Dynamic Modeling of Patients, Modalities and Tasks via Multi-modal Multi-task Mixture of Experts

Chenwei Wu[*], Zitao Shuai[*], Zhengxu Tang[*], Luning Wang, & Liyue Shen[†]

Department of Electrical and Computer Engineering, University of Michigan

[*]Equal Contribution

Introduction

Multi-modal multi-task learning holds significant promise in tackling complex diagnostic tasks in medical imaging. However, two key challenges exist:

- **Dynamic modality fusion:** The quality and amount of task-related information from different modalities varies across patient samples.
- **Modality-task dependence:** Different clinical tasks require dynamic feature selection and combination from various modalities.

Traditional methods use fixed fusion strategies, potentially underutilizing modalities with stronger diagnostic signals for specific patients or tasks.

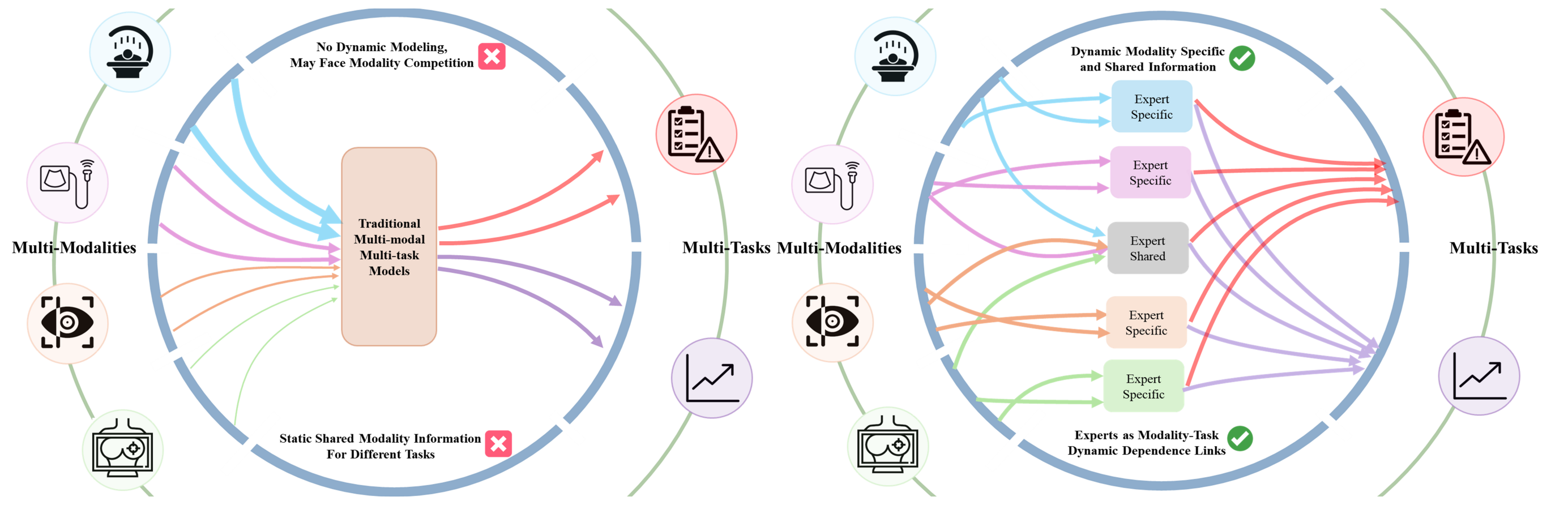
No Dynamic Modeling, May Face Modality Competition ✗
Multi-Modalities
Traditional Multi-modal Multi-task Models
Multi-Tasks
Static Shared Modality Information For Different Tasks ✗
Dynamic Modality Specific and Shared Information ✓
Multi-Modalities
Expert Specific
Expert Specific
Expert Shared
Expert Specific
Expert Specific
Multi-Tasks
Experts as Modality-Task Dynamic Dependence Links ✓

Figure 1. Traditional multi-modal multi-task modeling vs M^{4}oE in medical imaging

Main Contributions

- A novel framework for multi-modal multi-task learning, achieving both sample-dynamic modality fusion and dynamic modality-task dependence modeling
- Our MToE module provides a simple yet effective way to jointly model modality, expert, and task together
- Inspired by [1], we introduce a new loss to encourage experts to dynamically learn diverse patterns of task-dependent modality-shared information
- Extensive experiments on four benchmark datasets for breast cancer screening and retinal diagnosis demonstrate our method's effectiveness and generalization ability

Challenges in Medical Multi-modal Multi-task Learning

Challenge 1: Dynamic Modality Information

First mentioned in [2], clinical information between multi-modalities can be decomposed into:

- **Modality-specific** information unique to each modality
- **Modality-shared** information present across modalities

Challenge 2: Dynamic Modality-Task Dependence

Different diagnostic tasks require modality information to be fused differently. For example, in the mammography screening:

- **Breast Density Assessment:** FFDM provides higher-resolution details
- **Cancer Detection:** 2DS offers better visibility for calcified cancers in dense tissue

Problem Illustration

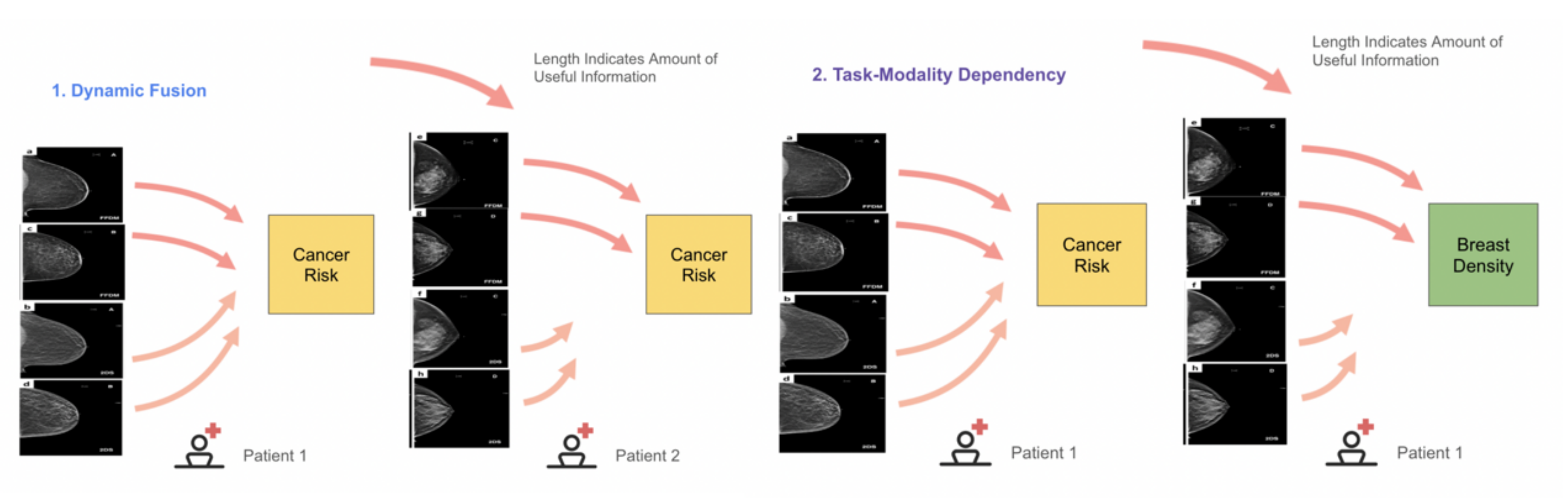
1. Dynamic Fusion
Length Indicates Amount of Useful Information
Cancer Risk
Patient 1
Cancer Risk
Patient 2
2. Task-Modality Dependency
Length Indicates Amount of Useful Information
Cancer Risk
Patient 1
Breast Density
Patient 1

Figure 2. Two Challenges in Real Clinical Process

Proposed Method: M^{4}oE

We propose **M^{4}oE**, a novel **M**ulti-modal **M**ulti-task **M**ixture of Experts framework for **M**edical diagnosis.

M^{4}oE consists of two main modules:

- **Modality-Specific MoE (MSoE):** Extracts and retains input-dependent modality-specific features
- **Modality-shared Modality-Task MoE (MToE):** Models shared modality information and modality-task dependence

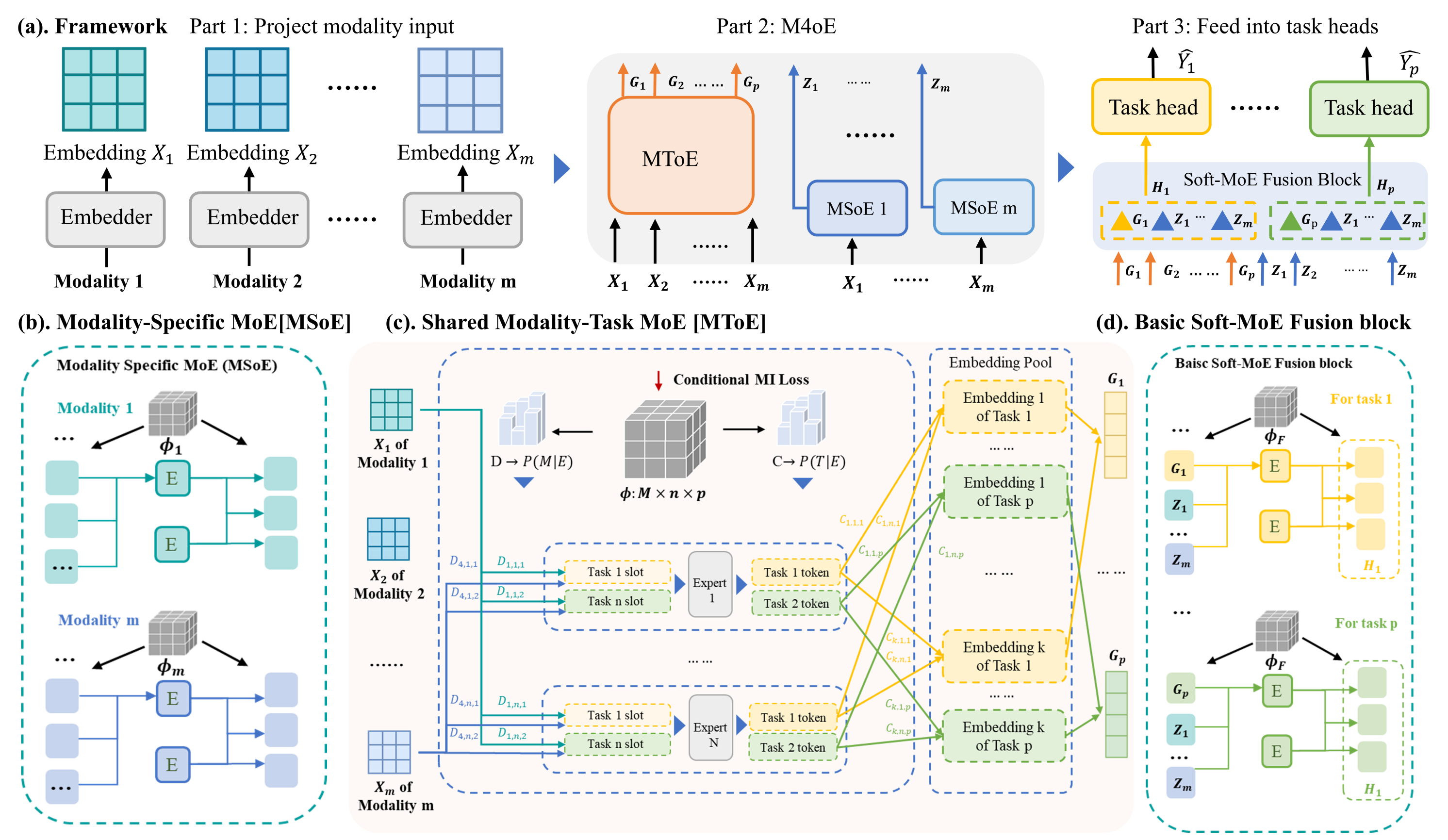

Figure 3. (a) Overall M^{4}oE framework (b) MToE block (c) MSoE block (d) Fusion block

Main Results

Setting	Method	EMBED			RSNA		VinDR		GAMMA	
		Risk	Density	BI-RADS	Density	BI-RADS	Density	BI-RADS	Glau.	Seg.
Multi-Task	Natural	84.8	83.7	73.9	77.1	66.4	89.0	70.4	89.2	87.2
	Medical	84.6	83.3	73.4	76.8	63.1	85.9	66.4	87.2	88.5
	M^{4}oE	**85.9**	**84.1**	**75.1**	**77.8**	**66.7**	**89.6**	**71.8**	**90.4**	**89.7**

Table 1. Performance comparison on benchmark datasets (accuracy %). SOTA represents best single-task medical AI methods. Natural and Medical represent best multi-task methods from natural domain and medical domain.

Key Findings

- Our M^{4}oE significantly outperforms all baseline methods across tasks and benchmarks
- Ablation studies indicate effectiveness of each of our key components, demonstrating meaningful modeling of specific and shared information
- MToE can be flexibly integrated with existing medical AI backbones to improve their performance

Analysis Visualizations

Modality Competition Analysis:

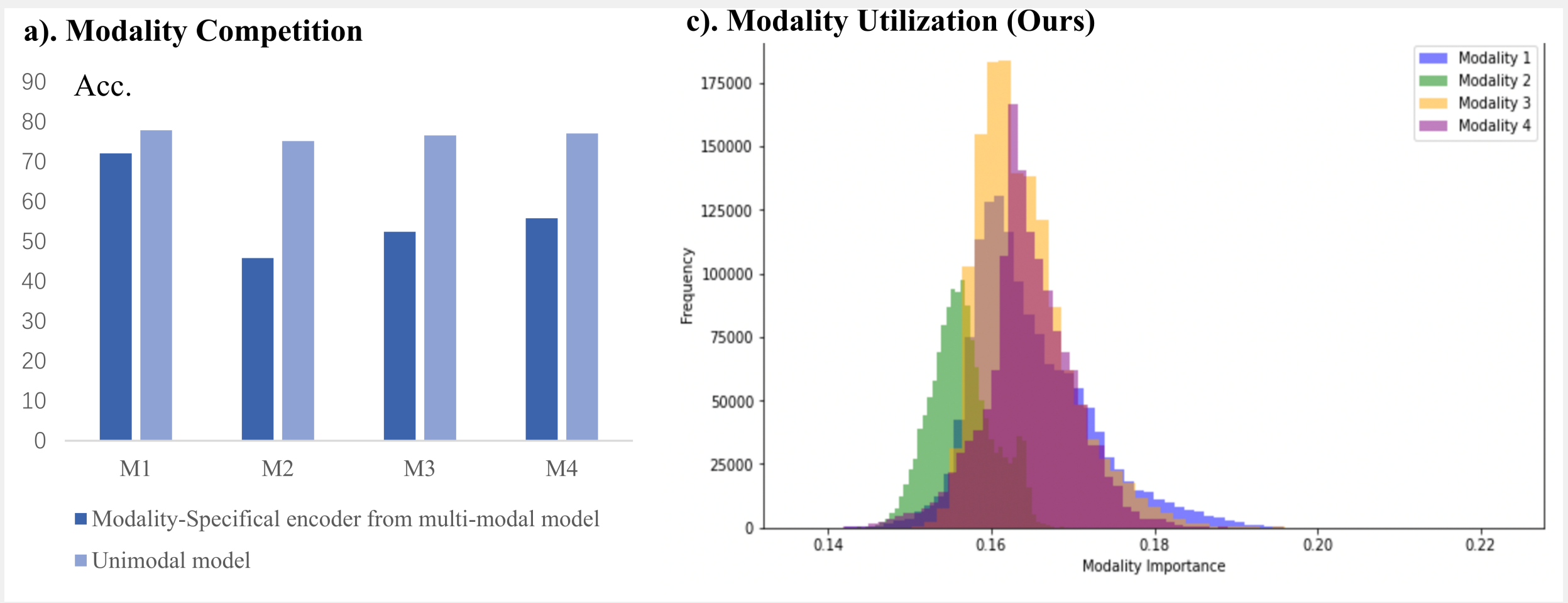

Figure 4. (a) Performance drop from modality competition (b) Modality utilization of M^{4}oE

Conclusion

- We proposed M^{4}oE, a Multi-modal Multi-task Mixture of Experts framework for medical imaging
- Our approach enables sample-adaptive dynamic modality fusion and modality-task dependence modeling
- M^{4}oE consistently outperforms SOTA methods across diverse medical imaging benchmarks
- Our framework approximates modality contribution and can be flexibly combined with different backbones

Future Work:

- Extending to handle missing modalities and missing labels [3]
- Applying to other medical imaging domains and tasks
- Exploring clinical deployment opportunities

References

[1] Chen Z, Shen Y, Ding M, et al. Mod-squad: Designing mixtures of experts as modular multi-task learners[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023: 11828-11837.

[2] Liang P P, Cheng Y, Fan X, et al. Quantifying & modeling multimodal interactions: An information decomposition framework[J]. Advances in Neural Information Processing Systems, 2023, 36: 27351-27393.

[3] Han X, Nguyen H, Harris C, et al. Fusemoe: Mixture-of-experts transformers for fleximodal fusion[J]. arXiv preprint arXiv:2402.03226, 2024.