

Semantic Feature Modulation for Mammographic Lesion Classification

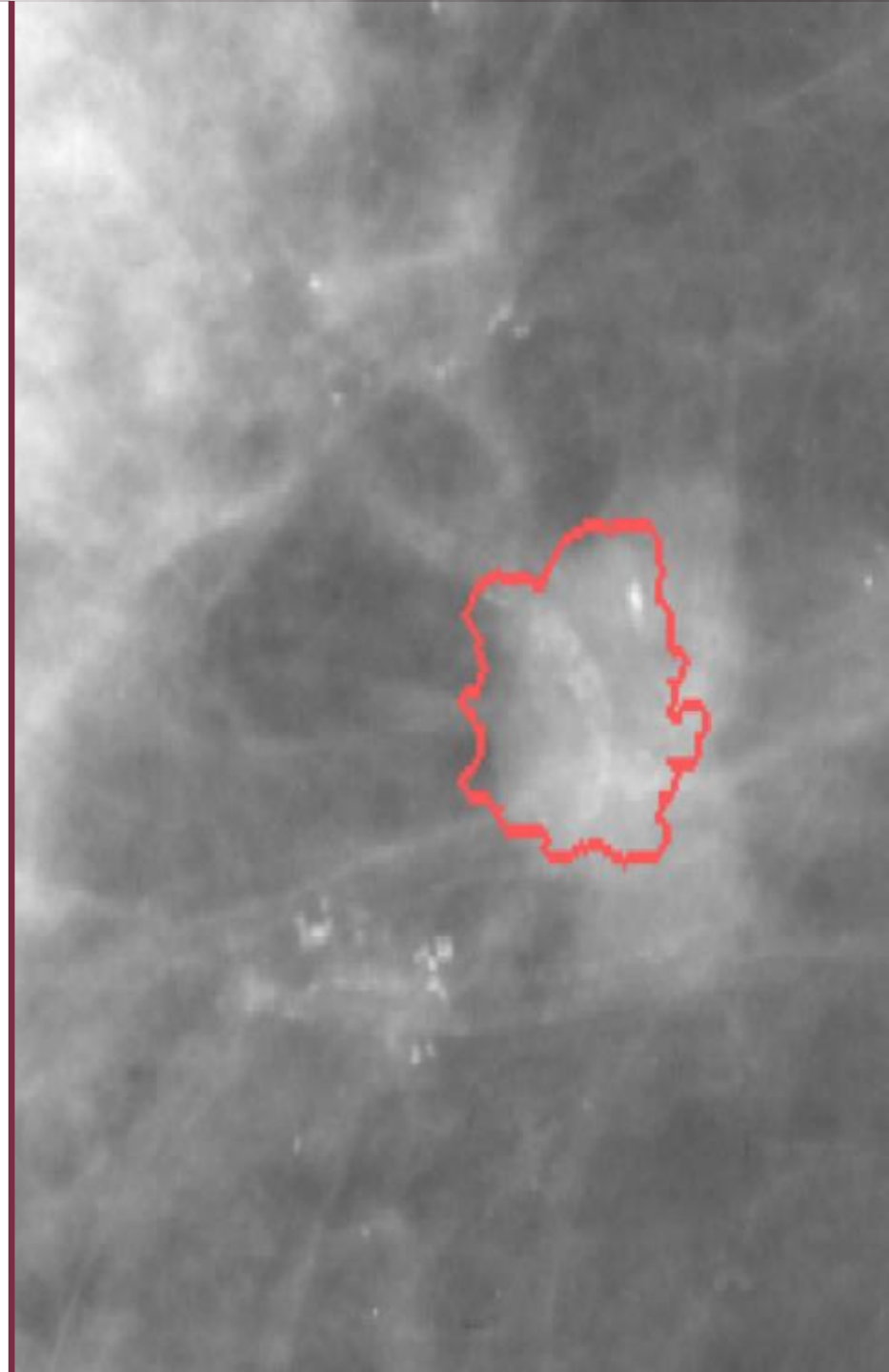
Shahar Mahpod and **Gil Ben-Artzi**

Computer Science School, Ariel University, Israel



Code

github.com/shahrma/semantic-mammo



Two ways to read the same mammogram

RADIOLOGISTS

Read through the BI-RADS lexicon

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

IMAGE-ONLY DEEP LEARNING MODELS

Read 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**

Scope



The lesion is already localized in the mammogram.



We have the BI-RADS descriptors (without assessment)



Goal: binary classification — predict benign vs. malignant

Why integrating BI-RADS is hard



Reading has inherent ambiguity

- Descriptor annotations vary across radiologists, despite a standardized lexicon. [1]
- This includes shape and margins, the descriptors our model relies on most. [2]



Disagreement between adjacent values

- Both across radiologists, and on repeat readings by the same radiologist. [3]

[1] Lee et al. Inter-reader Variability in the Use of BI-RADS Descriptors for Suspicious Findings on Diagnostic Mammography: A Multi-institution Study of 10 Academic Radiologists. *Academic Radiology*. 2017;24(1):60–66.

[2] Berg WA, Campassi C, Langenberg P, Sexton MJ. Breast Imaging Reporting and Data System: inter- and intraobserver variability in feature analysis and final assessment. *AJR American Journal of Roentgenology*. 2000;174(6):1769–1777.

[3] Redondo et al. Inter- and intraradiologist variability in the BI-RADS assessment and breast density categories for screening mammograms. *British Journal of Radiology*. 2012;85(1019):1465–1470. doi:10.1259/bjr/21256379.

Related work

The three ways descriptors are integrated — our baselines



Feature-level fusion / concatenation (Huang, Pareek, Seyyedi, Banerjee & Lungren, npj Digit. Med. 2020) → **our Concat baseline**

Channel gating (Hu, Shen, Albanie, Sun & Wu, TPAMI 2020) → **our SE-Gate baseline**

Auxiliary descriptor heads (Zhang, Vakanski & Xian, MLSP 2021; IEEE Access 2023) → **our MTL baseline**

None was applied to mammography, and all fuse the descriptors differently from ours.

Generic conditional normalization



FiLM (Perez et al., AAAI 2018) · AdaIN (Huang & Belongie, ICCV 2017) · de Vries et al. (NeurIPS 2017) **natural images**

Related work

Descriptors as input in mammography



Complexity-Efficient Deep Learning — descriptors as an input, several architectures (Ben-Artzi, ICPR 2026) **breast · mammography**
Deep BI-RADS Network — descriptors as an input (Ben-Artzi, Daragma & Mahpod, ICPR 2024) **breast · mammography**

What they share



Descriptors are integrated in a simple way: each is input as a single binary value representing the selected attribute.

The gap: clinical knowledge is lost in translation



BI-RADS discretizes a continuous range

- The radiologist perceives a continuous range.
- The observation is projected onto a discrete set of categories, and the model receives a hard label.
- For example, mass shape can be something between lobulated and irregular.



What no interface can express

0.7 lobulated + 0.3 irregular

- A one-hot input cannot hold a mixture.
- The forced categorical choice discards that mixture — and the mixture is clinically informative.
- There is no semantically transparent way to run what-if analysis over sub-categorical ambiguity.

The gap: clinical knowledge is lost in translation



Descriptors are context-dependent

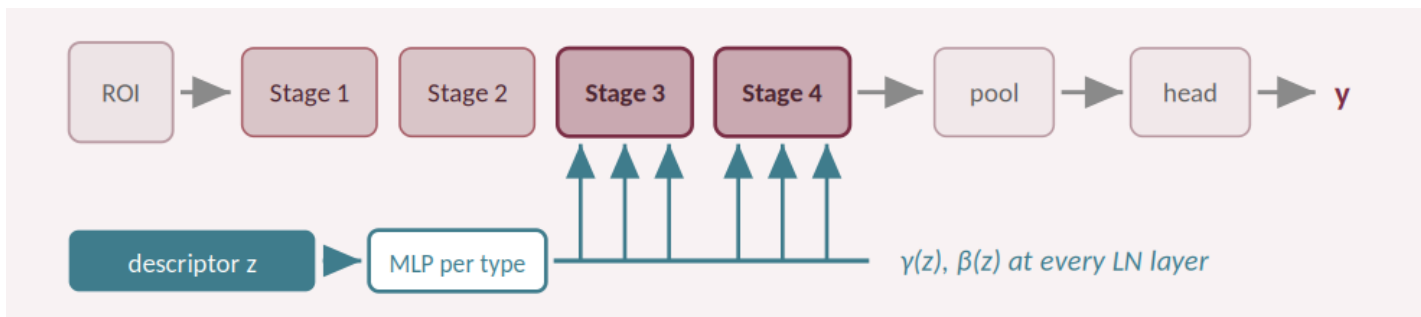
- Descriptors are processed as independent features.
- But their implications depend on what the surrounding image looks like.
- A static embedding is limited.



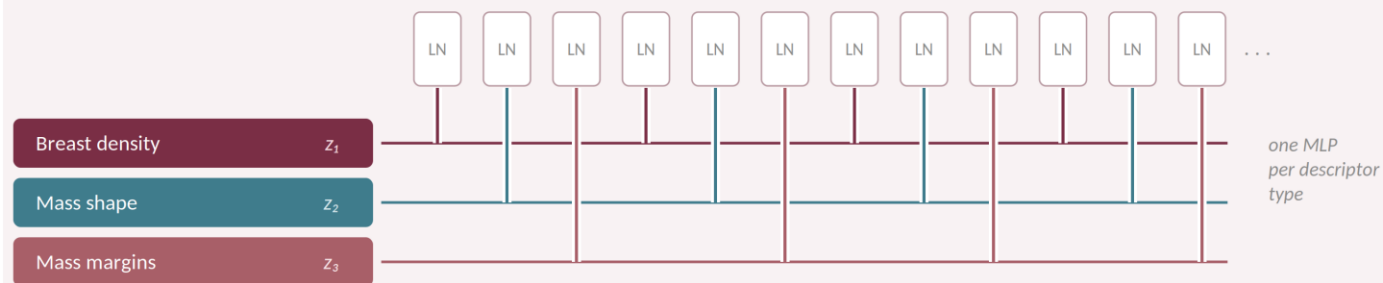
Descriptors have distinct roles

- Density describes the whole breast; margins describe the lesion edge.
- They are relevant at different depths, and each carries a different kind of information.
- A single fused vector can make it harder to learn the distinct role of each descriptor.

SFM — Semantic Feature Modulation



Interleaved conditioning — consecutive LN layers inside a stage



Discrete → continuous

Trained on one-hot descriptors. Queried between them, with no retraining.



Context-conditional

The conditioning axes are BI-RADS descriptors. The modulation parameters depend on the descriptor, while their effect depends on the image features.



Distinct roles

One network per descriptor type, at different depths

Setup: identical backbone and protocol for every method

Few public mammography datasets ship BI-RADS descriptors as metadata

CBIS-DDSM

CMMD

5

independent runs

mean $\pm$ std

6

descriptors

density, shape, margins,
morphology, distribution, lesion
type

BASELINES — ALL (EXCEPT BASE) USE THE SAME DESCRIPTORS

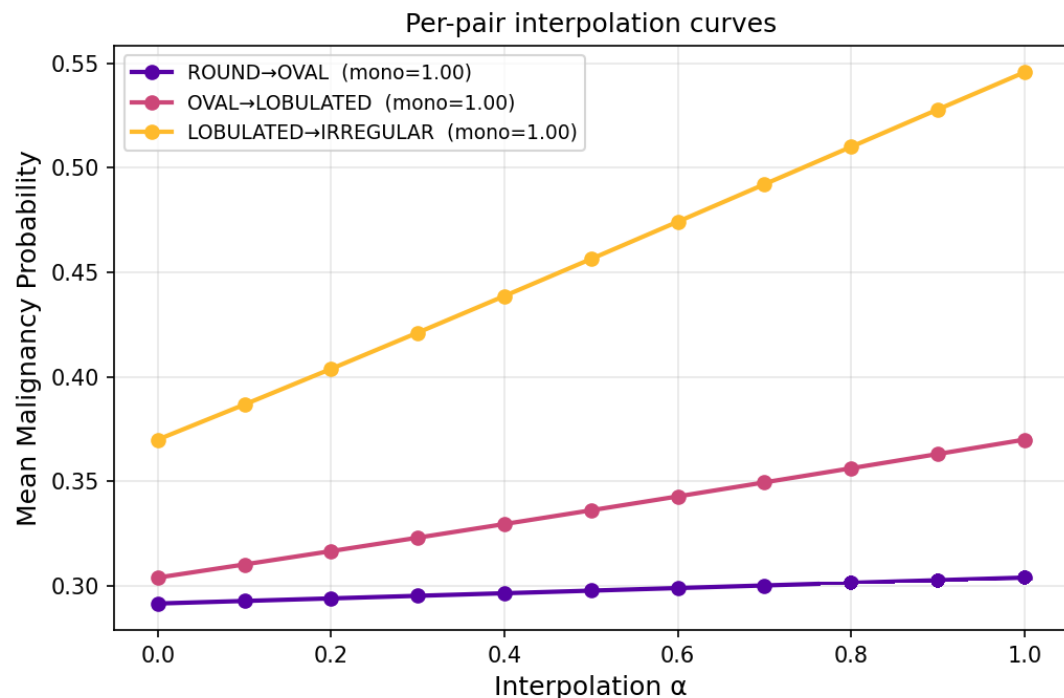
- Base — image only
- Concat — embeddings joined to pooled features
- SE-Gate — descriptor-conditioned channel gating
- MTL — auxiliary descriptor heads

SHARED PROTOCOL

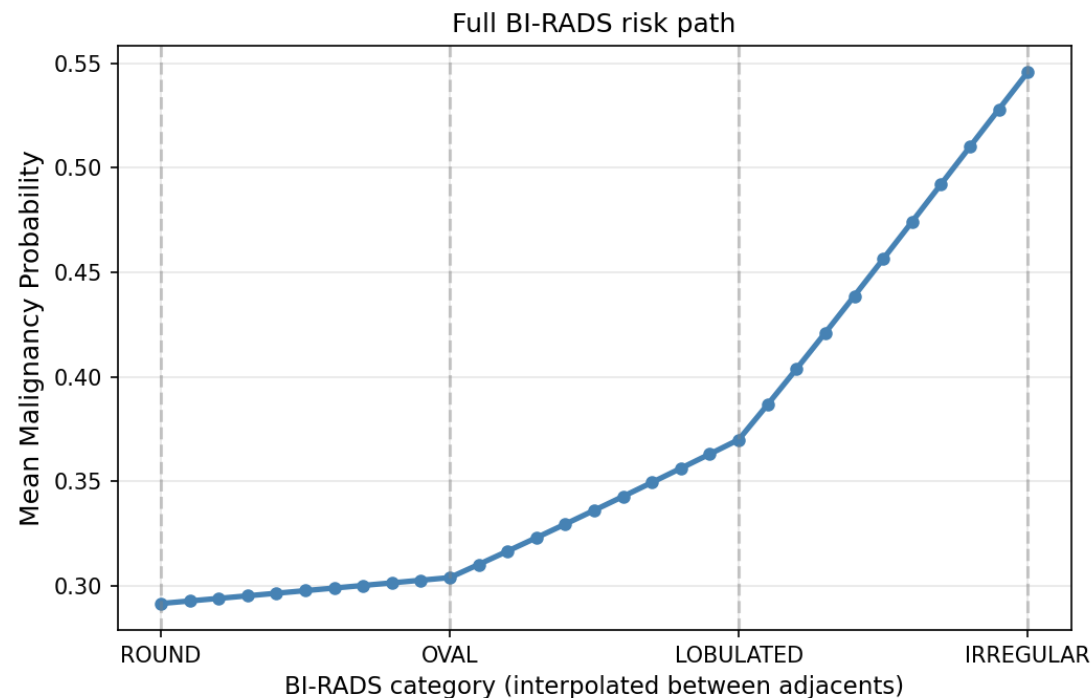
- ConvNeXt-Base, ImageNet-pretrained
- Single-view ROI crops from radiologist annotations
- Task: benign vs. malignant
- Three settings — mass only, calcification only, and joint (one model over both)

Trained on hard labels, queried with soft ones: an ordinal risk geometry nobody supervised

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.



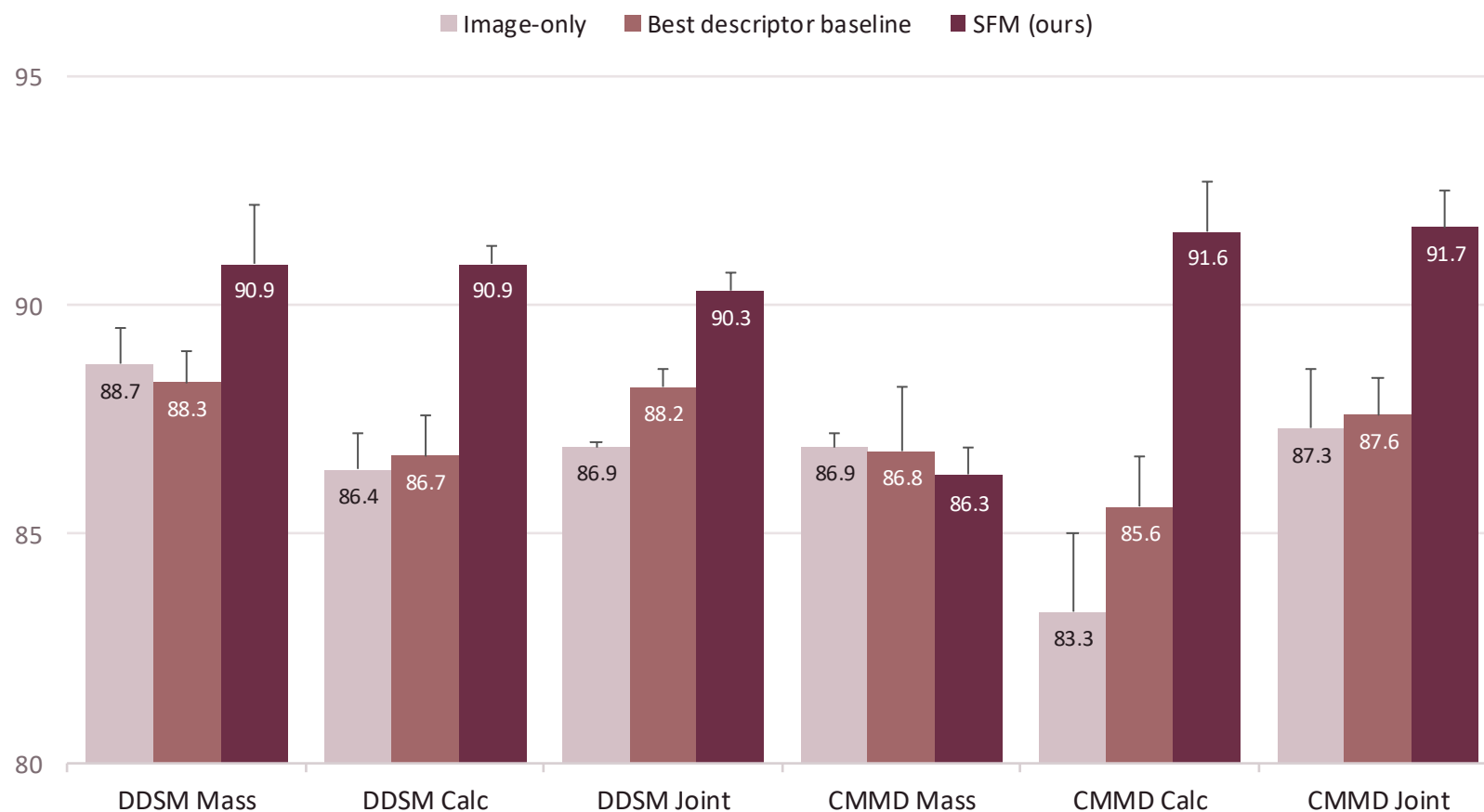
Across the path — uneven spacing. Near-flat ROUND→OVAL→LOBULATED, then a sharp rise into IRREGULAR: an unequal risk scale, learned with no ordinal supervision.

Trained on one-hot labels only. No soft ground truth exists for these mixtures — the interior is a generalization probe.

AUC across both datasets and all three lesion settings

"Best descriptor baseline" = the strongest of Concat / SE-Gate / MTL per setting — a deliberately conservative comparison.

Bars show the mean over 5 runs; whiskers show ± 1 std.



Mean: +3.7pp

Max: +8.3pp

5 / 6

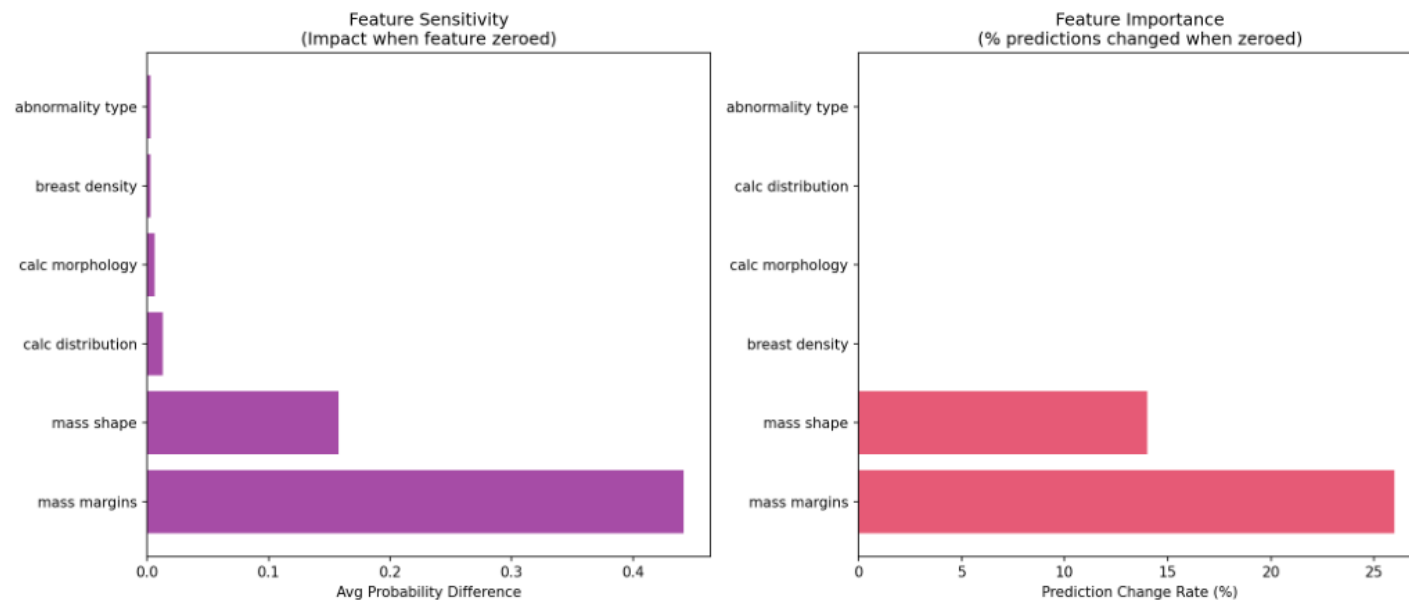
clear top AUC

CMMD mass: 86.3 ± 0.6 vs 86.9 ± 0.3

RESULT

The model relies on the descriptors

Perturbation analysis on the CBIS-DDSM joint model — one network over masses and calcifications together. One descriptor at a time is zeroed at inference, and we measure how far the predicted malignancy probability moves.



Probability shift and prediction-flip rate when each descriptor is zeroed (CBIS-DDSM, joint model)

~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. The model successfully infers them.

LIMITATIONS

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

THE SETTING TODAY



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.



WHAT WE ARE DOING ABOUT IT

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

Key takeaways



TAKE THREE THINGS

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

Thank you

Questions welcome



LOOKING FOR COLLABORATION

Breast radiology and neuroradiology

Shahar Mahpod · Gil Ben-Artzi · Computer Science School, Ariel University

gilba@g.ariel.ac.il

<https://gil-ba.com>

Computing resources provided by the Ariel HPC Center, Ariel University.



github.com/shahrma/semantic-mammo