

Bipartition Mutual Information Dynamics in Hybrid Quantum Neural Networks: An Exact Information-Theoretic Analysis

Neeshu Rathi*¹ and Monika Agarwal¹

¹Centre for Applied Research in Electronics, Indian Institute of Technology Delhi

Abstract

Understanding how variational quantum circuits encode and reorganize information during gradient-based training remains a fundamental open problem in quantum machine learning. We present an exact mutual information analysis of a 5-qubit Hybrid Quantum Neural Network (HQNN) trained on binary Iris classification, tracking Von Neumann bipartition mutual information $I(A:B)$ across all unique qubit bipartitions throughout the full training trajectory. Our architecture combines a Dense Angle feature map, an XX Hamiltonian evolution layer, and a RealAmplitudes variational ansatz optimized via SPSA, achieving 97% training accuracy over 150 iterations. Mutual information is computed exactly from the circuit density matrix at each checkpoint, ensuring that entropic measures reflect true quantum correlations without sampling approximations.

We identify a two-phase information-plane trajectory: a fitting phase in which classification accuracy rises rapidly while bipartition MI remains approximately stable, followed by an entanglement reorganization phase in which accuracy saturates and MI increases monotonically across most partitions. This temporal decoupling — accuracy improvement preceding entanglement growth — constitutes direct empirical evidence that the information bottleneck principle extends to multi-qubit variational circuits on classical data. Partition-resolved Pearson correlation analysis reveals strong topology dependence: non-local 2—3 partitions achieve MI-accuracy correlations up to $r = 0.95$, while the readout qubit partition yields systematically weaker correlation ($r = 0.45$), indicating that classification-relevant information concentrates in bulk non-local entanglement rather than at the measurement site. These findings provide a principled, information-theoretic framework for interpreting HQNN learning dynamics and guiding topology-aware circuit design.