UNIVERSITY OF HAWAI'I
CANCER CENTER

# Concept Bottleneck Models with Expert Ontologies for the Diagnosis of Cancer from Medical Imaging

Dissertation proposal defense by Arianna Bunnell

NCI Cancer Center
A Cancer Center Designated by the National Cancer Institute

# Overview

- Explainable AI (XAI) is valuable for high-risk scenarios such as medical diagnosis.
- We may consider clinical users of diagnostic AI models as *domain* experts.
- Concept-based XAI can provide a priori ontological alignment between expert users and AI models.
- How can we tune the use of concepts to provide maximum explainability, faithfulness, and performance?

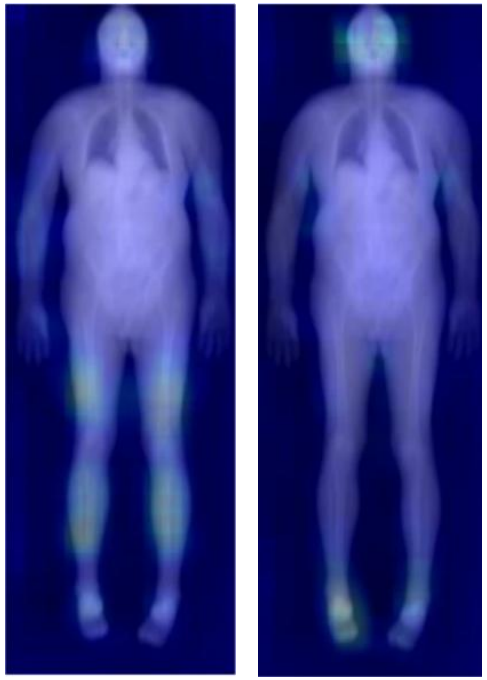
# Proposal Contents

- Background
  - XAI for domain experts
  - Concept-based XAI
- Research Objectives
- Proposed tasks
  - Malignancy classification on mammography
  - Lesion detection on breast ultrasound
  - Segmentation and classification on dermascopy
- Outcomes
- Timeline

# XAI for *Domain* Experts

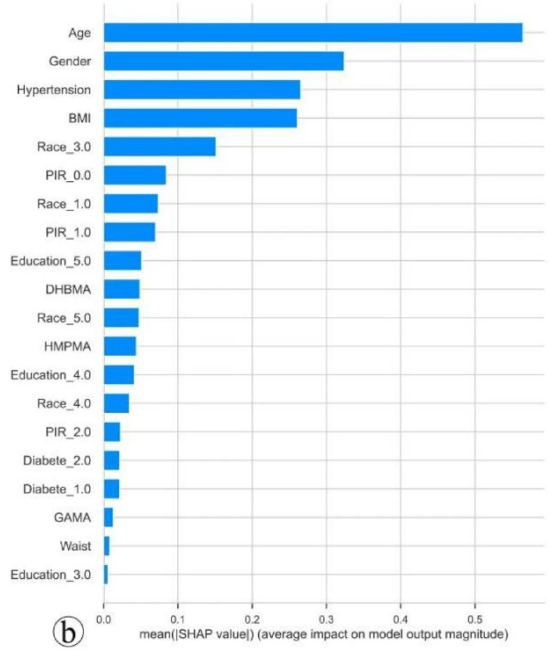
- Many XAI methods are designed for a) the layperson or b) without considering an explicit audience.

**GradCAM**



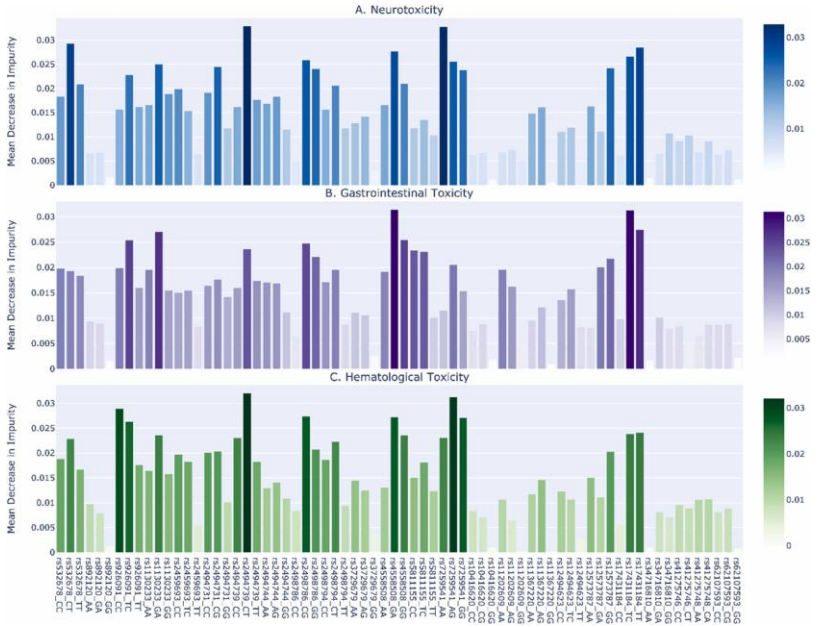
Glaser, et al. *2022.*
10.1038/s43856-022-00166-9

**SHAP**




Zheng, et al. *2025.*
10.1038/s41598-025-23050-7

**Method-specific (MDI)**




Guo, et al. *2023.*
https://doi.org/10.1016/j.biopha.2023.114518

# Dichotomous Scales of XAI

| | | | |
|---|---|---|---|
| **Global** | **Intrinsic** | **Model-specific** | ***Expert*** |
| Explanation of the entire model | Baked into the model design | Only valid for a single model style | Explanations interpretable for domain experts |
| Explanation of a single example or even single region of an example | Use a separate model/method after training | Can be applied to many different model types | Explanations interpretable for anyone |
| **Local** | **Post-hoc** | **Model-agnostic** | ***Lay*** |



Adadi, A., & Berrada, M. (2018). Peeking inside the black-box: a survey on explainable artificial intelligence (XAI). *IEEE access*, *6*, 52138-52160.

# Concept-based XAI

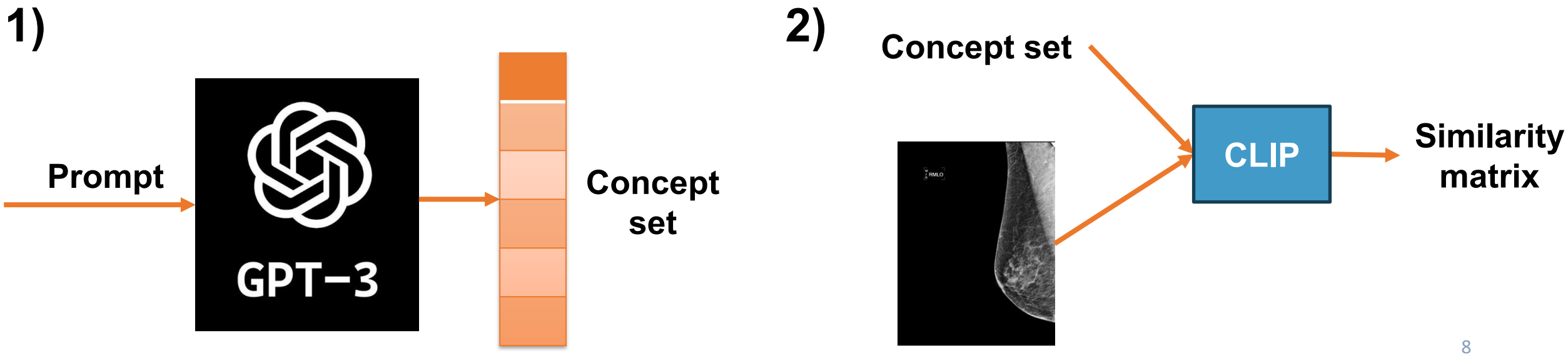
- Concept-based XAI methods seek to align the decision-making process of an AI model with semantically meaningful *concepts.*
- Concept-based methods are well-suited for expert use.

# Concept-based XAI

- $f: X \rightarrow Y$ is standard supervised learning.
- In concept models, $f(X) = (h \circ g)(X)$
  - $g: X \rightarrow L$ where $L$ is a representation of concepts
  - $h: L \rightarrow Y$ where $Y$ are class labels.
- Users can optionally modify representations $L$ at test-time to **steer** model predictions $Y$ in intrinsic methods.

# Concept Discovery Methods

- Concept discovery methods uncover dataset- and task-specific concepts within the latent space in an unsupervised fashion, without requiring predefined concept labels.
- Not directly applicable for expert use due to the lack of clear alignment with existing expert ontological systems.

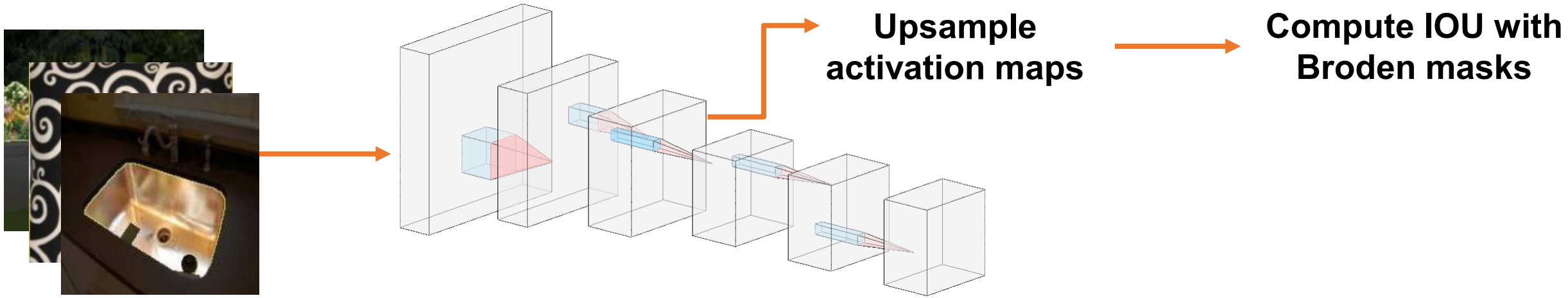





Oikarinen, T., Das, S., Nguyen, L. M., & Weng, T. W. (2023). Label-free concept bottleneck models. *arXiv preprint arXiv:2304.06129*.

# Probe-based Methods

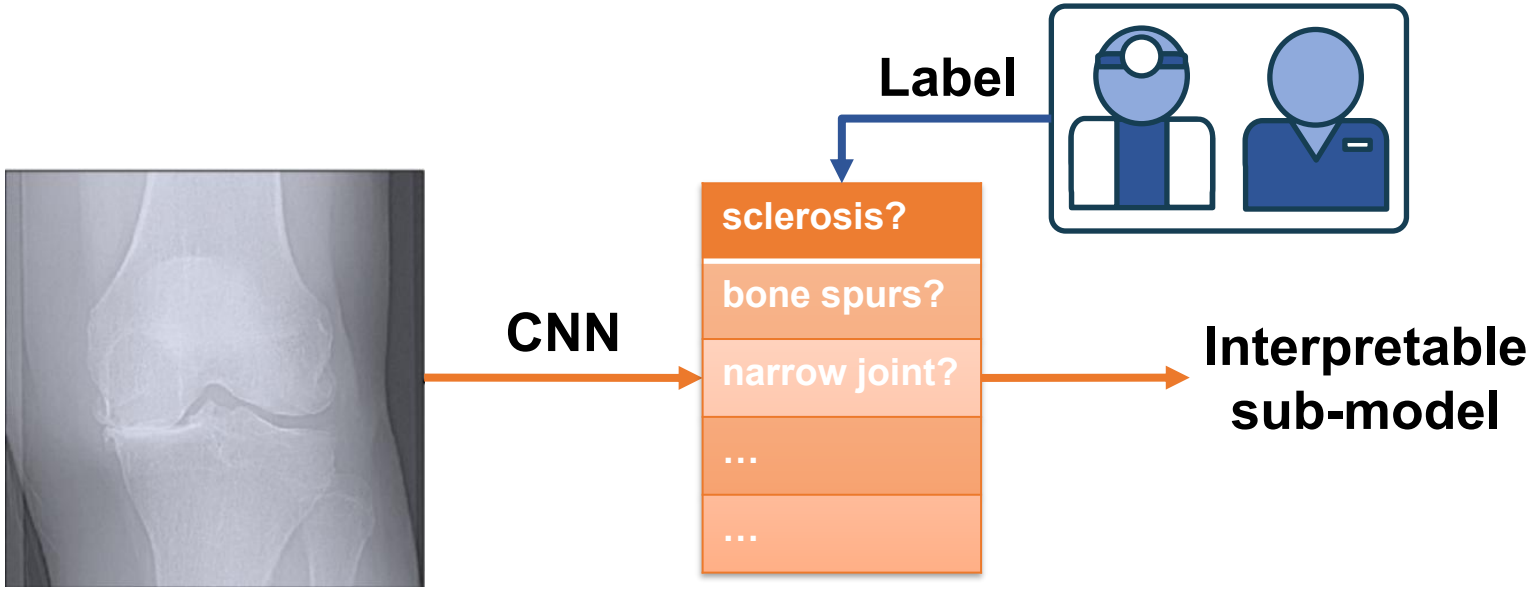
- Probe-based methods use representations from a probe dataset which is labeled with concepts and transfer them to the dataset of interest.
- Resulting explanations are usually not steerable and limited to general explanations (shapes, colors, etc.)




Bau, D., Zhou, B., Khosla, A., Oliva, A., & Torralba, A. (2017). Network dissection: Quantifying interpretability of deep visual representations. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 6541-6549).

# A priori Concept Methods

- A priori concept methods use concept labels on the dataset of interest to train in a supervised or semi-supervised fashion.
- Most directly useful for expert use, but do not circumvent any labeling cost like probe-based or concept discovery methods.

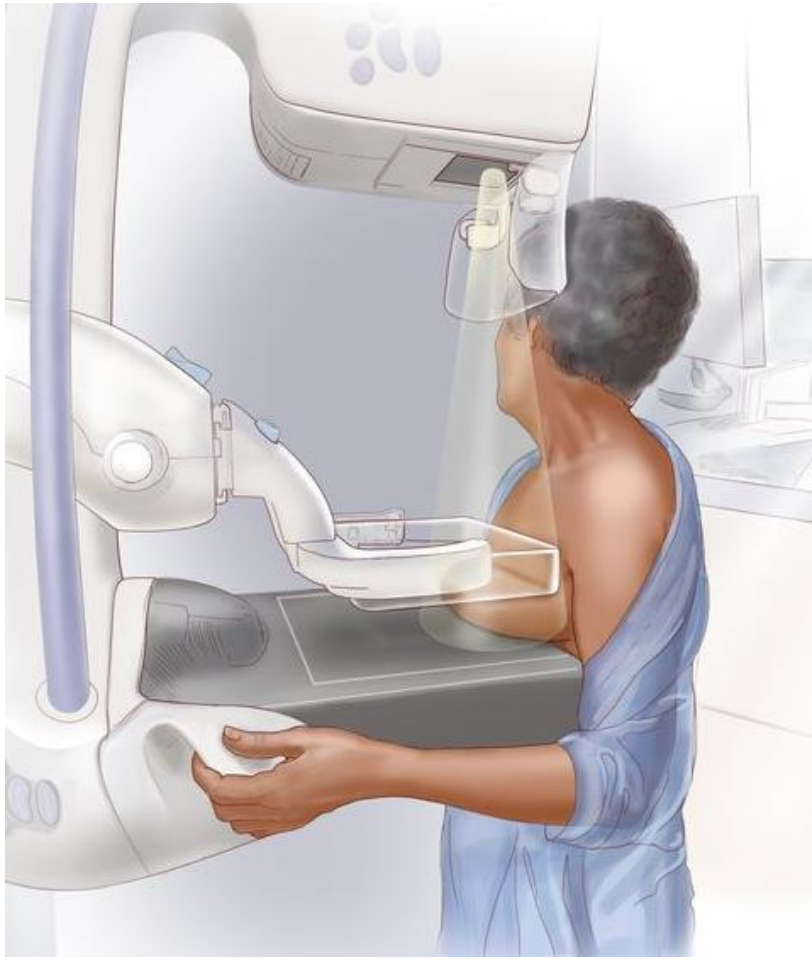



Koh, P. W., Nguyen, T., Tang, Y. S., Mussmann, S., Pierson, E., Kim, B., & Liang, P. (2020, November). Concept bottleneck models. In *International conference on machine learning* (pp. 5338-5348). PMLR.

# Task I – Mammography


https://www.openmed.co.in

- We investigate the use of clinical expert CBMs for mass and calcification localization and exam-level diagnosis on mammography imaging.
- Mammography imaging is obtained by passing a low-dose X-ray being passed through compressed breast tissue.
- Non-suspicious fatty tissue typically appears radiolucent on the mammogram and higher-density glandular tissue or masses appear brighter white.

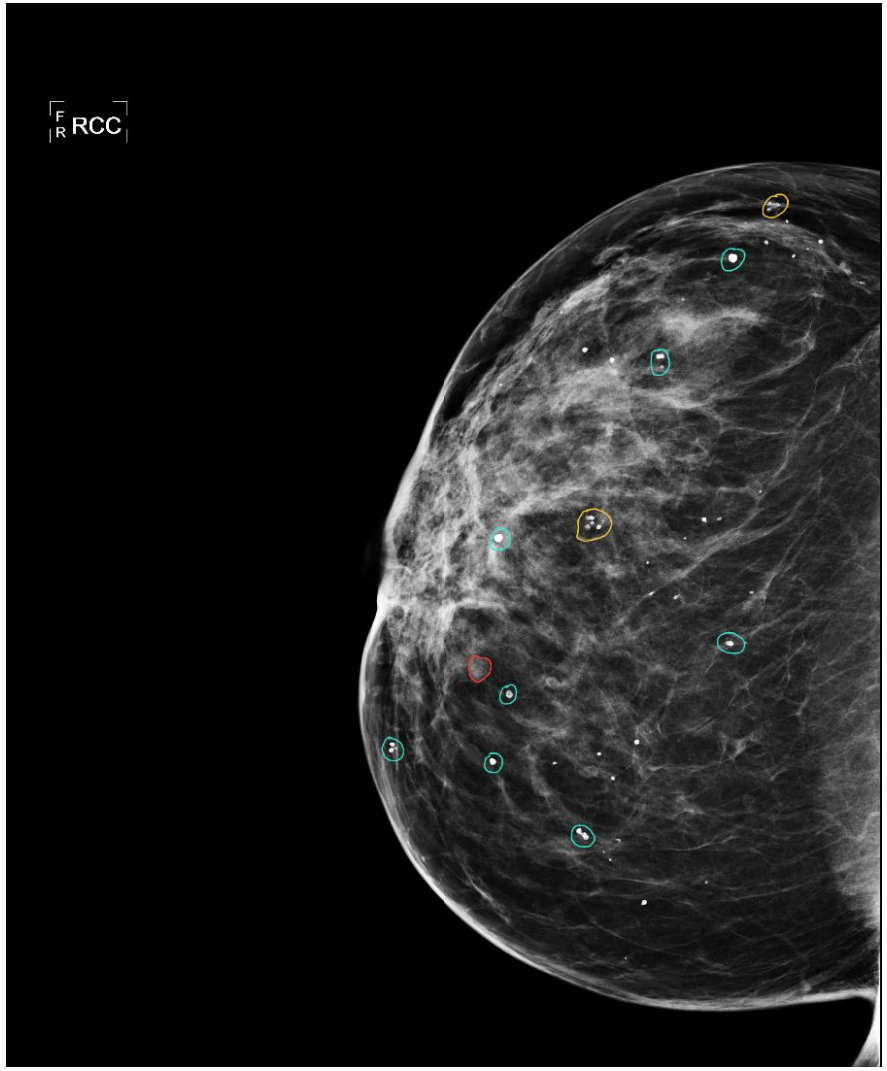


# Concept Set I – Mammography

- Masses and calcifications on mammography imaging are classified for reporting according to the ACR BI-RADS Masses and Calcifications lexica.

| Category | Attribute | Categories |
|---|---|---|
| Masses | Shape | Oval; Round; Irregular |
| | Margin | Circumscribed; Obscured; Microlobulated; Indistinct; Spiculated |
| | Density | High density; Equal density; Low density; fat-containing |
| Calcifications | Morphology | Skin; Vascular; Coarse or “popcorn-like”; Round; Rim; Dystrophic; Milk of calcium; Suture; Amorphous; Coarse heterogeneous; Fine pleomorphic; Fine linear branching |
| | Distribution | Diffuse; Regional’; Grouped; Linear; Segmental |
| Architectural distortion | | Yes; No |
| Skin retraction | | Yes; No |

# Examples I – Mammography




| | |
|---|---|
| **Shape** | Round |
| **Margin** | Microlobulated |
| **Density** | Equal |

| | |
|---|---|
| **Morphology** | Dystrophic |
| **Distribution** | Clustered |

| | |
|---|---|
| **Morphology** | Coarse |
| **Distribution** | Diffuse |

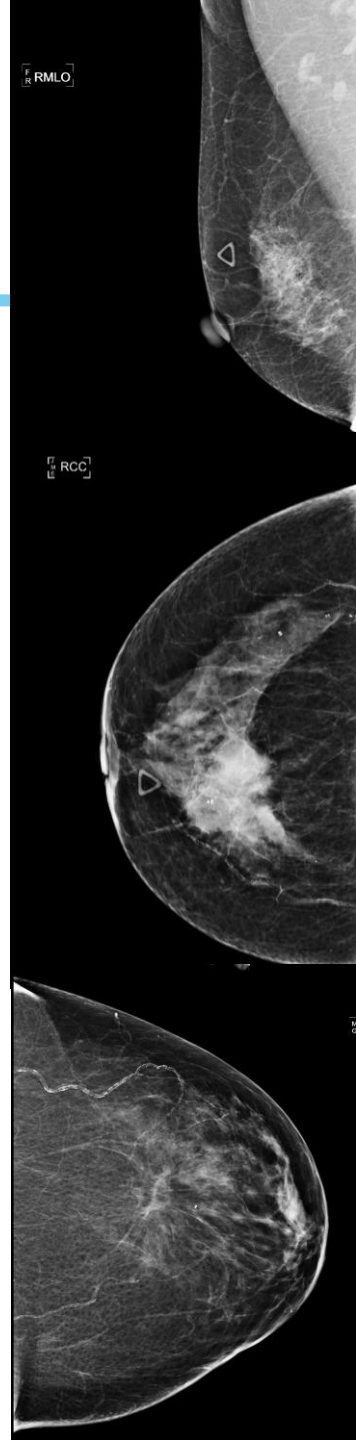
# Examples I – Mammography


| | |
|---|---|
| **Morphology** | Vascular |
| **Distribution** | Linear |

| | |
|---|---|
| **Morphology** | Coarse |
| **Distribution** | Clustered |

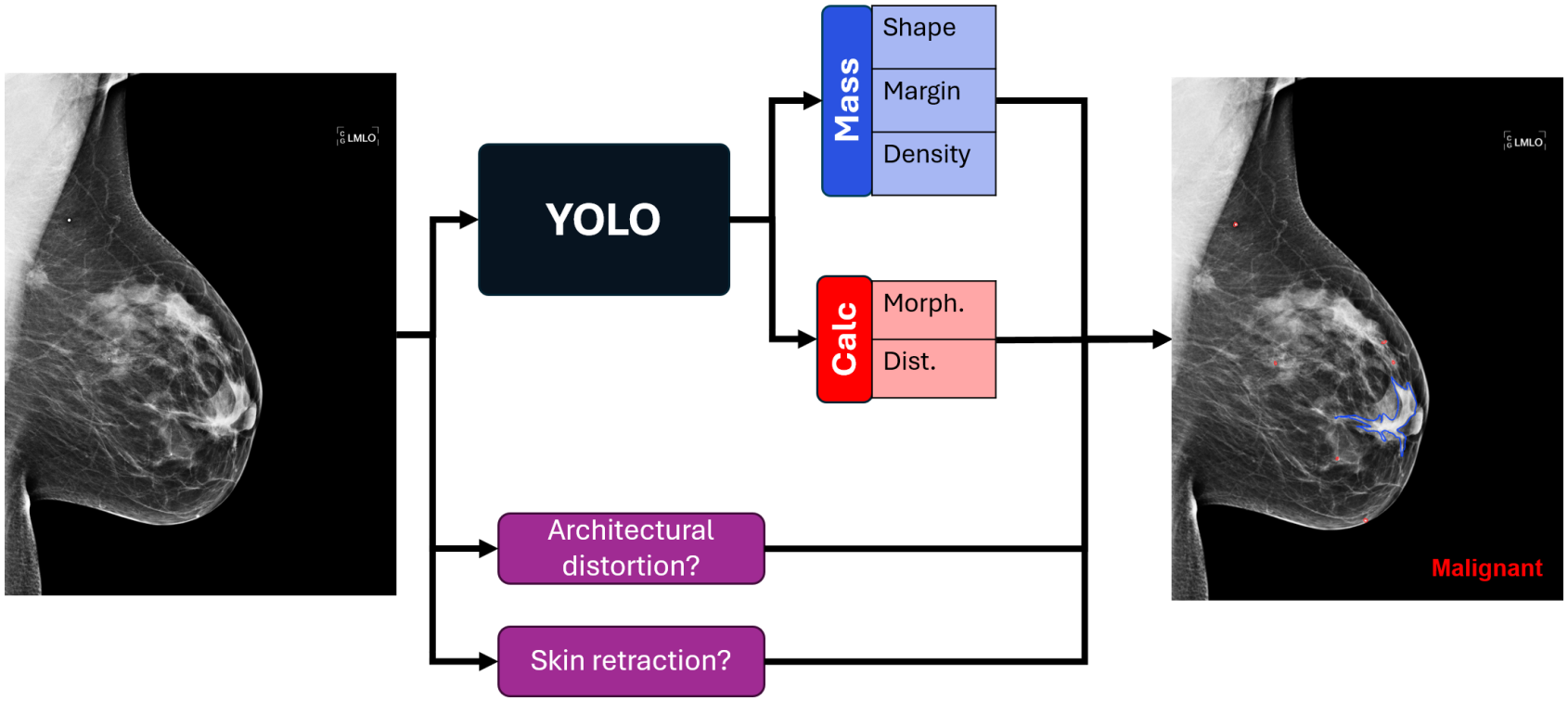
# Dataset I – Mammography

- Data from Hawai‘i and Pacific Islands Mammography Registry.
- 771 cases and 2,313 matched controls.
- Matched by patient birth year, clinical breast density, and mammography machine type.
- Annotations provided by consulting radiologist and consulting oncologist.
- Labels have been generated for 1,500 women.



# Method I - Mammography

- Combine multiple object-level concepts with image-level concepts to come to a final explainable image-level prediction.



# RQ I - Mammography

- What style of combination of object-level concepts into a single exam-level classification yields the best performance results?
- How does the choice of combination method effect the optimal choice of concept intervention style?

- *Novelty:* first CBM for mammography imaging and also first investigation into combination of multi-level concepts.

# Task II – Breast US

- We investigate the use of clinical expert CBMs for breast cancer detection and diagnosis on breast ultrasound (US) imaging.
- Breast US imaging displays the relative propagation, scatting, and reflection of sound waves through breast tissue.
- Boundaries between tissue types produce echoes which are reflected back to the transducer and converted into imaging.

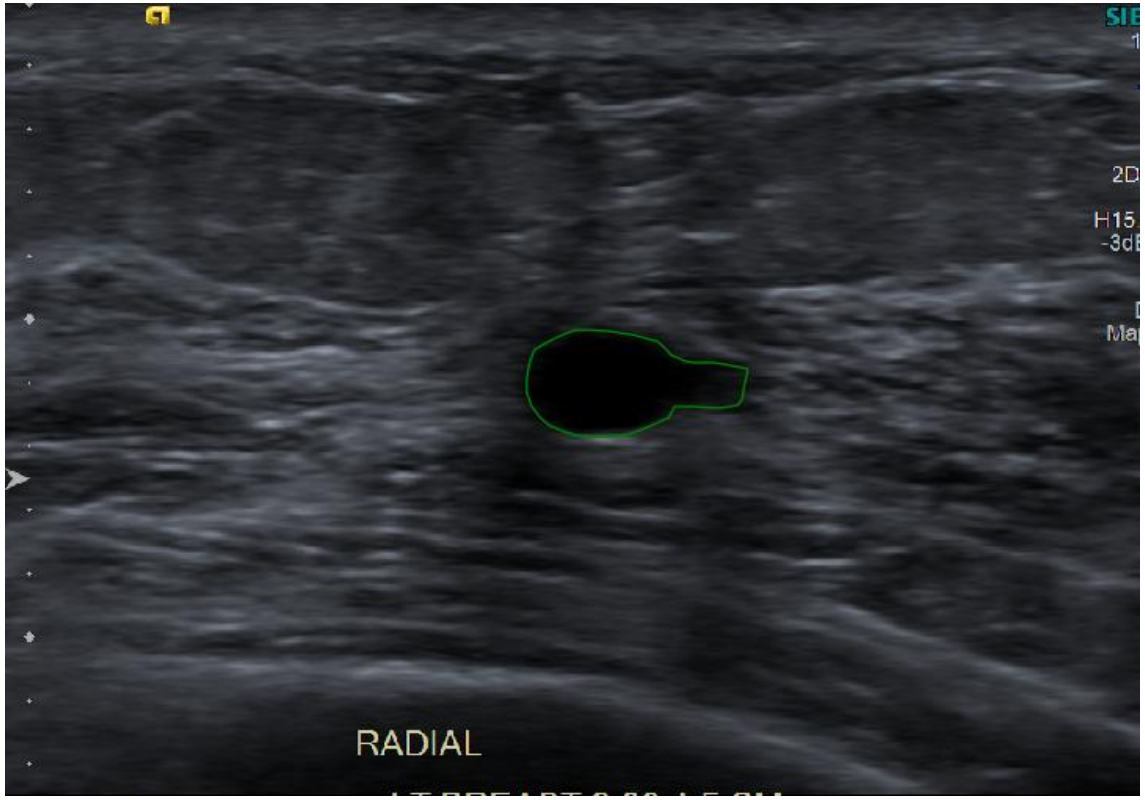
Muro Paz V. M. C., Hammond Castro R., Falcon L. (2017). Breast US technique and BIRADS lexicon for "beginners". C-2030. ECR 2017..

# Concept Set II – Breast US

- Lesions on breast ultrasound imaging are classified for reporting according to the ACR BI-RADS Masses lexicon for ultrasound.

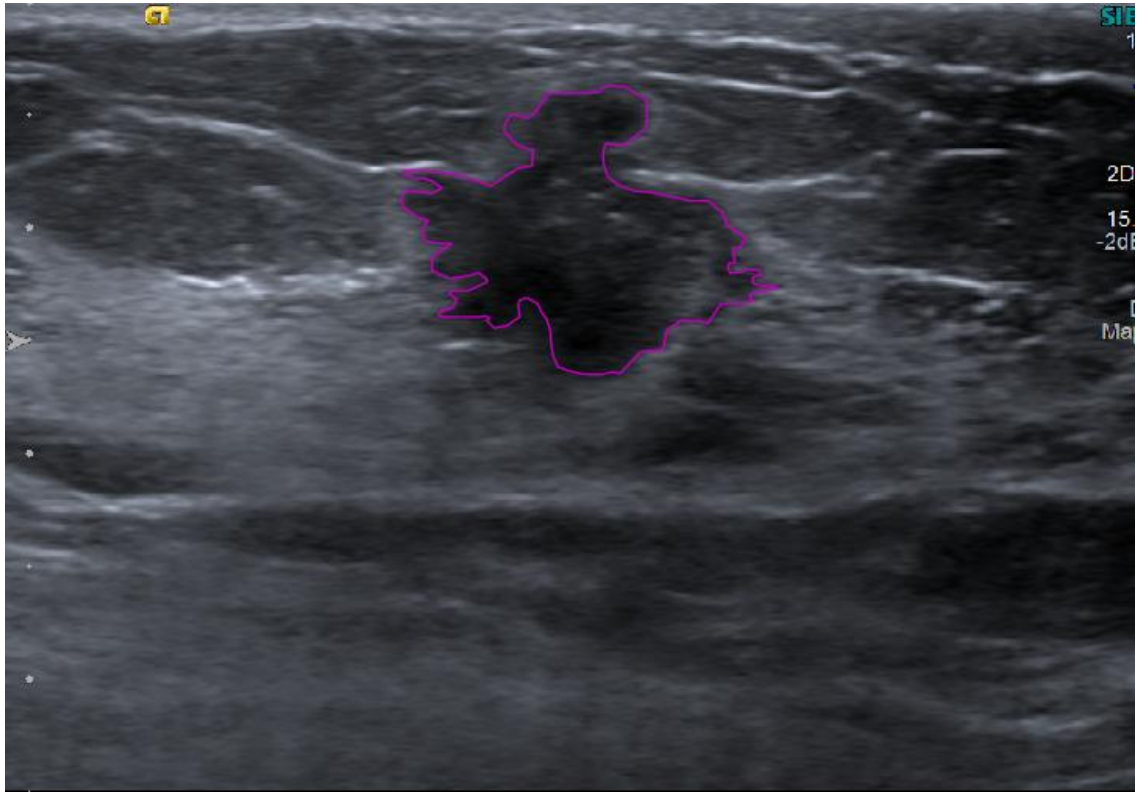
| Attribute | Categories |
|---|---|
| Shape | Oval; Round; Irregular |
| Orientation | Parallel; Not parallel |
| Margin | Circumscribed; Indistinct; Angular; Microlobulated; Spiculated |
| Echo pattern | Anechoic; Hyperechoic; Complex cystic and solid; Hypoechoic; Isoechoic, Heterogeneous |
| Posterior features | No posterior features; Enhancement; Shadowing; Combined pattern |

# Examples II – Breast US




| | |
|---|---|
| **Shape** | Oval |
| **Orientation** | Parallel |
| **Margin** | Circumscribed |
| **Echo pattern** | Anechoic |
| **Posterior features** | No posterior features |

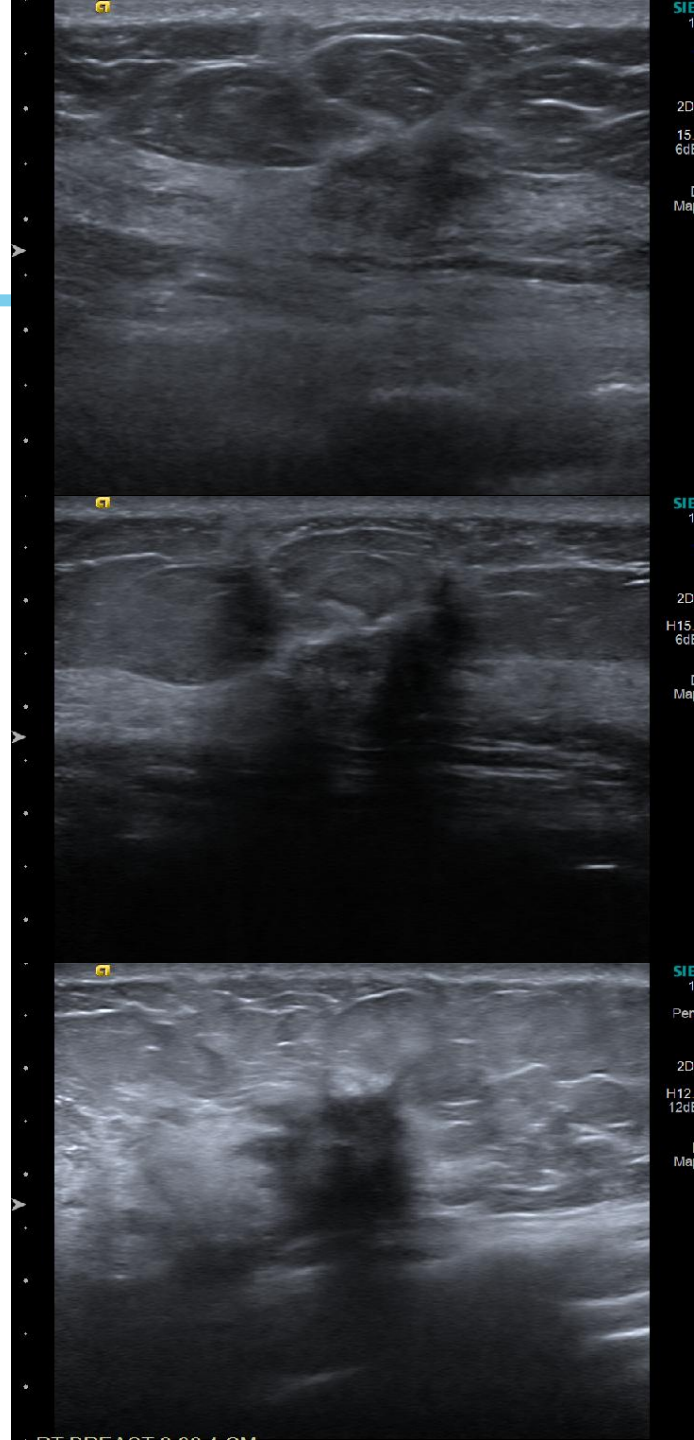
# Examples II – Breast US



| | |
|---|---|
| **Shape** | Irregular |
| **Orientation** | Not parallel |
| **Margin** | Microlobulated |
| **Echo pattern** | Heterogeneous |
| **Posterior features** | Enhancement |

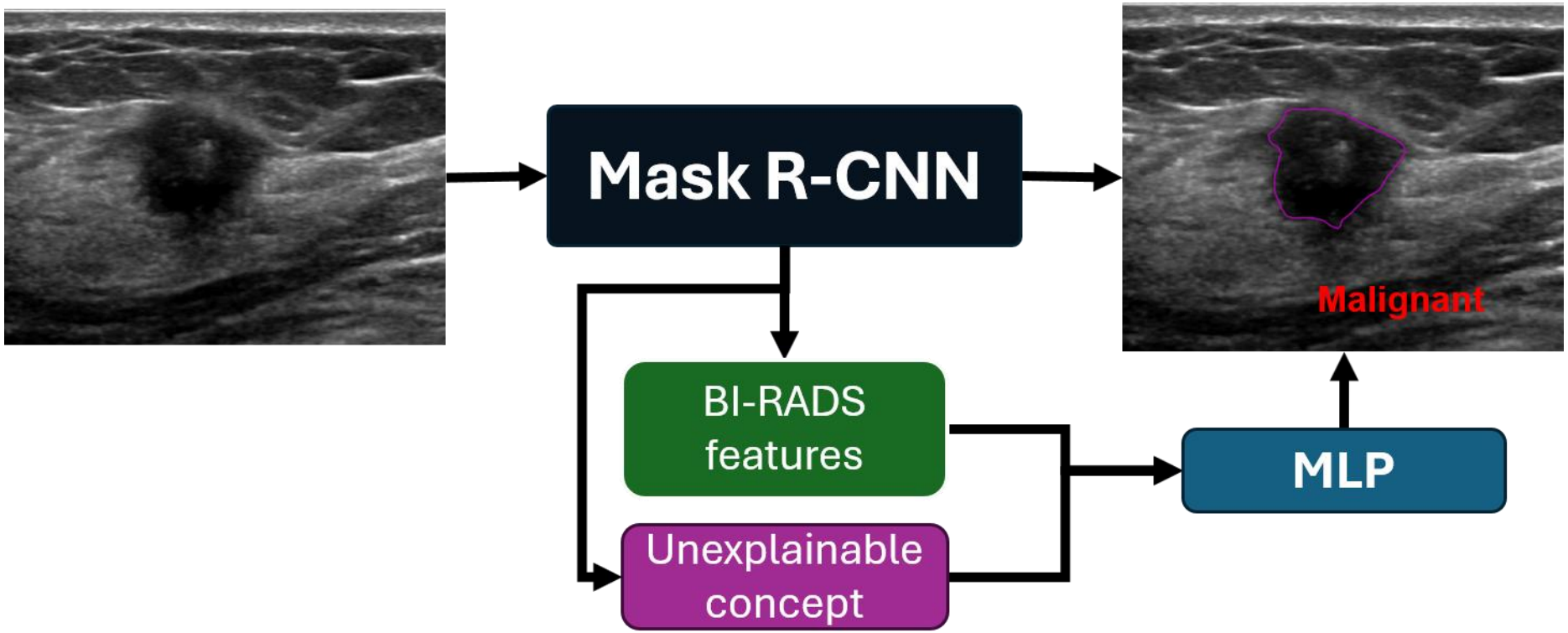
# Dataset II – Breast US

- Data from Hawai‘i and Pacific Islands Mammography Registry.
- 994 women with 8,854 images.
- Matched by patient birth year and ultrasound machine type.
- Split 70% train, 10% valid, 20% test.
- Annotations provided by consulting radiologist and verified by second consulting radiologist on ~10% of the included scans.



# Method II – Breast US

- Integrate CBM into established object detection architecture for object-level concept-based explanations.



# RQ II – Breast US

- How do various simulated concept intervention schemes effect the final classification output of the model?
- Can the choice of concept intervention scheme “save” the performance of a simpler model?
- *Novelty:* first CBM for breast US imaging and also first experimental evaluation into clinical concept evaluation.

# Task III – Dermascopy

- We investigate the use of clinical expert CBMs for skin cancer segmentation and diagnosis on dermascopy imaging.
- Dermascopy imaging is obtained by examining a skin lesion with a dermascope using a liquid or gel interface on the skin.
- Supplements naked eye examination of lesions and allows the dermatologist to see subsurface structures.


https://dermnetnz.org/cme/dermoscopy-course

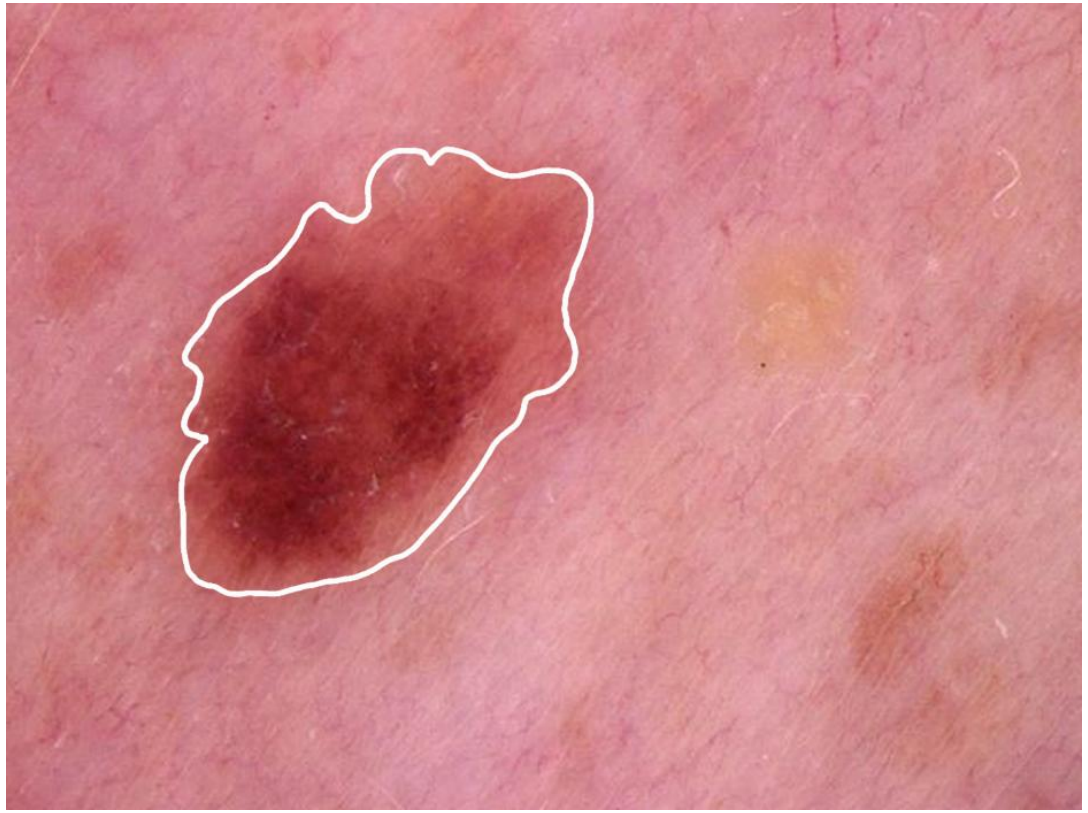
https://www.fotofinder.de/en/specialties/dermoscopy

# Concept Set III – Dermascopy

- Lesions on dermascopy imaging are classified for reporting according to the IDS ABCD criteria for dermascopy.

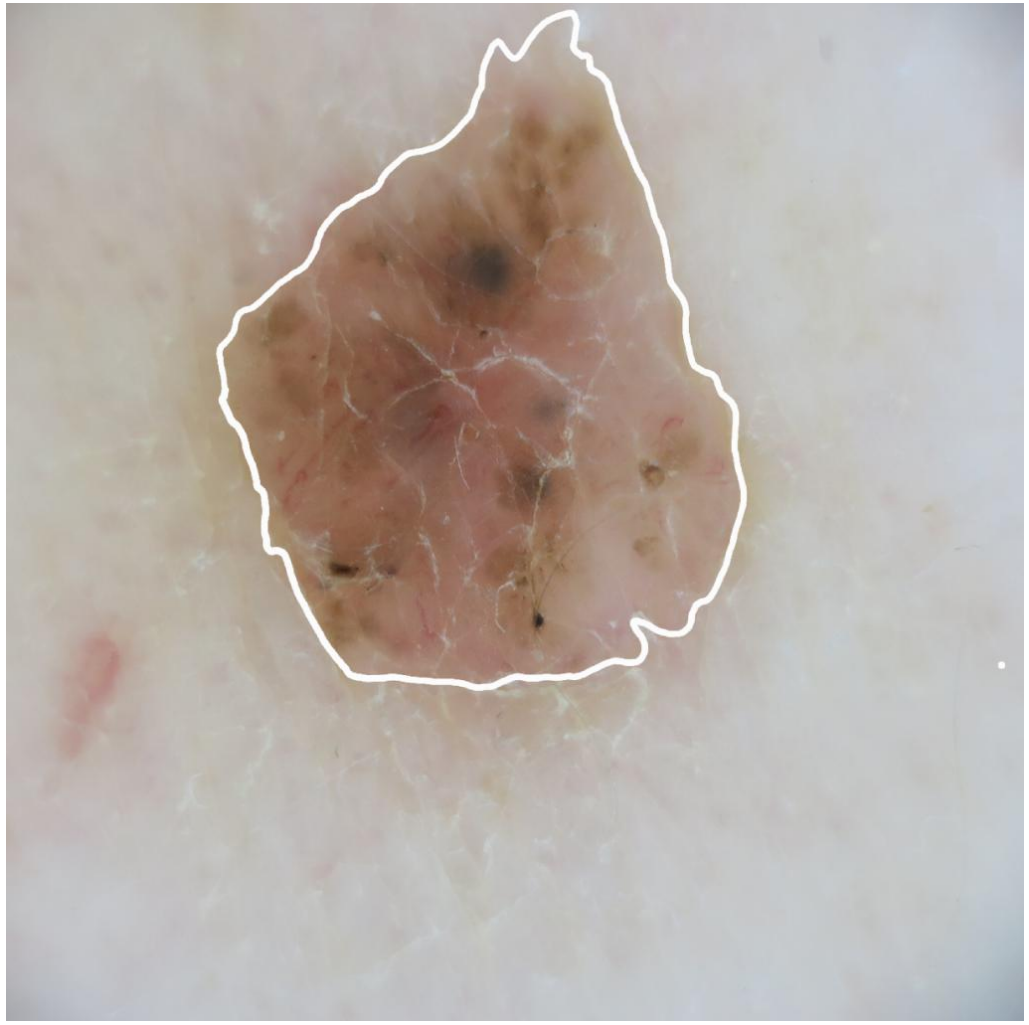
| Attribute | Categories |
|---|---|
| Asymmetry | 0 (no asymmetry); 1 (single axis); 2 (both axes) |
| Border | 0 (no octants with irregular pigment pattern cutoff); 1 – 7; 8 (entire lesion border displays irregular pigment pattern cutoff) |
| Color (count) | White; Red; Light brown; Dark brown; Blue-gray; Black |
| Dermascopic structures (count) | Structureless areas, Pigment networks; Branches streaks; Dots; Globules |

# Examples III - Dermascopy



| | |
|---|---|
| **Asymmetry** | 1 |
| **Border** | 0 |
| **Color** | 2 |
| **Dermascopic structures** | Pigment network<br>Dots |

# Examples III - Dermascopy



| | |
|---|---|
| **Asymmetry** | 2 |
| **Border** | 4 |
| **Color** | 5 |
| **Dermascopic structures** | Pigment network<br>Dots<br>Globules |

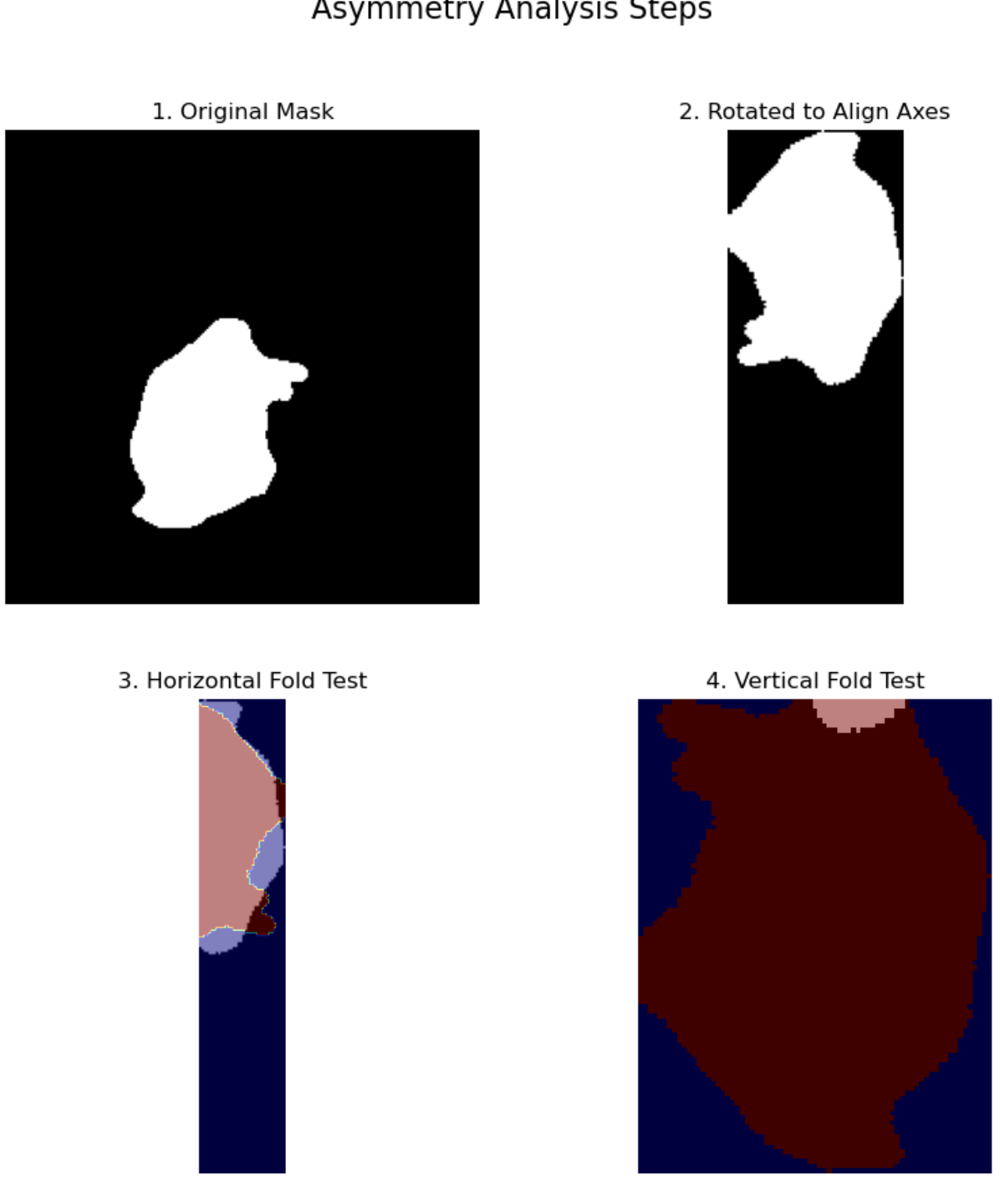
# Computational ABC

- *Stoecker et al. (1992)* propose an asymmetry measure from the lesion mask.
- *Kasmi et al. (2016)* propose a border method from the lesion mask and image.
- *Majumder et al. (2019)* propose a color method from the lesion mask and image.

Stoecker, W. V., Li, W. W., & Moss, R. H. (1992). Automatic detection of asymmetry in skin tumors. *Computerized Medical Imaging and Graphics*, *16*(3), 191-197.
Majumder, S., & Ullah, M. A. (2019). Feature extraction from dermoscopy images for melanoma diagnosis. SN Applied Sciences, 1(7), 753.
Kasmi, R., & Mokrani, K. (2016). Classification of malignant melanoma and benign skin lesions: implementation of automatic ABCD rule. IET Image Processing, 10(6), 448-455.

# Asymmetry

1. Align axes of image with principal axes of lesion mask.
2. Align lesion centroid with image center.
3. “Fold” lesion along both principal axes.
4. Measure non-overlapping proportion.

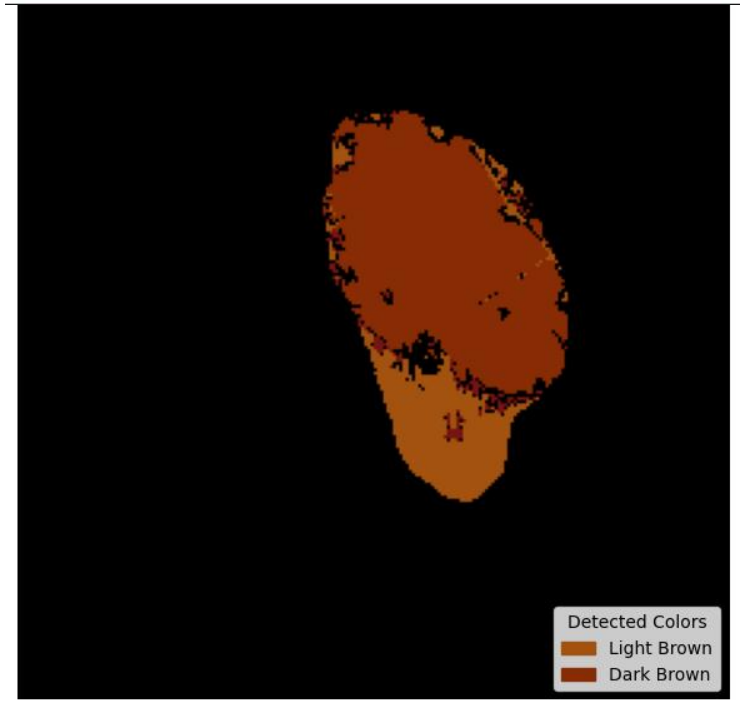


# Border

1. Convert lesion image to grayscale.
2. Calculate the mean grayscale difference along vector normal to border at each point.
3. Average means per octant.

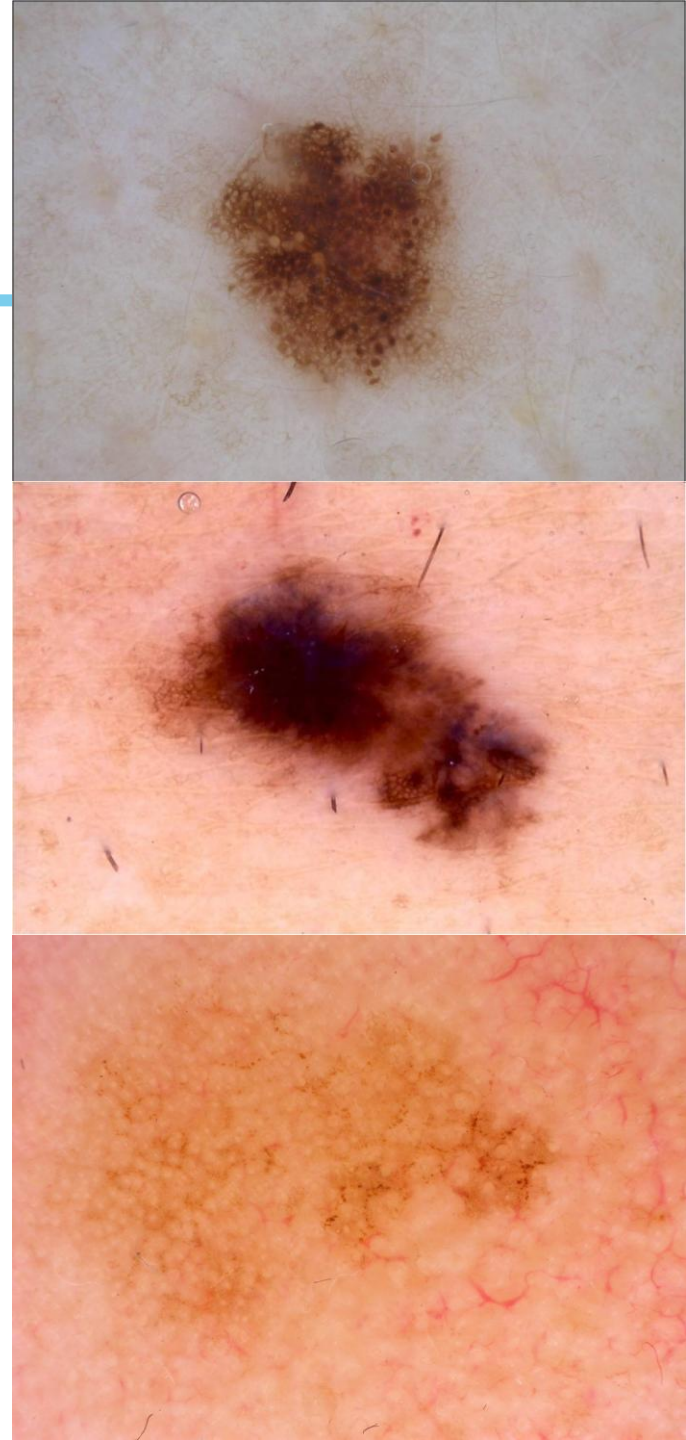
# Color

1. Erode lesion mask to exclude any non-lesion skin area.
2. Compare pixel values with predefined soft color ranges.
3. If more than 5% of the lesion area is a certain color, then that color is detected.



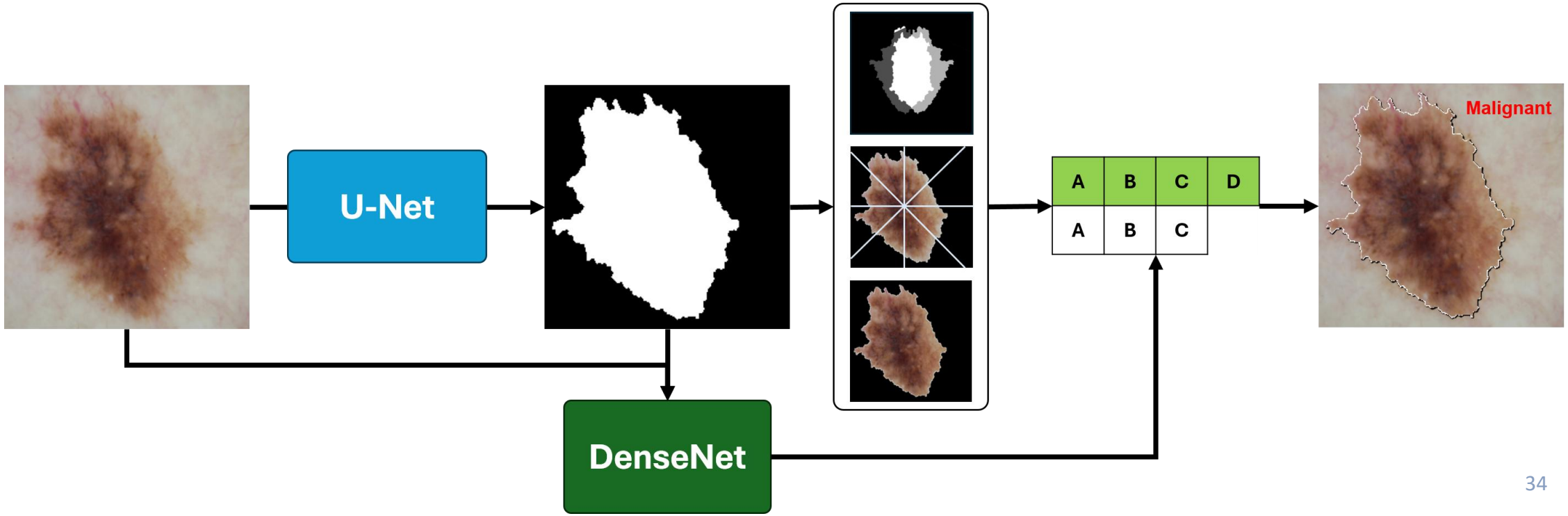
# Dataset III – Dermascopy

- Data International Skin Imaging Collaboration (ISIC) Archive and the PH2 dataset
- 9,638 image and mask pairs.
- 8,170 image/mask pairs with cancer status ground truth.
- 6,370 controls and 1,800 cases.
- Annotations provided by consulting dermatologist for 4,000 images, with ~20% verified by consulting oncologist.



# Method III – Dermascopy

- Combine concepts learned from experts with computationally-defined concepts in standard CBM architecture.



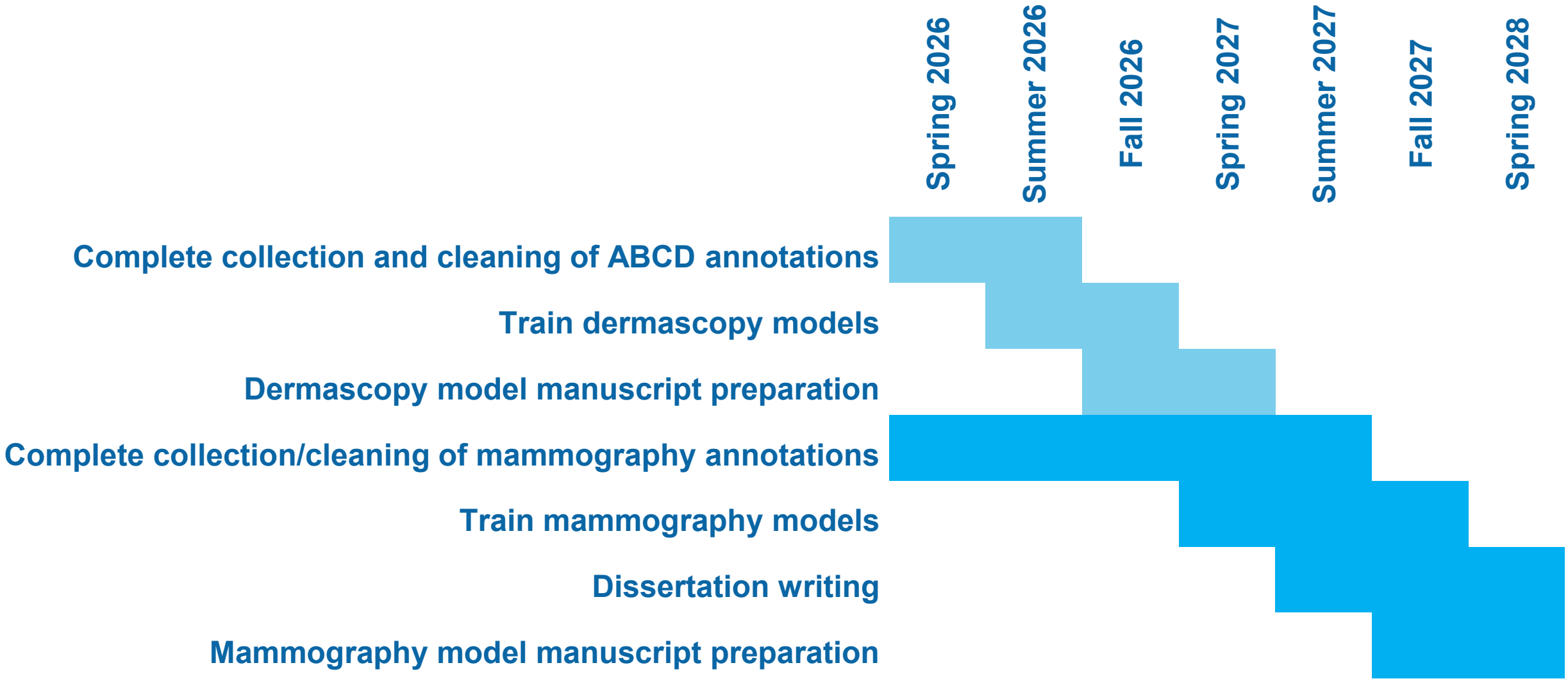
# RQ III - Dermascopy

- Does the addition of objective, computational versions of clinical concepts enhance the performance of concept models?
- Can we isolate certain areas where expert judgement differs from the objective definitions?
- *Novelty:* first mixture of computational/objective concepts with expert-defined concepts to isolate expert knowledge contribution.

# Outcomes

- Three trained concept models will be released.
- A single dataset with expert concept annotations (dermascopy) will be released.
- Add to literature by investigating more complex methods of using, combining, and intervening on concepts.
- Lead to future work with user studies investigating how expert users interact with concept models.

# Timeline


Spring 2026
Summer 2026
Fall 2026
Spring 2027
Summer 2027
Fall 2027
Spring 2028
Complete collection and cleaning of ABCD annotations
Train dermascopy models
Dermascopy model manuscript preparation
Complete collection/cleaning of mammography annotations
Train mammography models
Dissertation writing
Mammography model manuscript preparation

# Mahalo nui loa!