Text-Grounded Interpretability for Mammography Risk Prediction via Vision-Language Alignment

Shantanu Ghosh[1], Vedant Parthesh Joshi[2], Rayan Syed[1], Aya Kassem [1], Abhishek Varshney [1], Payel Basak [3], Weicheng Dai [1], Judy Wawira Gichoya [4], Hari M. Trivedi [4], Imon Banerjee [2], Shyam Visweswaran [5], Clare B. Poynton [3,6], Kayhan Batmanghelich[1]

[1]Dept. Of Electrical and Computer Engineering, Boston University, [2]Data Science, Analytics and Engineering, Arizona State University, [3]Boston University Chobanian & Avedisian School of Medicine, [4]Department of Radiology, Emory University, [5]Department of Biomedical Informatics, University of Pittsburgh, [6]Department of Radiology, Boston University Medical Campus

TL;DR: We build a breast cancer **risk predictor** with **vision** and **language interpretability** without compromising **performance**.

Why interpretable risk?

1. Risk prediction enables **personalized screening**
2. Risk-stratified screening allows for **earlier detection**
3. Interpretability increases **clinician's trust**

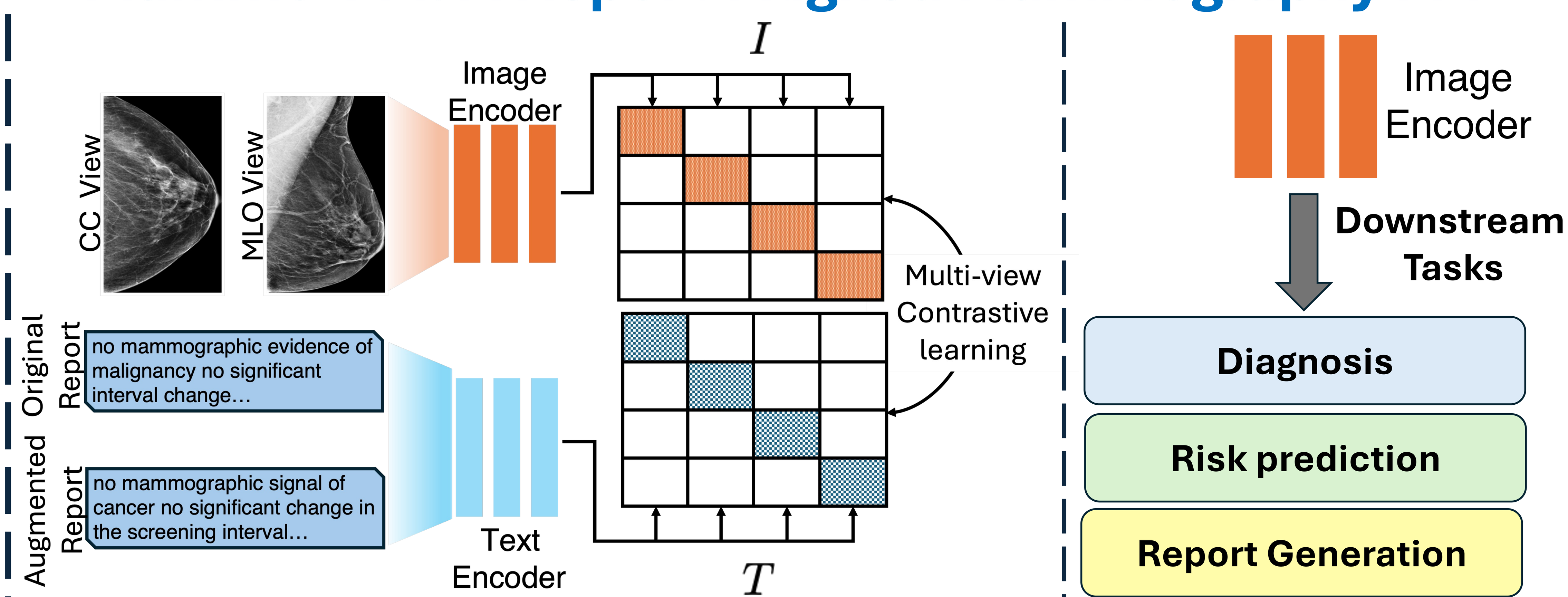
The gap:

1. MIRAI: accurate, but a **black box**
2. Existing methods: heatmap interpretability only
3. Clinicians reason using both **image** and **language**

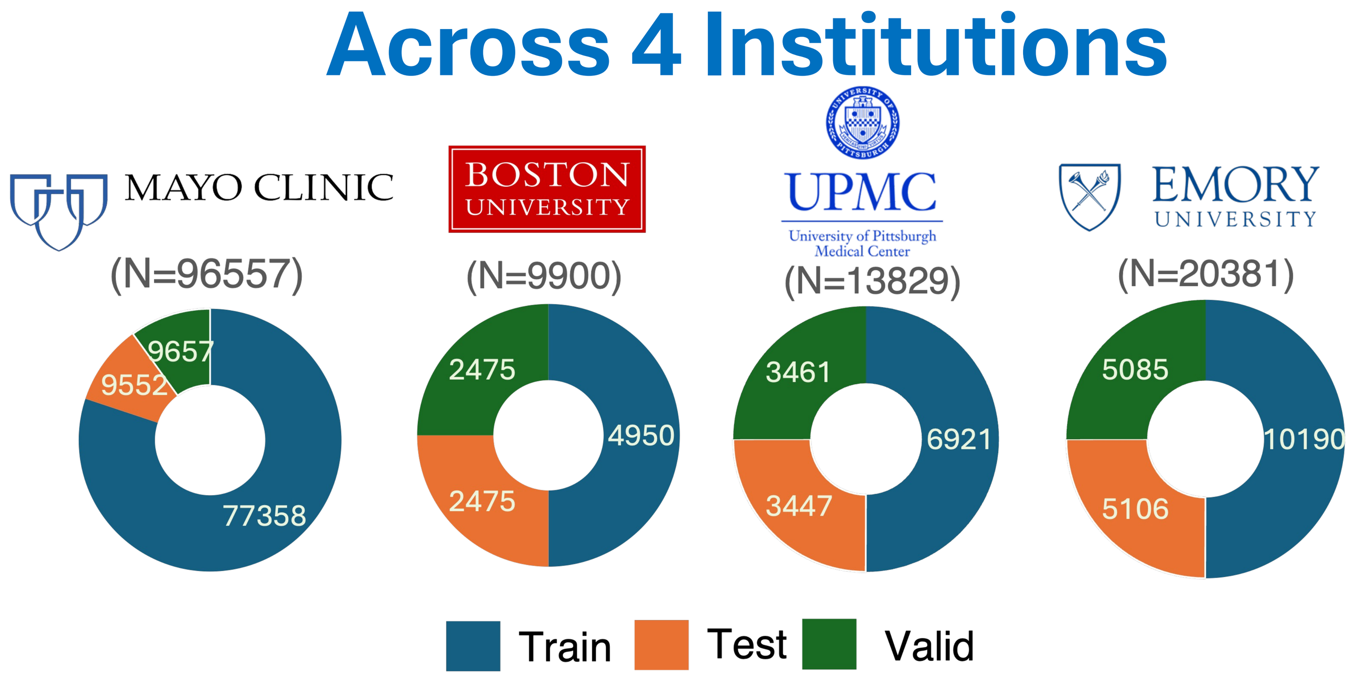
Our fix:

1. Build a **report-aligned** mammography FM
2. Ground risk in both **image** and **text**
3. Preserve **SOTA accuracy**

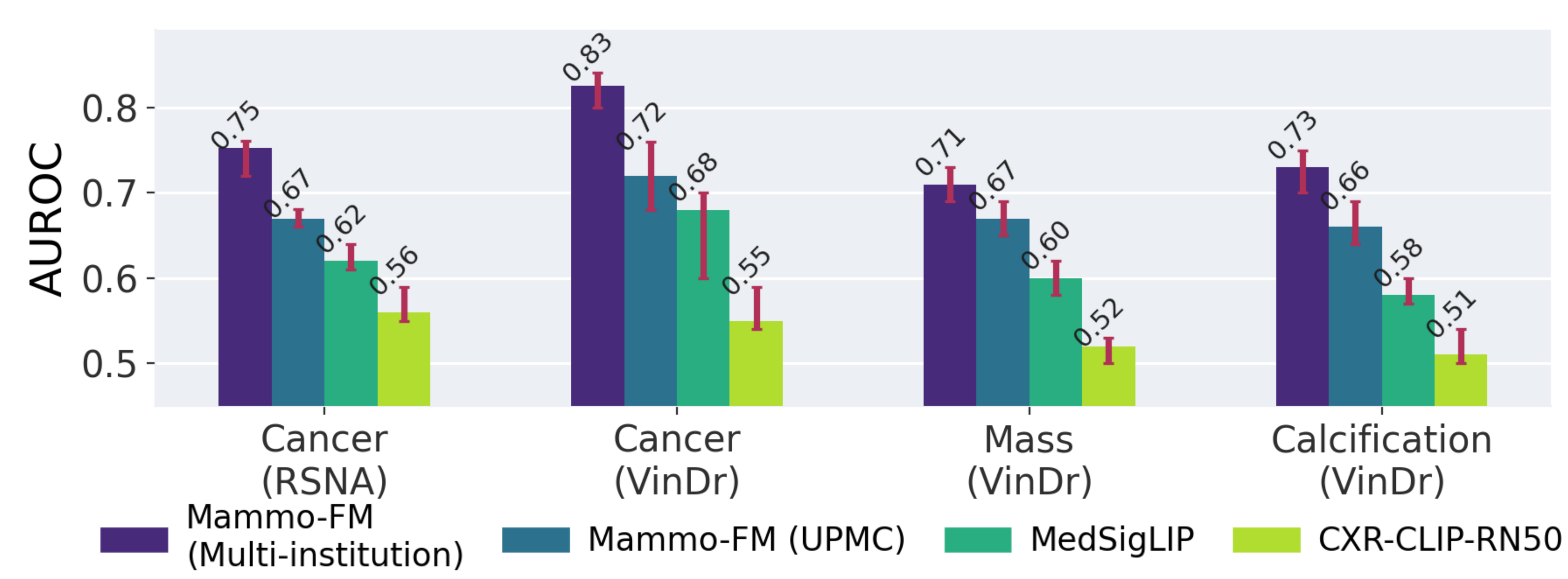
Mammo-FM: A Report-Aligned Mammography FM



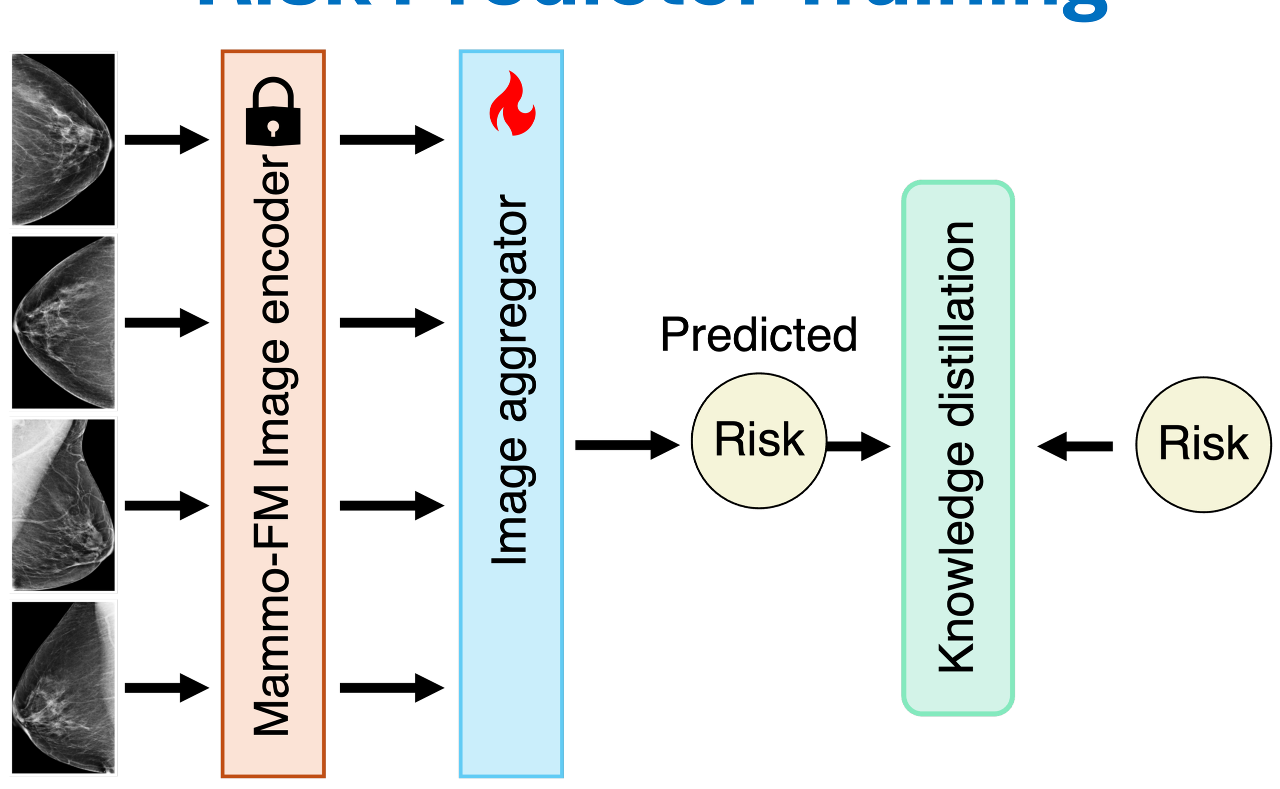
800K+ Pre-training Mammograms Across 4 Institutions



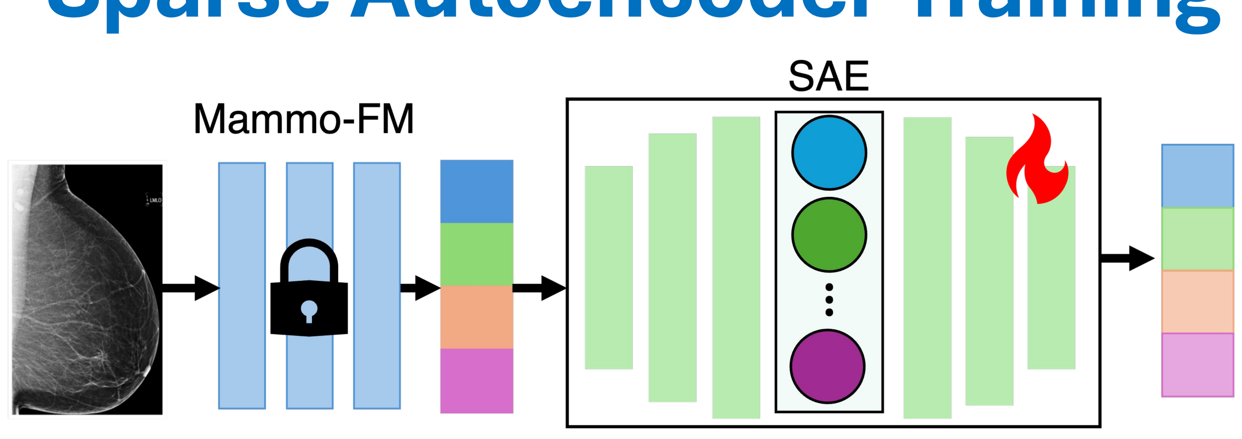
Zero-Shot Classification Results



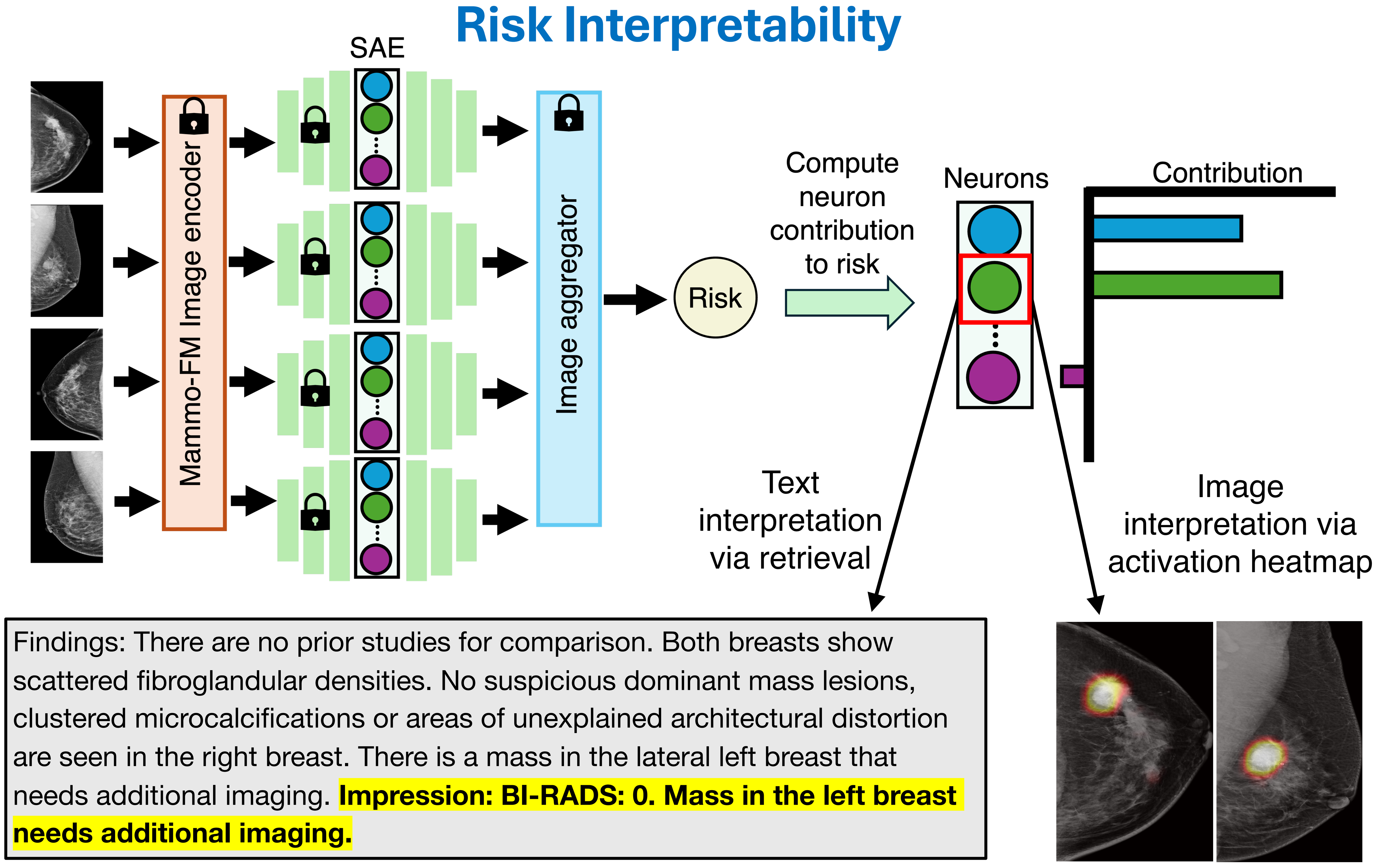
Risk Predictor Training



Sparse Autoencoder Training



Risk Interpretability



Risk Prediction Results