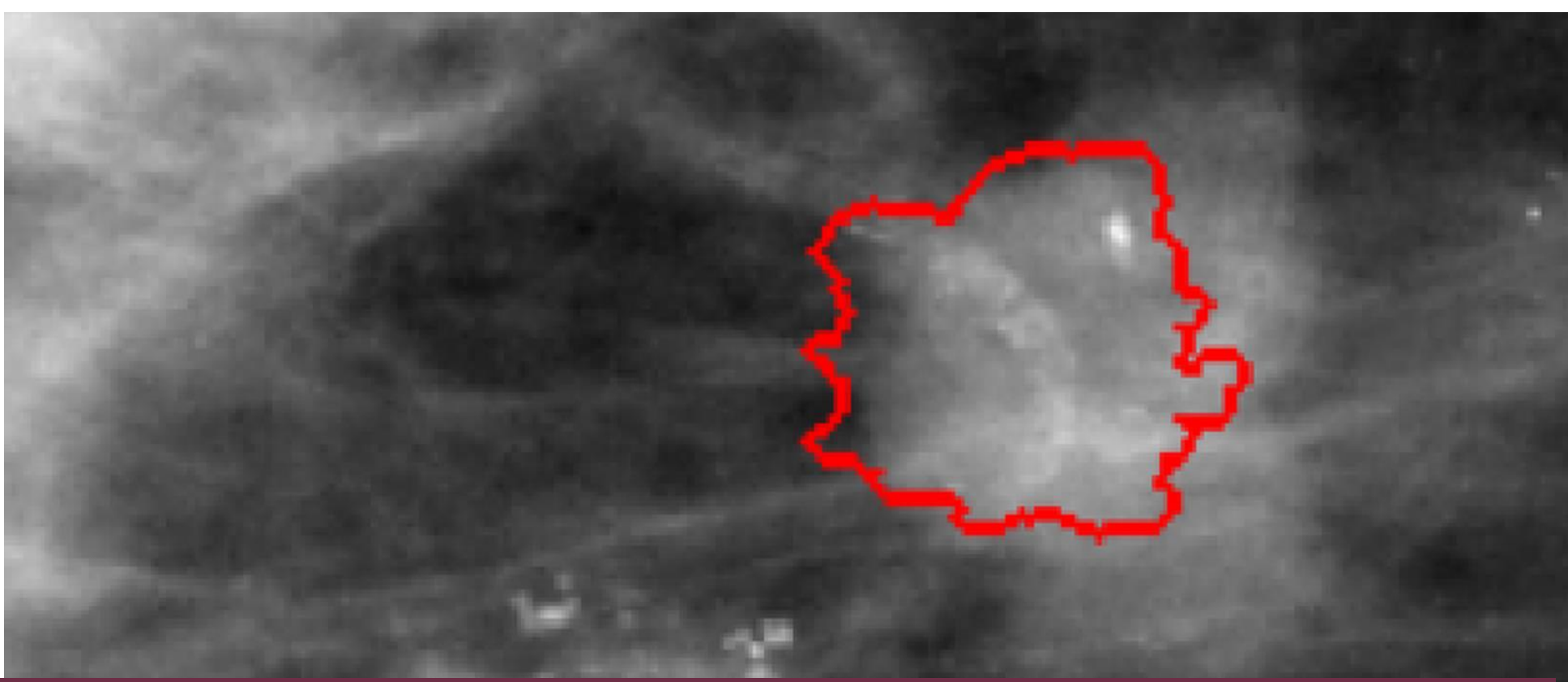
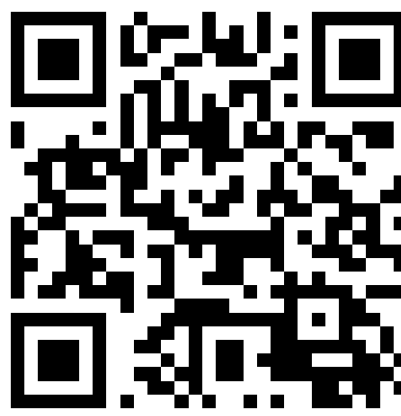


Semantic Feature Modulation

for Mammographic Lesion Classification



Shahar Mahpod and Gil Ben-Artzi

Department of Computer Science, Ariel University · gilba@ariel.ac.il · <https://gil-ba.com>

Code

github.com/shahrma/semantic-mammo

01 Abstract

- A new way to integrate expert knowledge into deep learning models, aligned with how radiologists actually read a mammogram.
- BI-RADS descriptors enter as a continuous embedding.
- An ordinal structure of clinical knowledge emerges.
- Improves AUC over image-only and descriptor baselines.

02 BACKGROUND

Two ways to read a mammogram

RADIOLOGIST

reads through the BI-RADS lexicon

- Breast density
- Mass shape · mass margins
- Calcification morphology · distribution
- Structured, auditable, communicable**

IMAGE-ONLY DEEP LEARNING MODELS

reads pixels

- ROI patch → benign / malignant
- Descriptors already in the record are discarded
- Nothing a reader can inspect or contest
- Nothing to interrogate when it disagrees**

03 Why integrating BI-RADS is hard

Reading has inherent ambiguity [1,2]

Disagreement is between adjacent values [3]

04 Prior work

Descriptors as input in mammography

- Deep BI-RADS Network — descriptors as input (Ben-Artzi, Daragma & Mahpod, ICPR 2024)
- Complexity-Efficient Deep Learning — several architectures (Ben-Artzi, ICPR 2026)
- All require hard descriptor labels, so none can represent ambiguity inside a category or support what-if analysis over it.

Three integration mechanisms — our baselines

- Feature-level fusion — Huang et al., npj Digit. Med. 2020 → our Concat baseline
- Channel gating — Hu et al., TPAMI 2020 → our SE-Gate baseline
- Auxiliary descriptor heads — Zhang et al., MLSP 2021 → our MTL baseline
- Conditional normalization — FILM (AAAI 2018), AdaIN (ICCV 2017), de Vries et al. (NeurIPS 2017)- natural images only
- None was applied to mammography, and all fuse the descriptors differently from ours.

06 Setup and baselines

Identical backbone and protocol for every method — all baselines except Base use the same descriptors.

CBIS-DDSM ~3.5K ROIs
CMMD ~1.6K ROIs
5 independent runs
6 descriptors

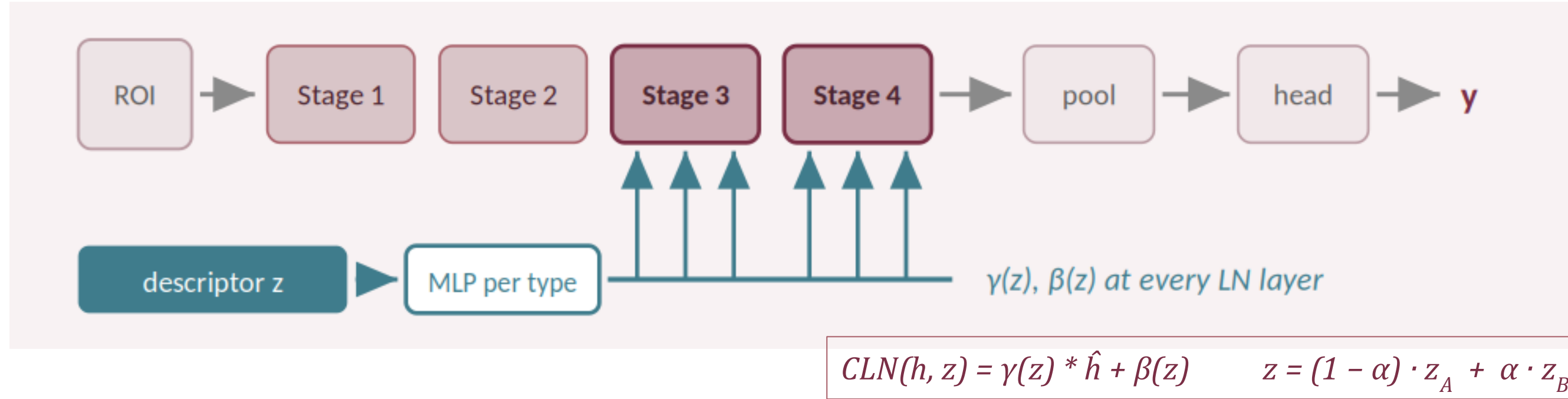
- ConvNeXt-Base, ImageNet-pretrained · single-view ROI crops from radiologist annotations
- Three settings — mass only, calcification only, and joint (one model over both)
- Descriptors: density, shape, margins, morphology, distribution, lesion type · mean ± std

Baseline
Base — image only
Concat — late fusion
SE-Gate — channel gating
MTL — auxiliary heads
SFM (ours) — conditional LN

07 METHOD

SFM — Semantic Feature Modulation

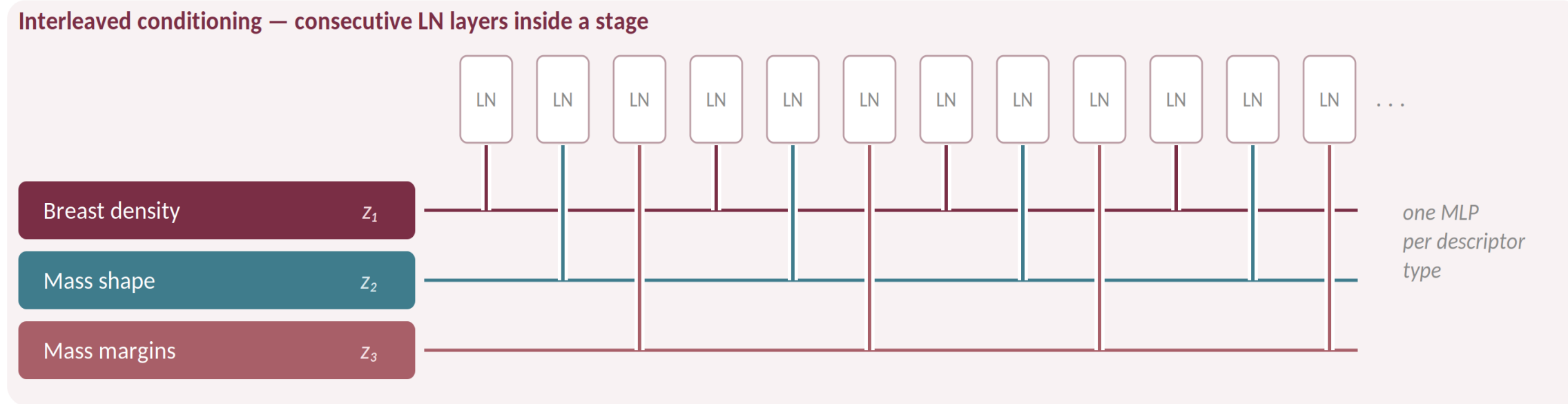
SCOPE The lesion is already localized in the mammogram. · We have the BI-RADS descriptors, without the assessment category. · Goal: binary classification — predict benign vs. malignant.



Discrete → continuous

Context-conditional

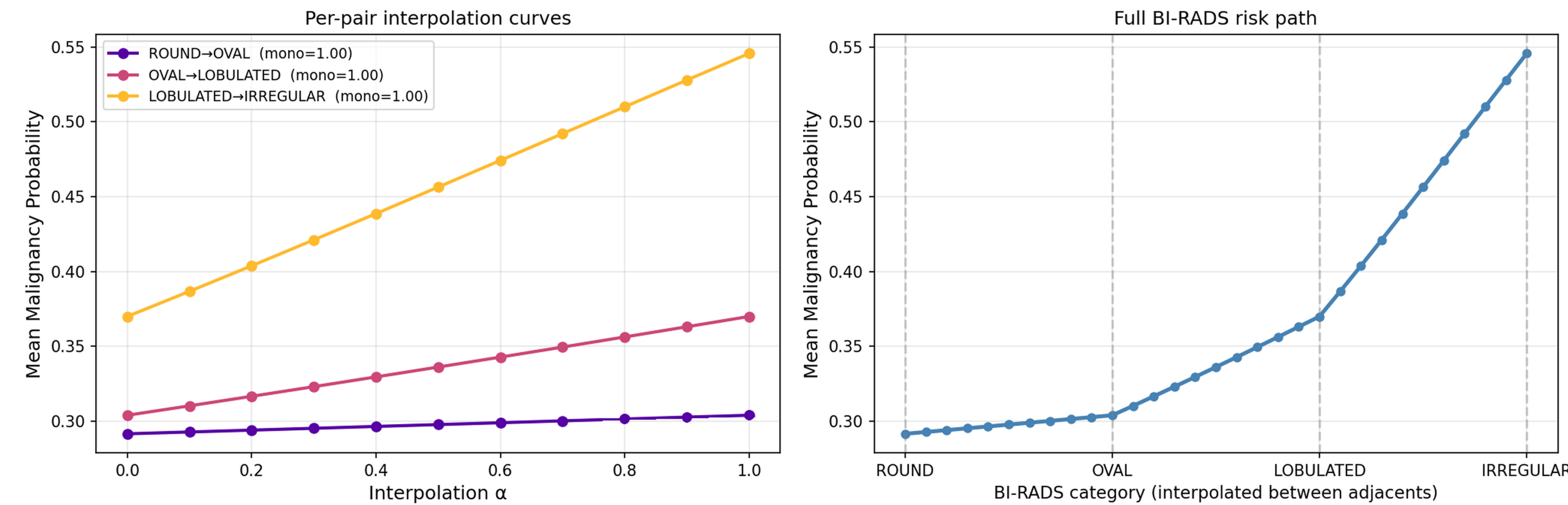
Distinct roles



08 RESULT

Emergent ordinal structure aligned with clinical reasoning

Trained on hard labels, queried with soft ones · mass-shape sweep across CBIS-DDSM test ROIs



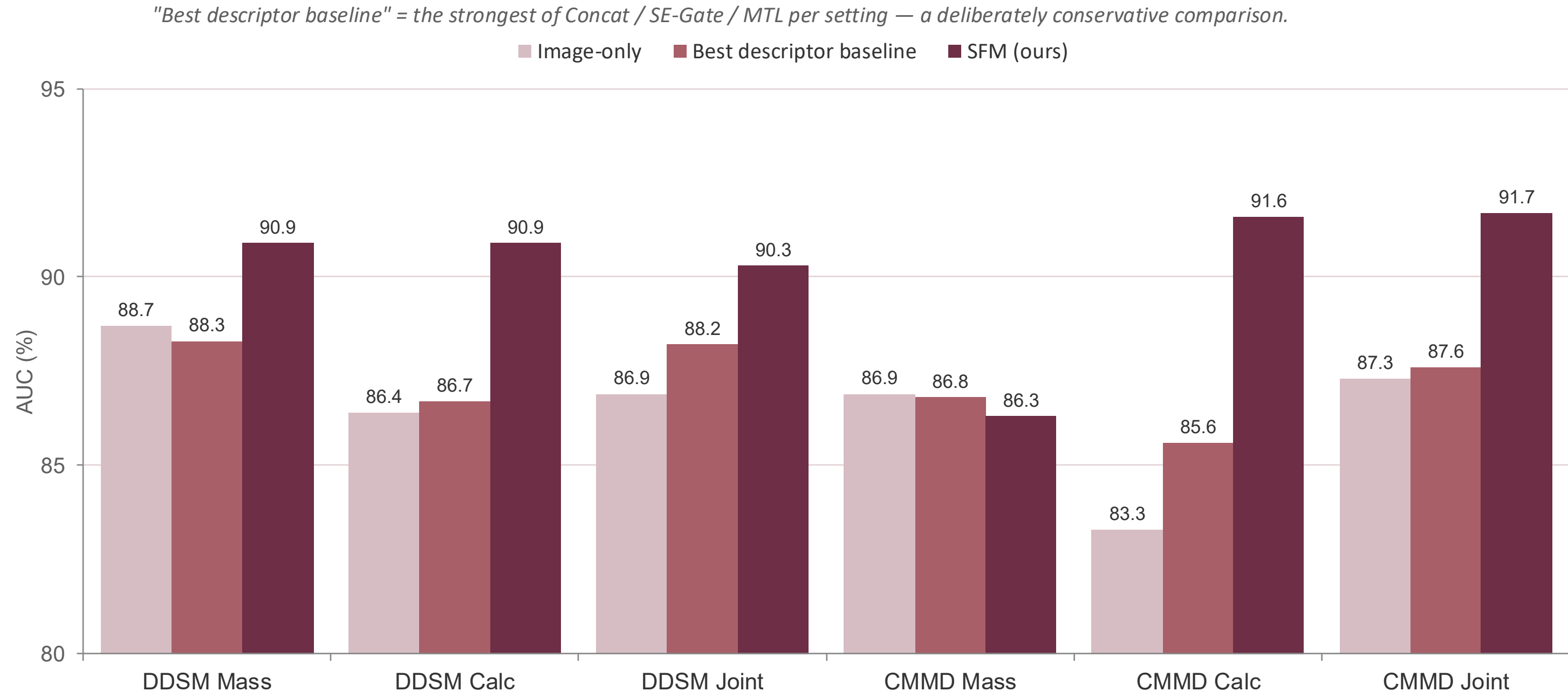
Between two categories — gradual, not rounded. A 60/40 mixture is not pushed to the nearer attribute. It receives its own risk value, so the mixture itself carries information.

Across the path — uneven spacing. Near-flat ROUND→OVAL→LOBULATED, then a sharp rise into IRREGULAR: an unequal risk scale, learned with no ordinal supervision.

Every adjacent pair increases monotonically (score = 1.00) despite training only on hard categorical labels.
Mean malignancy probability rises 0.29 → 0.55 for shape and 0.22 → 0.69 for margins.
Trained on one-hot labels only. No soft ground truth exists for these mixtures — the interior is a generalization probe.

09 RESULT

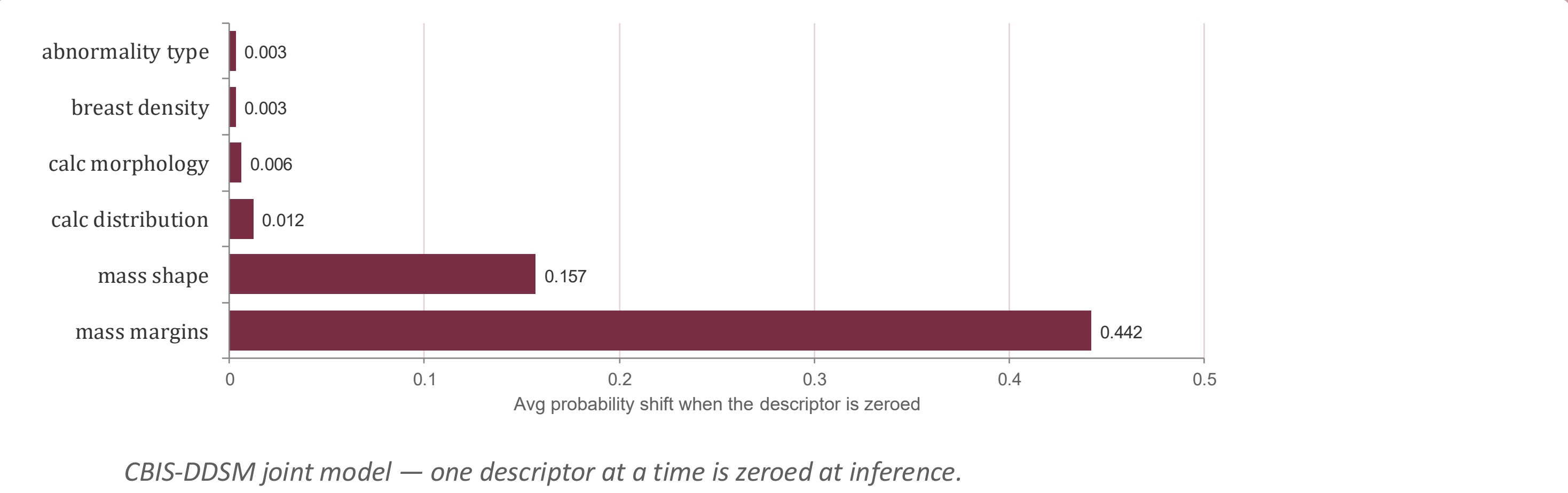
AUC across both datasets and all three lesion settings



Mean: +3.7pp
Max: +8.3pp
5 / 6 clear top AUC
CMMD mass: 86.9 ± 0.3 vs 86.3 ± 0.6

10 RESULT

The model relies on the descriptors



~ 25 % of joint-setting predictions change class when the descriptors are removed.

- Margins and shape dominate our perturbation analysis.
- Density and lesion type produce the smallest probability shifts in this model, which suggests the model infers them from the image.

11 TAKE AWAY

Take three things

- BI-RADS descriptors should be modeled continuously.
- Conditioning captures context-dependence better than static fusion
- Expert knowledge is an efficient way to improve models.

12 LIMITATIONS

What this study does not yet show — and what is next

- Descriptors come from a reader**
At inference a radiologist supplies the descriptors, so the model assists a reading rather than producing one.
Descriptors vary across readings
Training and inference rely on standard-quality descriptors.
- Lesion-level setting**
Annotated ROIs, not whole-image screening — one step away from the deployment task.

- Robust descriptor prediction directly from pixels**
Predicted from the image and re-injected into the model, removing the annotation dependency. **ONGOING**
- Automatic ROI recovery**
Accurate automatic lesion localization makes the pipeline fully end-to-end. **ONGOING**