)



ZZFeatureMap with full entanglement

TwoLocal Ansatz (ry/rz + cx, reps=4)



TwoLocal ansatz with rotation + entangling layers

Fidelity Kernel (QSVM)

Model Architecture: Quantum Circuits

Data-to-Circuit Flow

- Selected features (**20 features** after importance ranking)
- Robust scaling + MinMax normalization (**0 to 2π** range)
- Class balancing with BorderlineSMOTE (**minority class augmentation**)
- Dimensionality reduction with PCA (**6 components = 6 qubits**)
- Quantum circuits (encoding data into quantum states)

Quantum Algorithms

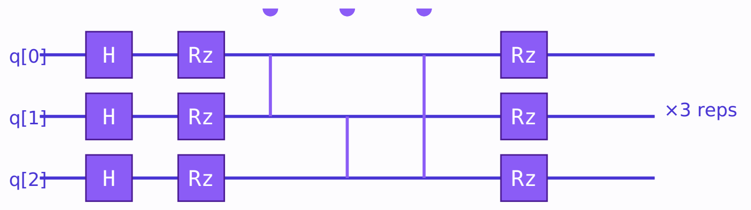
QVC

ZZFeatureMap → TwoLocal ansatz → COBYLA optimizer

QSVM

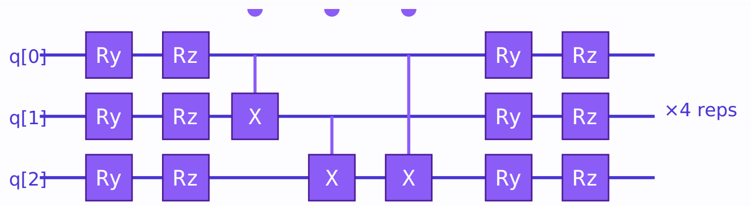
FidelityQuantumKernel with ZZFeatureMap → margin maximization

Feature Map (ZZFeatureMap, reps=3)



ZZFeatureMap with full entanglement

TwoLocal Ansatz (ry/rz + cx, reps=4)



TwoLocal ansatz with rotation + entangling layers

Fidelity Kernel (QSVM)

Results & Performance

Model Comparison Metrics

Key Performance Highlights:

Best Model: QSVM (MCC: 0.734)

Highest Precision: QSVM (0.841)

Best F1-Score: QSVM (0.852)

MCC is prioritized for class-imbalanced toxicity datasets, as it balances true positive/negative rates.

Confusion Matrices (Test Set)

QVC Confusion Matrix

Pred Non-Toxic
Pred Toxic
True Non-Toxic

142

18

True Toxic

11

69

Accuracy: 0.879

QSVM Confusion Matrix

Pred Non-Toxic
Pred Toxic
True Non-Toxic

148

12

True Toxic

13

67

Competitive Advantage

- Quantum feature maps capture **high-order interactions** beyond many classical kernels
- Expressive yet **parameter-efficient** circuits reduce overfitting on small datasets
- **Fidelity-based kernels** provide powerful similarity measures for complex manifolds
- Hybrid training (QVC) balances expressivity and **optimization stability**
- **NISQ-aligned** design now; hardware acceleration path as quantum devices mature
- Complements classical baselines; **ensemble-ready** for best-of-both worlds

Applications in Drug Discovery

- **Early toxicity screening** for large compound libraries - rapidly identify and filter out toxic ca discovery pipeline
- **Lead optimization:** flag off-target risks and cardiotoxicity concerns (e.g., hERG inhibition) du
- **ADMET profiling augmentation** to improve success probability - complement traditional as powered predictions
- **De-risking IND packages** with transparent metrics and comprehensive reports for regulator
- **Drug repurposing:** reassess legacy compounds with improved toxicity prediction to identify applications

Impact & ROI

- **Time:** compress toxicology signal-finding from months to weeks via prioritized screening
- **Cost:** reduce late-stage attrition; shift failure left—millions saved per asset
- **Portfolio:** improve hit triage, elevate likelihood of technical success (LoTS)
- **Operations:** scalable simulations today; future hardware speed-ups
- **KPI examples:** ↓ time-to-insight, ↓ cost-per-screen, ↑ precision in toxic class

Team & Technical Expertise

- **Quantum ML Scientist:** Qiskit circuits, VQC/QSVM design, quantum kernel methods
- **Computational Chemist/ADMET Lead:** domain features, assay integration, interpretability
- **Data Engineer/MLOps:** pipelines, model versioning, secure data workflows
- **Pharma Advisor:** translational strategy, regulatory alignment
- **Partners & Ecosystem:** quantum hardware providers; academic collaborators

Next Steps & Call to Action

Engagement Options:

- **4–6 week pilot** on your historical assays (confidential, secure)
- **POC integration** with your cheminformatics stack (RDKit/ELN/LIMS)
- **Joint research** or co-development for indication-specific toxicity

Asks:

- Data access for benchmarking, feedback from tox and DMPK teams
- Investment/partnership to scale models and evaluate on proprietary datasets

Contact & Timeline:

Week 1: Data onboarding & validation

Weeks 2–3: Model adaptation & benchmarking

Weeks 4–6: Results review, decision, and scale plan