Proceedings of the Thirty-Fifth International Joint Conference on Artificial Intelligence (IJCAI-26)
Doctoral Consortium Track

Towards Reliable Multimodal Clinical Decision Support: From Quality-Aware Fusion to Patient Digital Twins

Linpeng Sun
Department of Computer Science, Texas Tech University
linsun@ttu.edu

Abstract

Multimodal deep learning for clinical decision support frequently fails in real-world deployments due to severe data missingness and sensor noise. Standard fusion and Mixture-of-Experts (MoE) architectures assume all modalities are uniformly informative, causing routing collapse and miscalibrated confidence when faced with corrupted inputs. My doctoral research addresses this fragility by shifting the paradigm from generative imputation to uncertainty-guided abstention. As a foundational step, I introduce QA-MoE (Quality-Aware MoE), which explicitly decouples reliability estimation from the routing process. By quantifying epistemic uncertainty, the model dynamically filters noisy modalities before fusion, ensuring sparse and stable inference. Building on this static robustness, my ongoing and future research extends to longitudinal and structured clinical data. I am developing PathFlow-EM, which integrates Expectation-Maximization to handle hidden variable distributions in unevenly sampled time-series, and Uncertainty-Guided Dynamic Topological Fusion (U-DTF), which uses topological data analysis to gate features based on evolving patient graphs. Ultimately, this research trajectory aims to deliver robust, calibration-aware architectures that form the computational basis for reliable Patient Digital Twins.

1 Introduction and Research Vision

Deep learning models for clinical decision support increasingly rely on multimodal inputs, integrating electronic health records (EHR), medical imaging, and continuous vitals. While theoretical performance is strong, real-world deployment frequently fails. Mainstream multimodal fusion and MoE gating mechanisms [Shazeer *et al.*, 2017; Zhou *et al.*, 2022] operate under the assumption that all available modalities provide uniform benefits. In practice, clinical data suffers from severe structural missingness, sensor noise, and distribution shifts. When weakly informative or corrupted modalities are allowed to perturb the routing process, the network experiences routing collapse and miscalibrated confidence.

My doctoral research addresses the gap between idealized architectures and messy clinical reality. Instead of relying on generative imputation [Yoon *et al.*, 2018], which risks hallucinating clinical features, my work shifts the paradigm towards uncertainty-guided abstention and dynamic structural adaptation. The overarching vision of my PhD is to develop robust inference frameworks that can gracefully handle missingness and temporal shifts, ultimately serving as the computational foundation for predictive Patient Digital Twins.

1.1 Current Work: Quality-Aware Multimodal Routing

To prevent noisy modalities from degrading overall predictions, I developed QA-MoE, a modular architecture that explicitly decouples reliability estimation from the routing process. As illustrated in Figure 1, traditional MoEs mix feature extraction and routing, making them vulnerable to corrupted inputs. In QA-MoE, an Evidential Quality Scorer first quantifies the epistemic uncertainty of each modality. This score guides a Stability-Enhanced Subset Selector to dynamically filter out noisy signals before they reach the experts. For cases of severe structural missingness, I introduced a Ternary Expert Aggregation branch that acts as a fail-safe. This mechanism ensures that the model maintains sparse, interpretable, and stable fusion even when critical modalities drop out unexpectedly.

1.2 Ongoing and Future Directions: Temporal and Topological Dynamics

Building on the static robustness of QA-MoE, my current focus expands to longitudinal patient trajectories and continuous physiological tracking. To address the inherent irregularity and sparsity of clinical events, I am developing PathFlow-EM, a framework that integrates Expectation-Maximization to explicitly infer hidden variable distributions across complex, unevenly sampled clinical time-series data. In parallel, to scale robust sequence modeling to high-frequency clinical streams, I have designed Phase-Mamba, a linear state-space architecture for continuous physiology forecasting. Evaluated on the MIMIC-IV and PhysioNet 2019 datasets for cross-center validation, Phase-Mamba regulates information flow by applying a rigorous filtering threshold to the normalized Shannon Entropy of the system dynamics rather than a

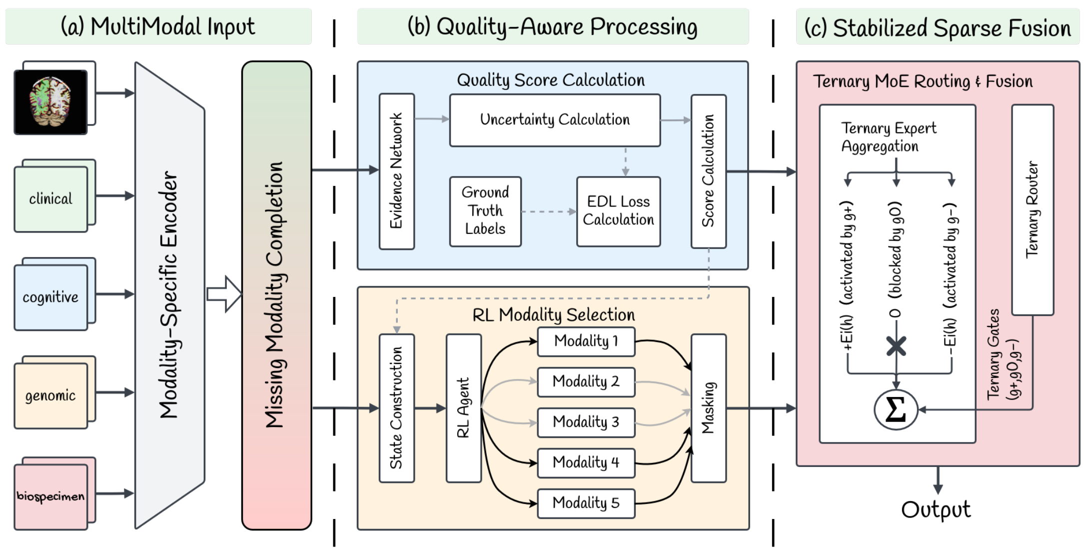

Figure 1: Overview of the QA-MoE architecture, which explicitly decouples reliability estimation from routing by dynamically filtering noisy multimodal clinical inputs through an evidential quality scorer to achieve robust, stabilized fusion via a ternary MoE.

standard gating coefficient, thereby effectively isolating unpredictable physiological noise. Concurrently, to handle the high-dimensional structural relationships within multi-omics and patient similarity networks, my future work involves U-DTF. By applying topological data analysis [Hensel *et al.*, 2021] to track multi-scale persistent homological features, U-DTF aims to gate features based on evolving graph structures rather than treating patient data as isolated snapshots.

2 Evaluation Strategy and Broader Impacts

Evaluating clinical models requires metrics that go beyond basic accuracy, prioritizing calibration and computational efficiency under duress.

2.1 Empirical Validation on Clinical Benchmarks

My frameworks are tested on heavily noisy, real-world datasets. For instance, QA-MoE was evaluated on the ADNI dataset for Alzheimer's staging (combining imaging and cognitive scores) and the MIMIC-IV dataset for Length-of-Stay prediction (dense EHR time-series). The results demonstrate that QA-MoE not only outperforms strong multimodal baselines in predictive accuracy but also maintains low Expected Calibration Error (ECE) under simulated missingness ranging from 10% to 50%. Furthermore, bypassing noisy experts on the fly significantly cuts down unnecessary computation overhead. Specifically, we compare QA-MoE against state-of-the-art multimodal fusion baselines, including standard MoE routing, multimodal gating architectures, and traditional non-MoE models. Under severe simulated missingness, QA-MoE reduces the ECE while simultaneously achieving the highest accuracy, F1, and AUC scores.

2.2 Societal Impact and Ethical Considerations

As the healthcare sector moves towards continuous patient monitoring, the ability to synthesize multimodal streams without catastrophic failure from missing data becomes a prerequisite for creating reliable Patient Digital Twins. However, deploying such models introduces strict ethical constraints. In medical AI, an overconfident incorrect prediction is far more dangerous than an abstention. By explicitly modeling epistemic uncertainty at the modality level, my architecture prevents the system from making high-confidence guesses on severely incomplete patient profiles. Future iterations of this work will also incorporate privacy-preserving constraints into the topological graphs to ensure that learning from patient similarity networks does not compromise individual confidentiality or amplify historical biases.

References

[Hensel *et al.*, 2021] Felix Hensel, Michael Moor, and Bastian Rieck. A survey of topological machine learning methods. *Frontiers in Artificial Intelligence*, 4:681108, 2021.

[Shazeer *et al.*, 2017] Noam Shazeer, Azalia Mirhoseini, Krzysztof Maziarz, Andy Davis, Quoc Le, Geoffrey Hinton, and Jeff Dean. Outrageously large neural networks: The sparsely-gated mixture-of-experts layer. *arXiv preprint arXiv:1701.06538*, 2017.

[Yoon *et al.*, 2018] Jinsung Yoon, James Jordon, and Mihaela Schaar. Gain: Missing data imputation using generative adversarial nets. In *International conference on machine learning*, pages 5689–5698. PMLR, 2018.

[Zhou *et al.*, 2022] Yanqi Zhou, Tao Lei, Hanxiao Liu, Nan Du, Yanping Huang, Vincent Zhao, Andrew M Dai, Quoc V Le, James Laudon, et al. Mixture-of-experts with expert choice routing. *Advances in Neural Information Processing Systems*, 35:7103–7114, 2022.