ARTIFICIAL INTELLIGENCE PRECISION HEALTH INSTITUTE
UNIVERSITY OF HAWAIʻI

# Learning a Clinically-Relevant Concept Bottleneck for Lesion Detection in Breast Ultrasound

HIPIMR website

Read the paper!

**Arianna Bunnell**[1,2], Yannik Glaser[2], Dustin Valdez[1], Thomas Wolfgruber[1], Aleen Altamirano[3], Carol Zamora González[3], Brenda Y. Hernandez[1], Peter Sadowski[2], and John A. Shepherd[1]
[1]University of Hawaiʻi Cancer Center, [2]University of Hawaiʻi at Mānoa, [3]Instituto Radiodiagnóstico Managua, Nicaragua

## Introduction

- AI-enabled breast ultrasound (BUS) has the potential to speed up reading and improve workflow for resource-limited scenarios.
- Explainable AI (XAI) can improve radiologist acceptance of AI-enabled BUS by providing verification and explanation of lesion malignancy decisions, acting as a second reader for BUS exams.
- Concept bottleneck models (CBM) [1] seek to align intermediate model representations with human-defined concepts such that the activation of a particular node in the bottleneck layer indicates concept activation.
- The BI-RADS masses lexicon for BUS is defined by the American College of Radiology [2] to standardize reporting of BUS lesions. The BI-RADS masses lexicon contains 5 properties to describe lesions in BUS: shape, orientation, margin, echo pattern, and posterior features.
- Our overall hypothesis is that CBMs which contain clinically-relevant concepts (BI-RADS masses lexicon) can perform with state-of-the-art accuracy in lesion detection from BUS while allowing radiologist intervention for *steerable* XAI decisions.

## Methods

- We propose to integrate a CBM [1] into a Mask RCNN [3] with a ResNet-101 backbone [4, 5], creating **BI-RADS CBM** (see **Figure 1**). Models are implemented in PyTorch [6] using the Detectron2 [7] library.
- BI-RADS CBM 1) detects a lesion in a BUS image; 2) predicts the BI-RADS masses lexicon; and 3) uses the BI-RADS masses lexicon to predict whether the lesion is cancerous.
- BUS images were collected from the Hawaiʻi and Pacific Islands Mammography Registry (HIPIMR) and cleaned using an automatic preprocessing pipeline [8].
- Data were randomly split into training (70%), validation (10%), and testing (20%) by case-control group (**Table 1**). Cases were matched to controls on BUS machine type and birth year.
- To minimize concept leakage, we train BI-RADS CBM in 3 stages. In Stage 1, the detection backbone network is fine-tuned to detect lesions only. In Stage 2, a classification head is trained to predict the BI-RADS masses lexicon concepts. In Stage 3, the final part of the model is trained to predict cancer from the BI-RADS masses lexicon concepts.

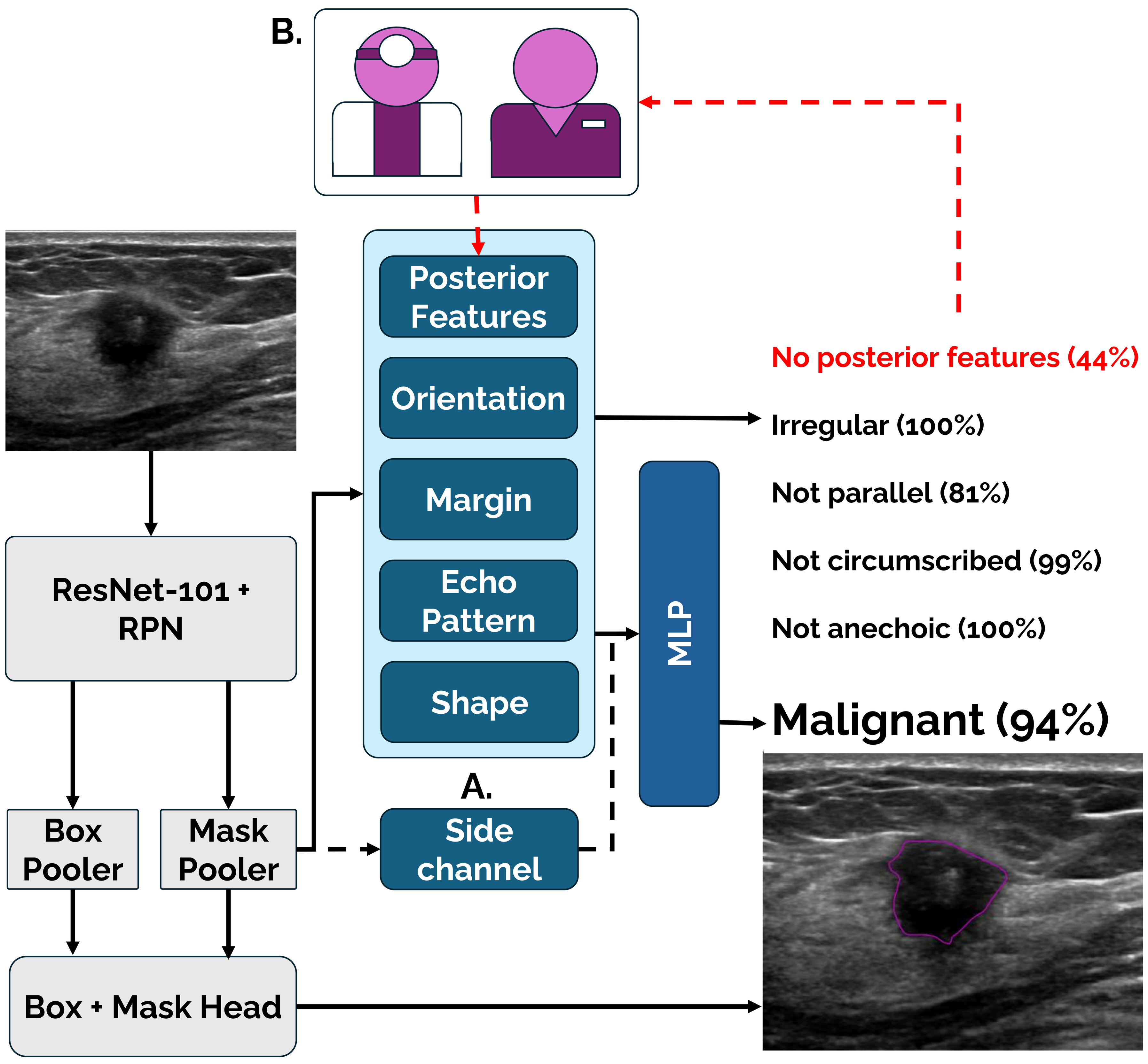


**Figure 1:** An overview of BI-RADS CBM, including the Mask-RCNN underlying structure and the BI-RADS concept bottleneck sub-network. **A.** highlights the side channel, trained for cancer classification only. **B.** highlights concept-level corrections which can be made by the expert reader in the clinic.

| | Overall | Train | Validation | Test |
|---|---|---|---|---|
| Women, N | 994 | 693 | 101 | 200 |
| Women with benign findings, N | 745 | 520 | 75 | 150 |
| Women with malignant findings, N | 249 | 173 | 26 | 50 |
| Mean no. of images/woman | 8.91 | 9.03 | 9.01 | 8.42 |
| Images, N | 8,854 | 6,260 | 910 | 1,684 |
| Images with benign findings, N | 6,555 | 4,587 | 661 | 1,307 |
| Images with malignant findings, N | 2,299 | 1,673 | 249 | 377 |
| Mean no. of lesion views/image | 1.24 | 1.26 | 1.21 | 1.17 |
| Lesion views, N | 5,648 | 4,203 | 573 | 872 |
| Lesion views w/benign findings, N | 3,579 | 2,626 | 369 | 584 |
| Lesion views w/malignant findings, N | 2,069 | 1,577 | 204 | 288 |

**Table 1:** Image-, patient-, and lesion-level counts for all data splits from the HIPIMR.

## Methods (cont.)

- We experiment with cancer head complexity by varying concept combination strategy (linear vs. non-linear) and model interpretability (clinical concepts only vs. with additional side channel).
- For ease of intermediate representation in BI-RADS CBM, we binarize the BI-RADS masses lexicon for each property into those classifications which are either indicative of malignancy or indicative of benignity.
- In experiments on steering with corrected concepts in BI-RADS CBM, concepts are corrected just until the correct class is predicted with either probability 0.51 (minimal) or 0.99 (maximal)

## Results

- The BI-RADS CBM detection backbone detected lesions with AP 0.469 for box-style detections on the testing set.
- BI-RADS CBM classifies the masses lexicon with AUROC 0.616, 0.921, 0.901, 0.842, and 0.916 for posterior features, echo pattern, shape, orientation, and margin, respectively at IOU=0.75.
- The best performing model without accounting for concept correction was the non-linear model with a side channel. See **Table 2.**
- When allowing for concept correction for incorrectly predicted concepts, the best performing model is the linear model with no side channel. See **Table 2.**

| Side channel? | Linear? | Correction? | $AUROC_{0.75}$ |
|---|---|---|---|
| ✗ | ✓ | None | 0.861 |
| ✗ | ✓ | Minimal | 0.885 |
| ✗ | ✓ | Maximal | 0.841 |
| ✗ | ✗ | None | 0.862 |
| ✗ | ✗ | Minimal | 0.874 |
| ✗ | ✗ | Maximal | 0.814 |
| ✓ | ✗ | None | 0.871 |
| ✓ | ✗ | Minimal | 0.872 |
| ✓ | ✗ | Maximal | 0.845 |
| N/A | N/A | N/A | 0.876 |

**Table 2:** Performance characteristics for the cancer classification task, with and without concept correction on the testing set. Gray represents the baseline model.

## Conclusion

- BI-RADS masses lexicon concept intervention is possible on BUS imaging and increases cancer classification performance.
- The complexity of the cancer head and the non-explainable side channel both improved performance when intervention was not permitted. However, the linear cancer head retained the best performance when concepts were corrected at test time.
- CBMs which contain clinically-relevant concepts can perform with state-of-the-art accuracy in lesion detection from BUS

## References

[1] Koh, P.W., et al., Concept Bottleneck Models. (2020). [2] D'Orsi, C., L. Bassett, and S. Feig, Breast imaging reporting and data system (BI-RADS). Breast imaging atlas, 4th edn. American College of Radiology, Reston, (2018). [3] He, K., et al. Mask R-CNN. 2017. arXiv:1703.06870. [4] He, K., et al. Deep residual learning for image recognition. PROCEEDINGS OF THE IEEE CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION. 2016. [5] Lin, T.-Y., et al. Feature Pyramid Networks for Object Detection. 2017 IEEE CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION (CVPR). 2017. IEEE. [6] Paszke, A., et al., PyTorch: An Imperative Style, High-Performance Deep Learning Library, H. Wallach, et al., Editors. 2019. [7] Wu, Y., et al., Detectron2. https://github.com/facebookresearch/detectron2. [8] Bunnell, A., et al., BUSClean: Open-source software for breast ultrasound image pre-processing and knowledge extraction for medical AI. arXiv, (2024). https://hipimr.shepherdresearchlab.org/