QuantumConoMix: Benchmarking Shallow VQCs for Resource-Constrained Peptide Classification

Quantum technologies promise to reveal faint biological signals that classical methods overlook, from subtle bioelectrical activity to hidden molecular features [1,2,4,5]. As AI tools for protein modeling advance, preparing quantum-ready approaches is urgent—not only for drug discovery but also for managing dual-use risks in biosecurity [3]. We present **QuantumConoMix**, a dataset of conotoxin sequences enriched with physicochemical descriptors and ESM embeddings [6]. Conotoxins are short, venom-derived peptides whose limited residue length makes them challenging to classify. Yet, they are pharmacologically valuable as templates for pain therapeutics and ion channel modulators. Their potency and ease of synthesis demand robust automated screening [6,7]. To gauge near-term feasibility, we benchmark shallow variational quantum classifiers (VQCs), among the most practical architectures for constrained quantum hardware [8,9]. Using physicochemical features alone [10], classical models achieved up to 0.93 accuracy while a two-qubit VQC rarely exceeded 0.60 [8,11]. Adding ESM-2 embeddings [12] transformed performance: classical baselines surpassed 0.97 accuracy, and the VQC improved to 0.69 with ZZ feature maps and COBYLA optimization. Ablation revealed hydrophobicity and charge as the strongest physicochemical signals, while embeddings dominated once introduced, with ESM_58/93 (LogReg) and ESM_62/198 (combined) as top contributors (Figure 1). Results are reported on held-out test data. **QuantumConoMix** thus establishes a benchmark and case study for how feature design shapes classical–quantum tradeoffs, offering a path toward quantum-enhanced classification in computational biology and biosecurity [13,14].

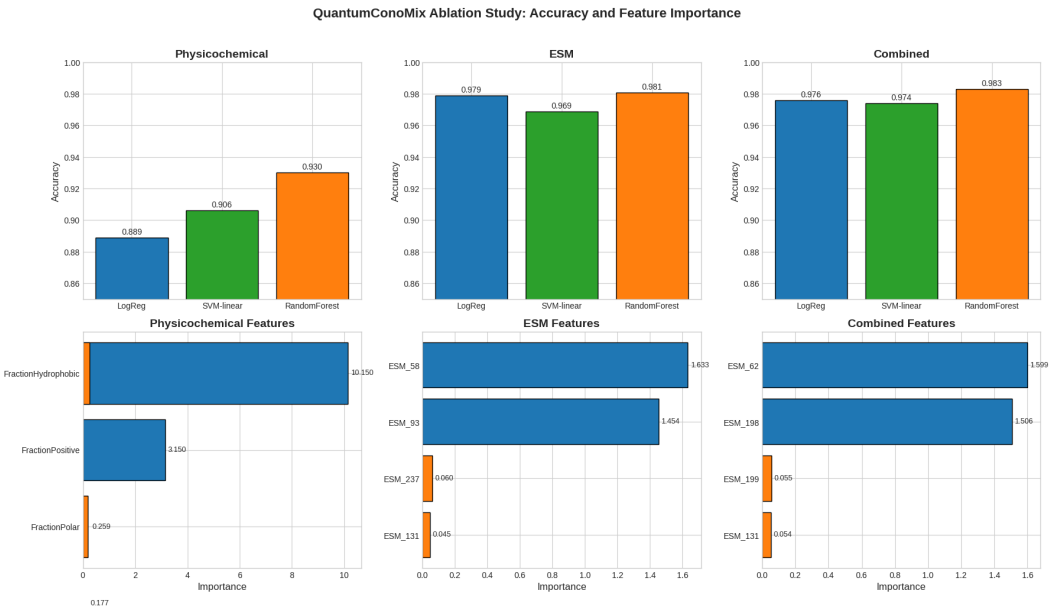

Figure 1: Ablation study of QuantumConoMix (2912 sequences, 331 features). Classical models reached >0.98 accuracy with ESM, highlighting ESM_58/93, while physicochemical models emphasized hydrophobicity and charge. Combined features prioritized ESM_62/198. The top two features from each category were used as qubits in VQC experiments.

References: [1], [2], [3], [4], [5], [6], [7], [8],[9],[10],[11],[12],[13],[14]

Submitted to 39th Conference on Neural Information Processing Systems (NeurIPS 2025). Do not distribute.