Mammo-CLIP: A Vision Language Foundation Model to Enhance Data Efficiency and Robustness in Mammography

Shantanu Ghosh[1], Clare B. Poynton[2], Shyam Visweswaran[3], Kayhan Batmanghelich[1,2]

[1]Dept. Of Electrical and Computer Engineering, Boston University

[2]Boston University Chobanian & Avedisian School of Medicine

[3]Intelligent Systems Program (ISP), University of Pittsburgh

MICCAI 2024 Marrakesh MOROCCO

BOSTON UNIVERSITY | BATMAN LAB

TL;DR: A vision language model trained on both mammogram-report pairs and mammogram-attribute datasets, enhancing data efficiency, robustness, and interpretability

Goal: Create an organ specific (breast) Foundation Model

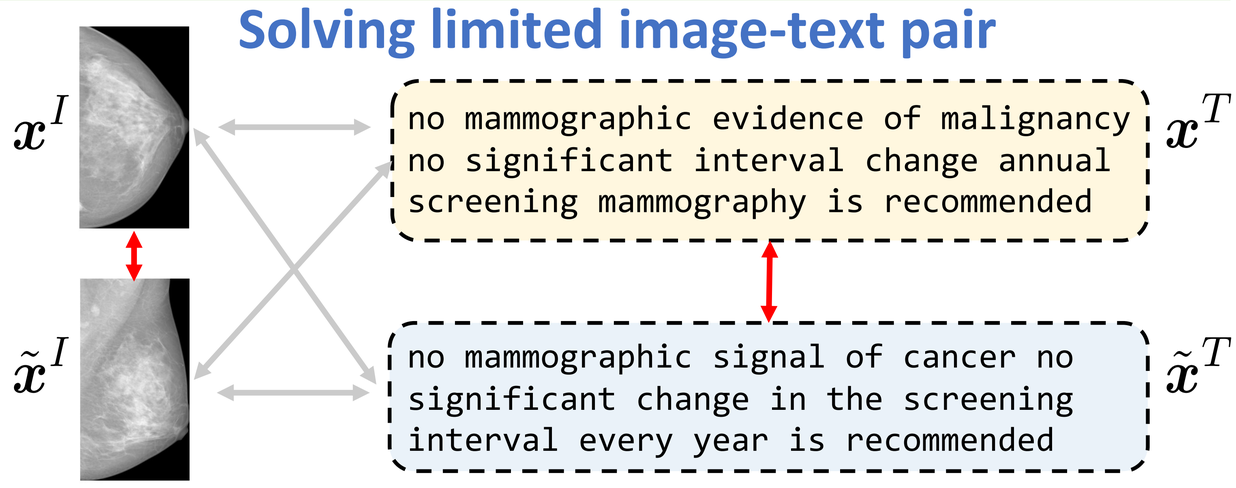
Why:

1. Improving data efficiency
2. Transparency/Interpretability
3. Improving Robustness

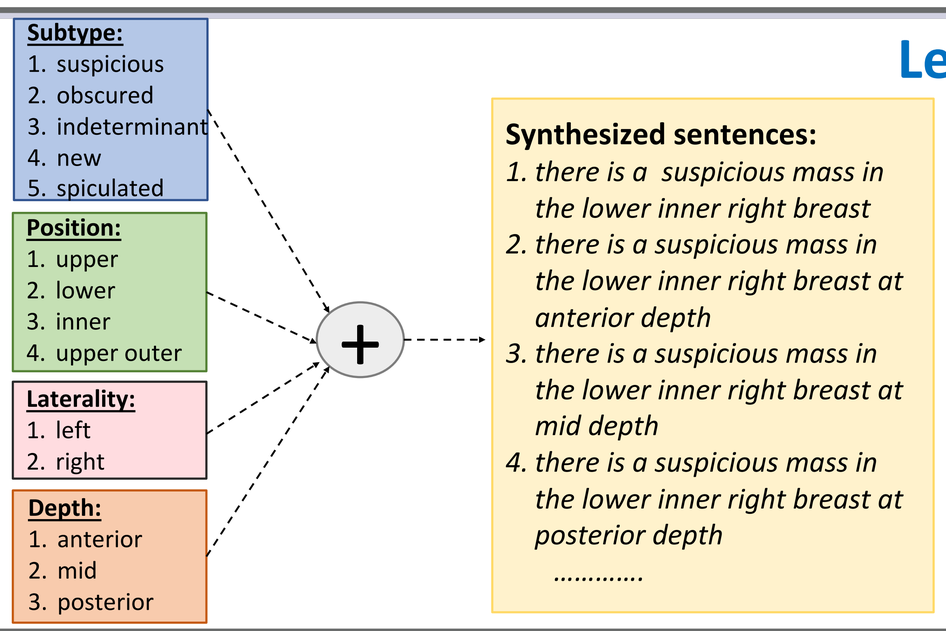
Challenges:

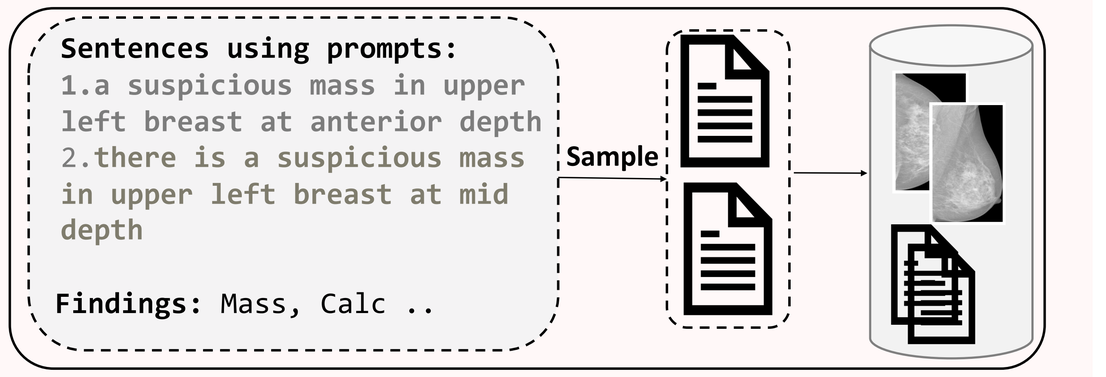
1. How to train w/ limited image-text pair (11k patients)
2. Leveraging image-label data (5k patients)
3. Rigorous evaluation (ZS, Linear probe, Finetuning)

Solving limited image-text pair

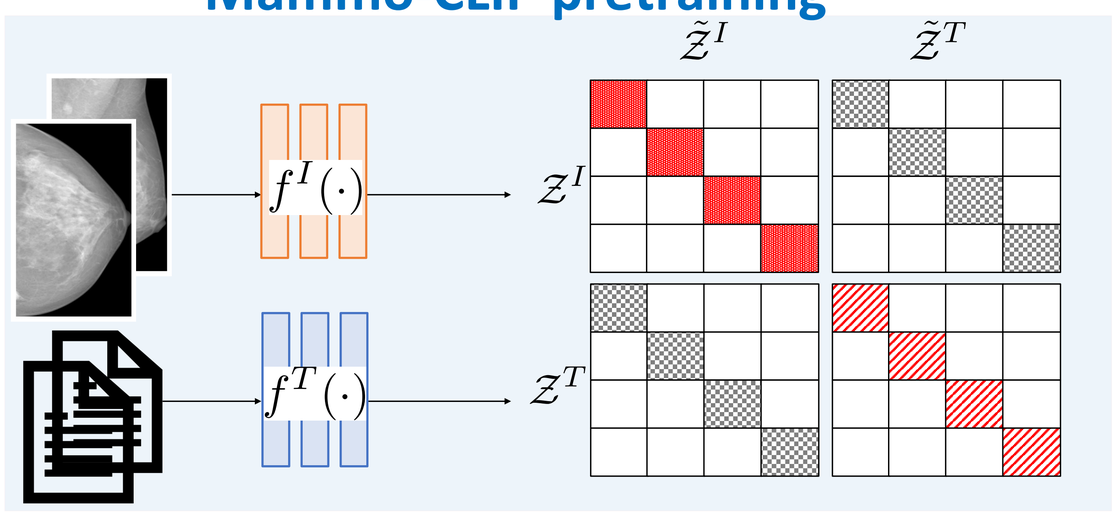


Leveraging image-label data

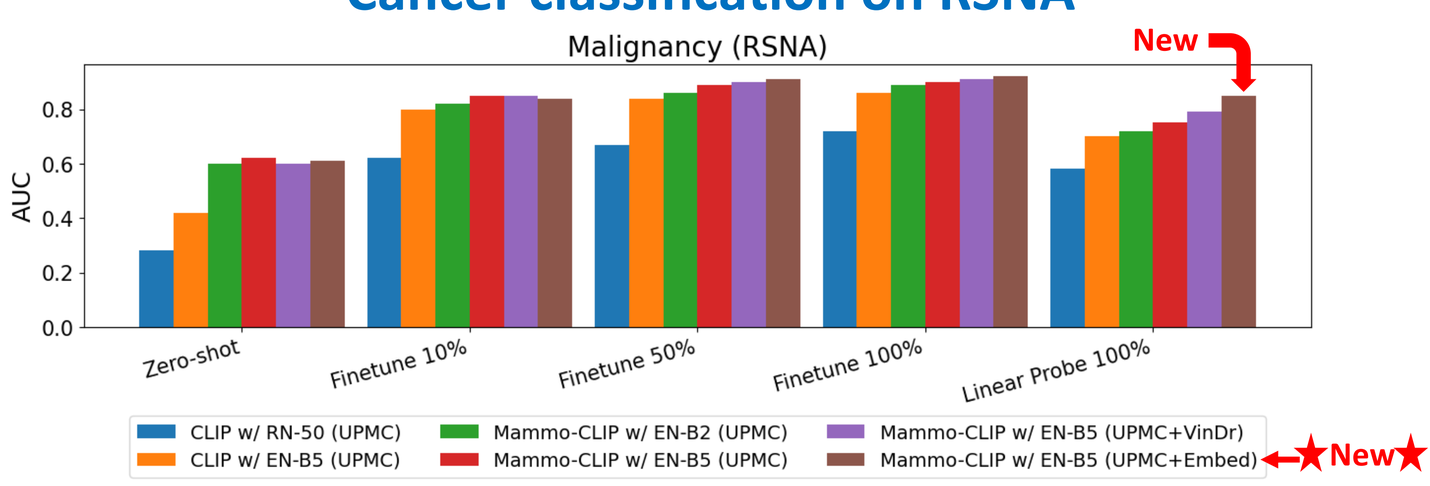




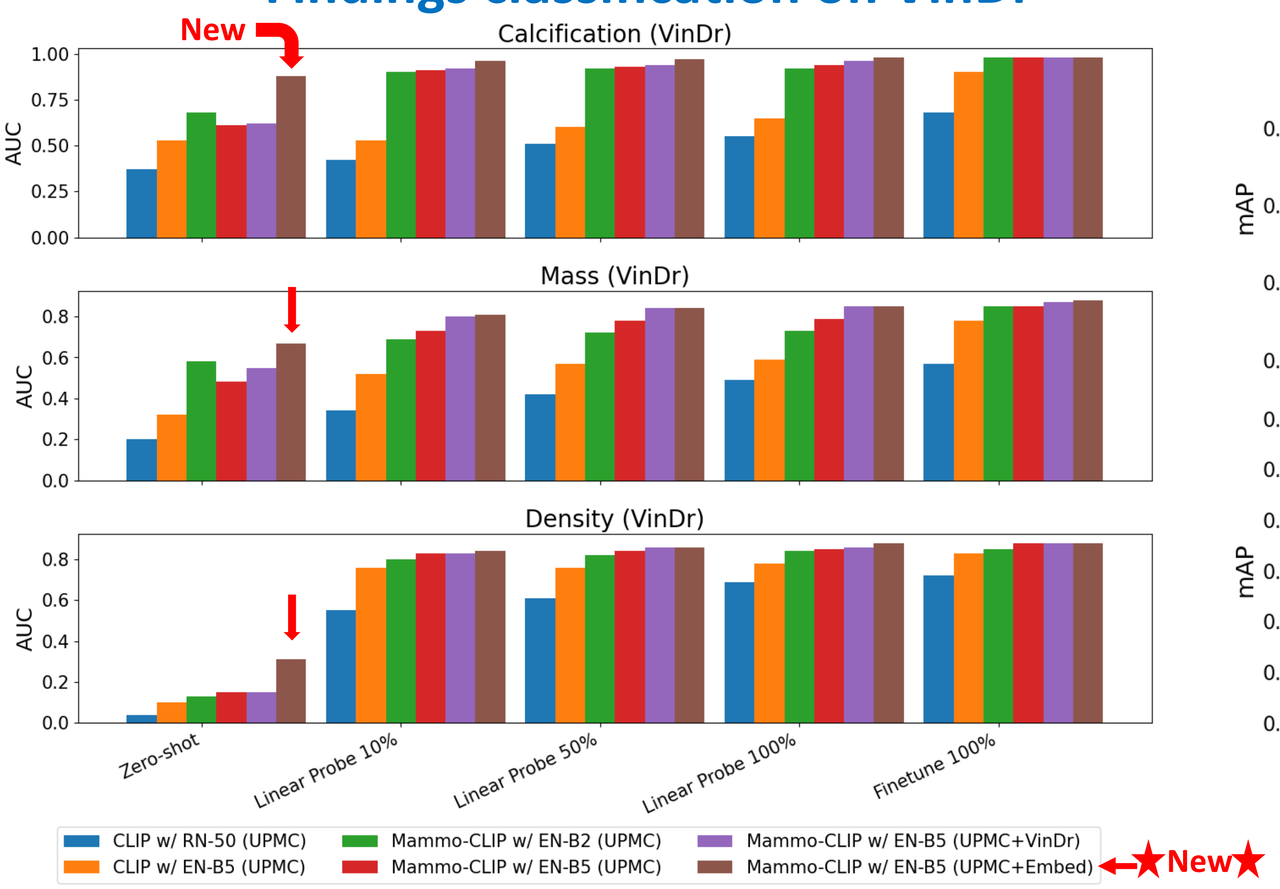
Mammo-CLIP pretraining



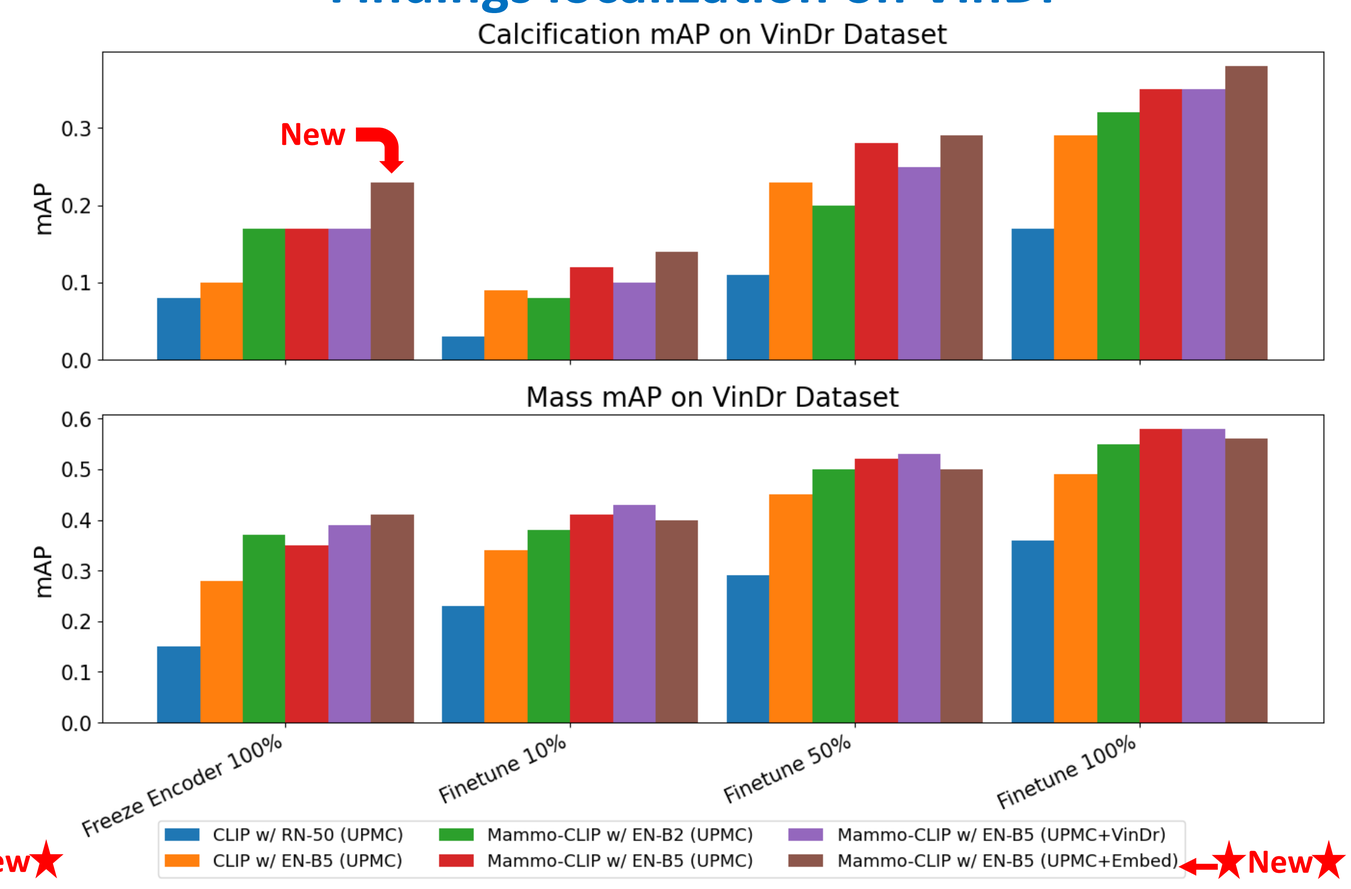
Cancer classification on RSNA



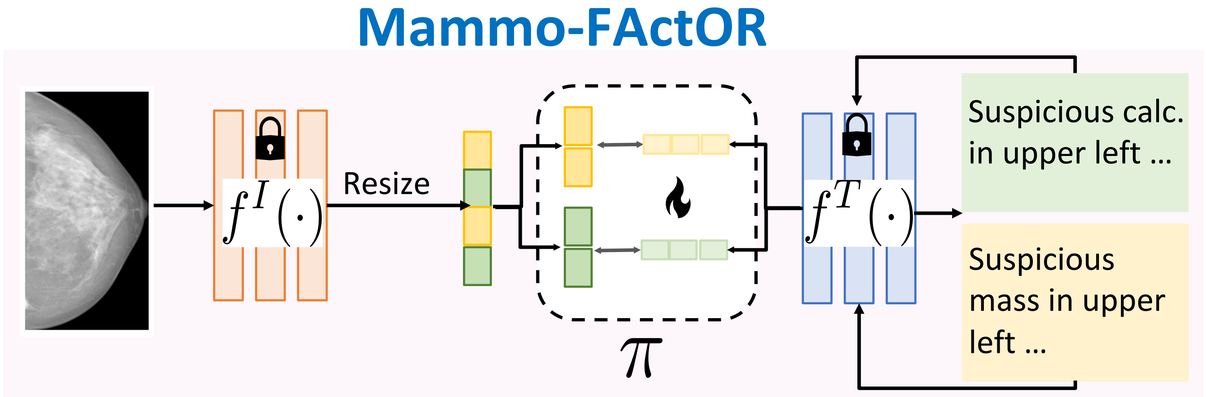
Findings classification on VinDr



Findings localization on VinDr



Mammo-FActOR



Mammo-FActOR localization