



Contribution ID: 41

Type: **Poster**

## Expressive and Scalable Quantum Fusion for Multimodal Learning

*Monday 1 December 2025 16:00 (1 hour)*

Multimodal learning, which integrates heterogeneous data modalities including text, vision, and sensor signals, has made remarkable progress. Yet, effectively capturing complex relationships across modalities remains a challenge, especially in settings with numerous input streams. Existing methods often restrict these interactions to remain computationally tractable: tensor-based models enforce low-rank constraints, while graph-based models rely on localized information flow, requiring deep networks to model long-range dependencies. In this work, we propose the Quantum Fusion Layer (QFL), a hybrid quantum-classical architecture that efficiently captures arbitrary-degree polynomial interactions across modalities with linear parameter scaling. We provide theoretical guarantees on QFL's expressivity, supported by a case study demonstrating a query complexity separation from classical tensor-based methods. Empirically, we benchmark QFL across a diverse set of multimodal tasks, ranging from low- to high-modality settings. In small-scale simulations, QFL consistently outperforms classical baselines, showing particularly strong improvements in high-modality scenarios. Specifically, in the best case, QFL achieves a 76% increase in ROC AUC while using only 5% of the parameters compared to a tensor-based method. Against graph-based approaches, QFL improves ROC AUC by 12%, while maintaining a comparable number of trainable parameters. These results provide both theoretical insight and empirical validation for the promise of hybrid quantum models as a scalable and expressive solution for complex multimodal learning tasks.

**Author:** NGUYEN, Tuyen (University of Technology Sydney)

**Presenter:** NGUYEN, Tuyen (University of Technology Sydney)

**Session Classification:** Poster Session

**Track Classification:** Topical Groups: Quantum Science and Technology