Part IV - B

Model calibration

Conformal Prediction

Adrian Galdran
RyC/ikerbasque Research Fellow
Tecnalia - Derio, Spain
adrian.galdran@tecnalia.com

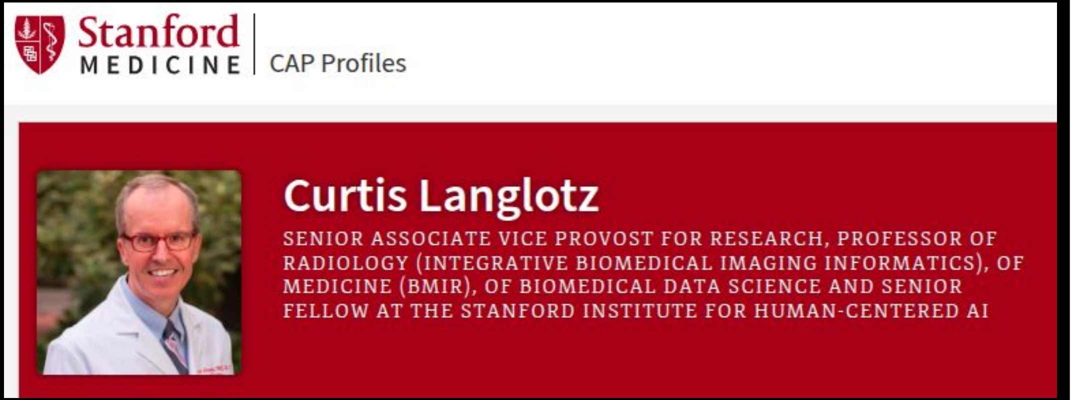
MICCAI2025
Daejeon
REPUBLIC OF KOREA

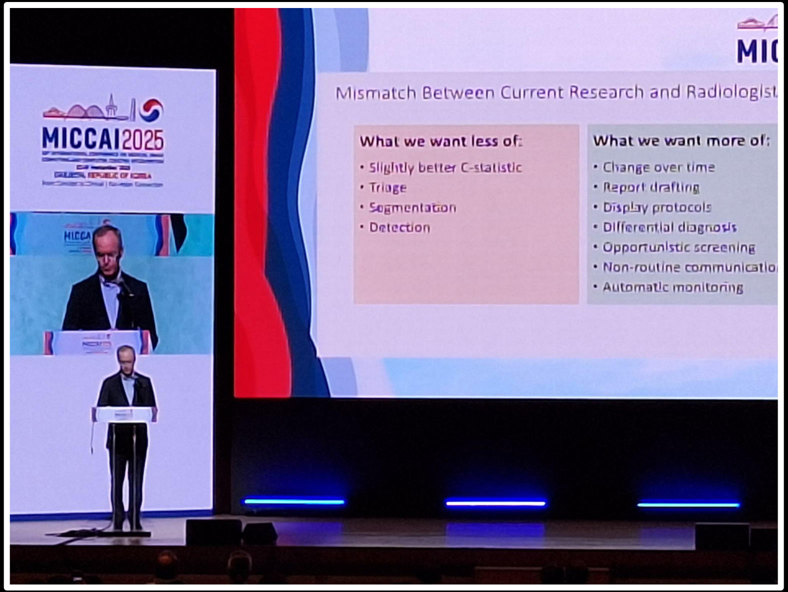
tecnalia
MEMBER OF BASQUE RESEARCH
& TECHNOLOGY ALLIANCE

Conformal Prediction - Contents

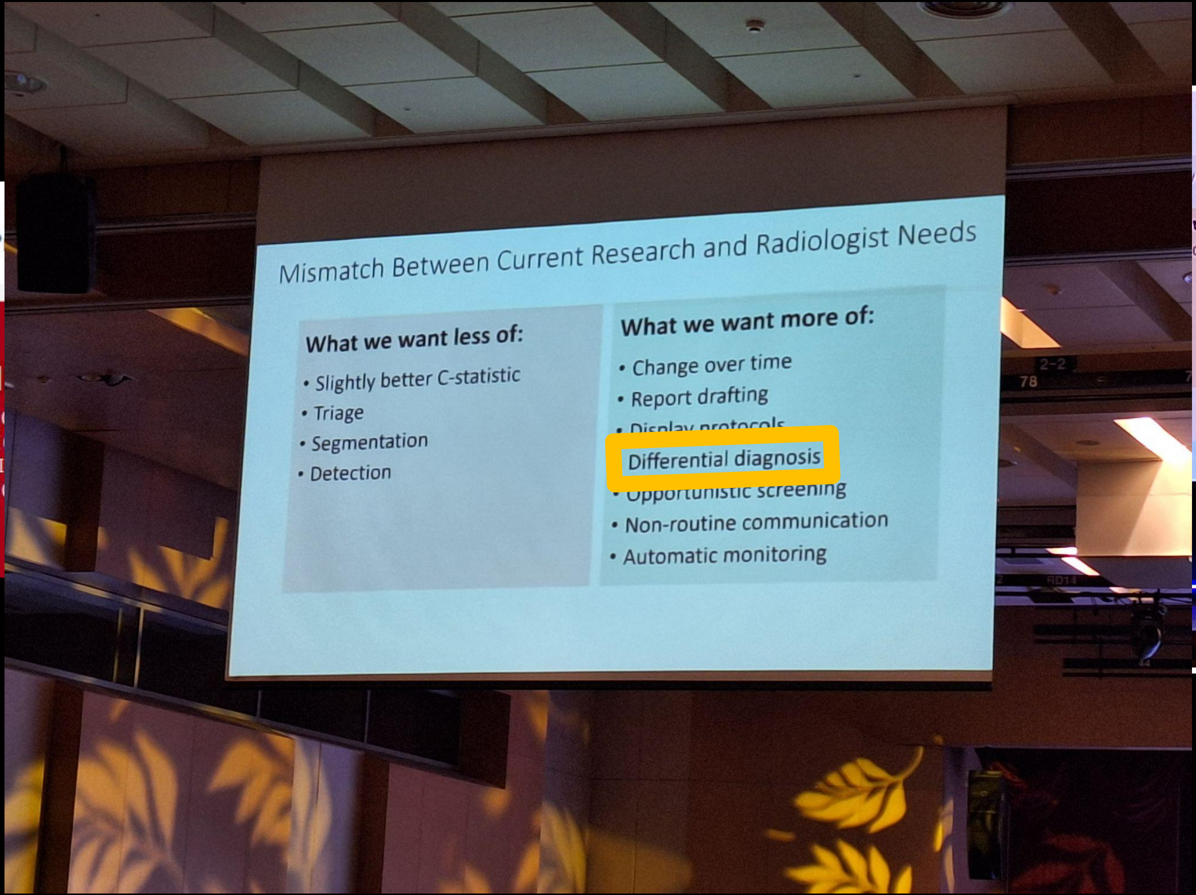
1. Motivation
2. Conformal Prediction: Ingredients
3. Conformal Predictions: Algorithm
4. Hands-On

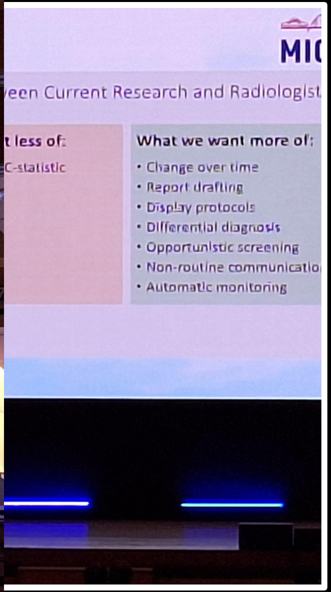
1. Motivation





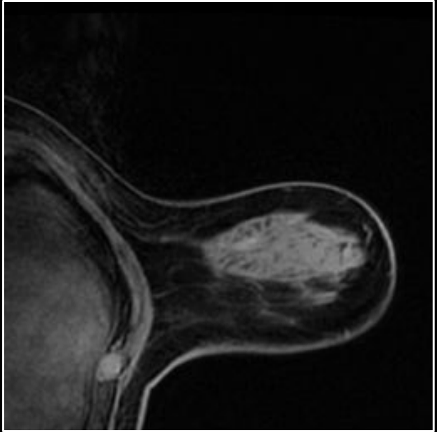
1. Motivation

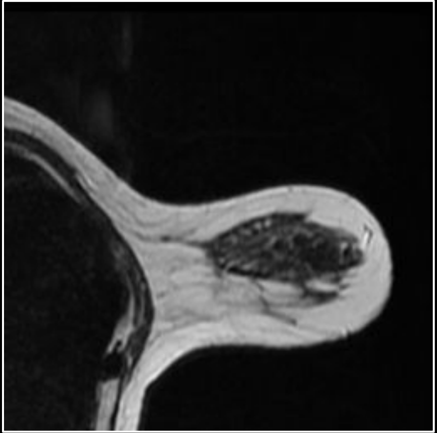


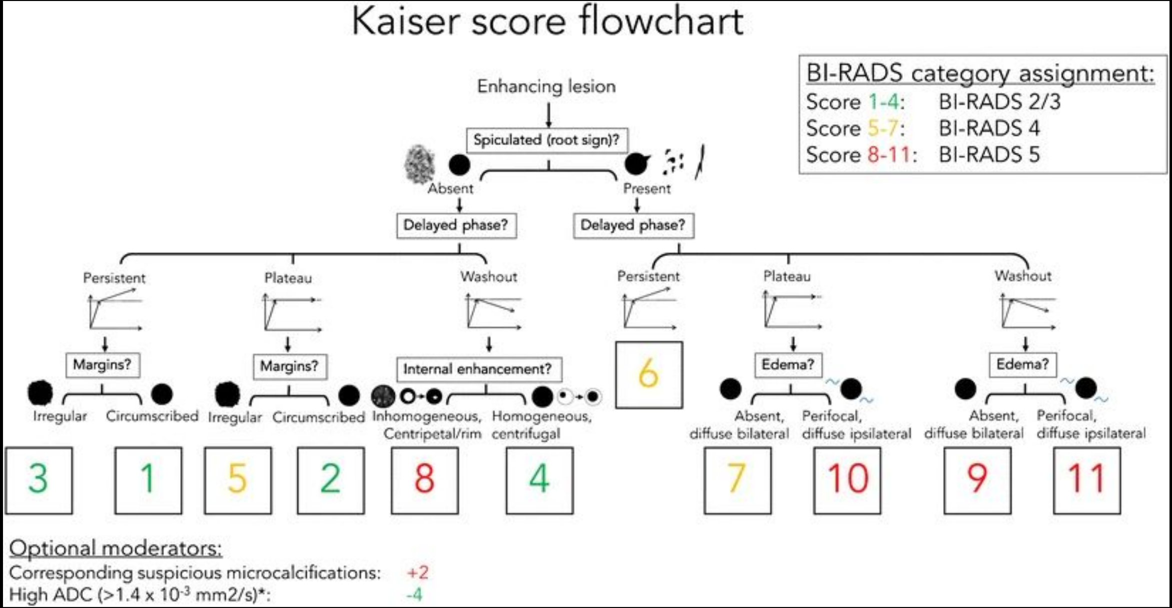


1. Motivation

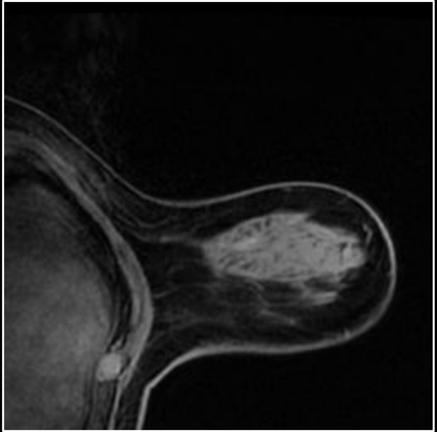
Dietzel M, Baltzer PAT. How to use the Kaiser score as a clinical decision rule for diagnosis in multiparametric breast MRI: a pictorial essay. Insights Imaging. 2018

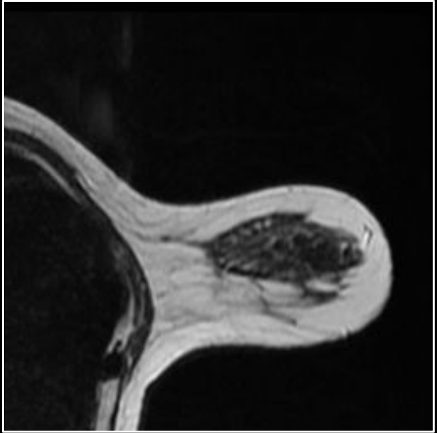




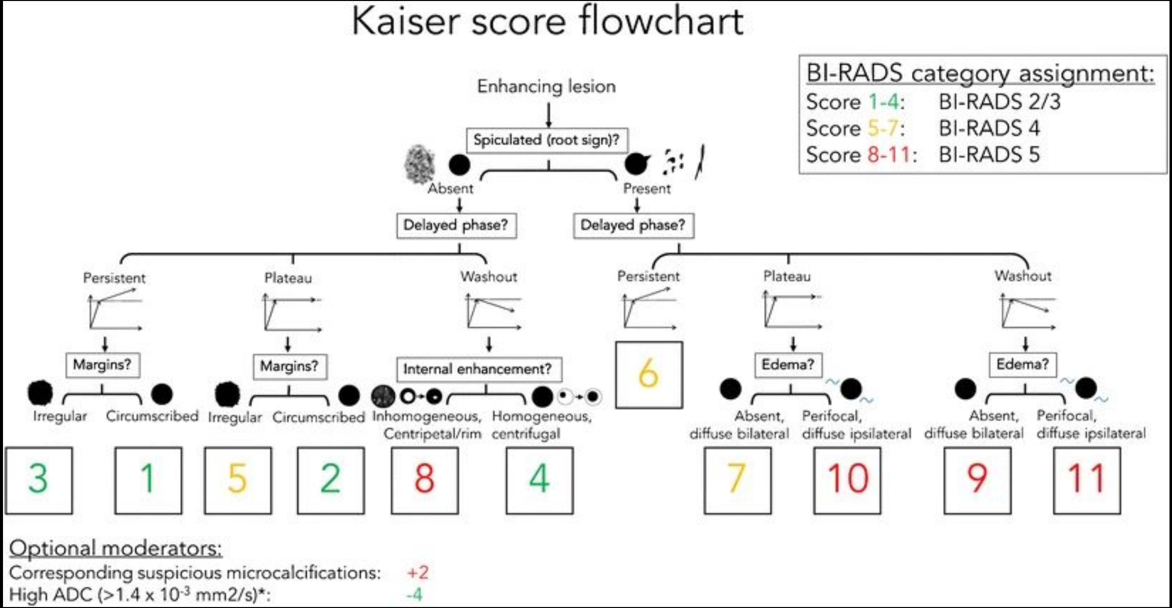


1. Motivation





Dietzel M, Baltzer PAT. How to use the Kaiser score as a clinical decision rule for diagnosis in multiparametric breast MRI: a pictorial essay. Insights Imaging. 2018



$$\mathbb{P}(\text{BI-RADS} \in \{\mathbf{7}, \mathbf{9}, \mathbf{10}, \mathbf{11}\}) \geq \mathbf{90\%}$$

1. Motivation: Differential Diagnosis

Step 1: Do we see an enhancing lesion?

- On MRI, when you inject contrast, suspicious lesions often "light up" because they have abnormal blood vessels.
- So the first question is: *does the lesion actually enhance?* If yes → move on.

Step 2: Is there a spiculated margin ("root sign")?

- **Spiculation** means the lesion has spikes or radiating lines extending into the surrounding tissue — like roots of a tree.
- This is a strong red flag for cancer. If it's present, you follow the *right-hand branch*.
- If absent (the lesion looks smooth/rounded), you go to the *left-hand branch*.

Step 3: What happens in the delayed phase (contrast wash-out curve)?

This is the part you asked about — it's the **time course** of contrast enhancement.

- After contrast injection, radiologists watch how bright the lesion gets over time.
- There are three typical "curves":
  - **Persistent:** keeps getting brighter and brighter with time.
    → Usually benign (think of a sponge slowly soaking water).
  - **Plateau:** gets bright quickly, then levels off.
    → Suspicious (like tissue that soaks fast but then "caps out").
  - **Washout:** gets bright early, but then *fades* as contrast drains away.
    → Very suspicious for cancer (because malignant tumors often have "leaky" vessels).

So the **"delayed phase" check** is: which of these three time-curves does the lesion follow?

Step 4: If persistent or plateau → check margins

- **Margins** = edge of the lesion.
- Smooth (circumscribed) edges → usually benign.
- Irregular/jagged edges → more worrisome.

Step 5: If washout → check internal enhancement pattern

- Inside the lesion, how does the contrast distribute?
- If it's patchy, rim-shaped, or irregular → higher suspicion.
- If it's uniform or "centrifugal" (from inside out), less worrisome.

Step 6: If spiculation was present (right side of the tree) → check delayed phase again, then edema

- If the lesion is spiculated and the curve is persistent/plateau/washout, you still branch down.
- In later branches, radiologists look for **edema** — swelling in the tissue around the lesion.
- Edema often shows up in malignancy, so its presence increases the suspicion.

Putting it together

At the bottom of the tree, you land on a **Kaiser score number (1–11)**.

- **Low scores (1–4)**: likely benign (BI-RADS 2/3).
- **Middle scores (5–7)**: indeterminate but suspicious (BI-RADS 4).
- **High scores (8–11)**: very suspicious, likely malignant (BI-RADS 5).

2. Conformal Prediction: Ingredients

Suppose you've got an FDA-approved diagnostic model

$$\hat{\mathcal{M}}_y(\boldsymbol{x}) \sim \mathbb{P}(\boldsymbol{Y} = \boldsymbol{y} \,|\, \boldsymbol{X} = \boldsymbol{x}) \qquad \boldsymbol{y} \in \{\mathbf{1}, \ldots, \boldsymbol{K}\} = \boldsymbol{\mathcal{Y}}$$

2. Conformal Prediction: Ingredients

Suppose you've got an FDA-approved diagnostic model

$$\hat{\mathcal{M}}_y(\boldsymbol{x}) \sim \mathbb{P}(\boldsymbol{Y} = \boldsymbol{y} \,|\, \boldsymbol{X} = \boldsymbol{x}) \qquad \boldsymbol{y} \in \{\mathbf{1}, \ldots, \boldsymbol{K}\} = \boldsymbol{\mathcal{Y}}$$

$$\hat{\mathcal{M}}_y(\boldsymbol{x}^*) = \mathbf{7}$$

2. Conformal Prediction: Ingredients

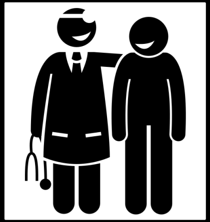
Suppose you've got an FDA-approved diagnostic model

$$\hat{\mathcal{M}}_y(\boldsymbol{x}) \sim \mathbb{P}(\boldsymbol{Y} = \boldsymbol{y} \,|\, \boldsymbol{X} = \boldsymbol{x}) \qquad \boldsymbol{y} \in \{\mathbf{1}, \ldots, \boldsymbol{K}\} = \mathcal{Y}$$

$$\hat{\mathcal{M}}_y(\boldsymbol{x}^*) = \mathbf{7}$$

$$\hat{\mathcal{M}}_y(\boldsymbol{x}^*) \in \{\mathbf{7}, \mathbf{8}, \mathbf{9}\}$$

$$\boldsymbol{with} \;\; \boldsymbol{p} = \mathbf{0.9}$$



2. Conformal Prediction: Ingredients

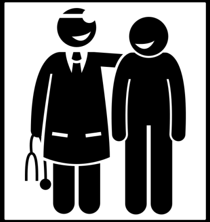
Suppose you've got an FDA-approved diagnostic model

$$\hat{\mathcal{M}}_y(\boldsymbol{x}) \sim \mathbb{P}(\boldsymbol{Y} = \boldsymbol{y} \,|\, \boldsymbol{X} = \boldsymbol{x}) \qquad \boldsymbol{y} \in \{\mathbf{1}, \ldots, \boldsymbol{K}\} = \mathcal{Y}$$

$$\hat{\mathcal{M}}_y(\boldsymbol{x}^*) = \mathbf{7}$$

$$\hat{\mathcal{M}}_y(\boldsymbol{x}^*) \in \{\mathbf{7}, \mathbf{8}, \mathbf{9}\}$$

$$\boldsymbol{with} \;\; \boldsymbol{p} = \mathbf{0.9}$$



Predict a set $\mathcal{T}_{x^*} \subseteq \mathcal{Y}$ which contains y^* with high p

$$\textbf{Coverage}: \; \mathbb{P}(y^* \in \mathcal{T}_{x^*}) \geq 1 - \alpha$$

MICCAI2025 Daejeon REPUBLIC OF KOREA

2. Conformal Prediction: Ingredients

Prediction Sets : $x^* \mapsto \mathcal{T}_{x^*} \subseteq \mathcal{Y} = \{1, \ldots, K\}$

Coverage : $\mathbb{P}\left(y^* \in \mathcal{T}_{x^*}\right) \geq 1 - \alpha$

MICCAI2025 Daejeon REPUBLIC OF KOREA

2. Conformal Prediction: Ingredients

Prediction Sets : $x^* \mapsto \mathcal{T}_{x^*} \subseteq \mathcal{Y} = \{1, \ldots, \mathcal{K}\}$

Coverage : $\mathbb{P}\left(y^* \in \mathcal{T}_{x^*}\right) \geq 1 - \alpha$

Angelopoulos & Bates. A Gentle Introduction to Conformal Prediction and Distribution-Free Uncertainty Quantification, arXiv:2107.07511, 2021

2. Conformal Prediction: Ingredients

Prediction Sets : $x^* \mapsto \mathcal{T}_{x^*} \subseteq \mathcal{Y} = \{1, \ldots, \mathcal{K}\}$

Coverage : $\mathbb{P}(y^* \in \mathcal{T}_{x^*}) \geq 1 - \alpha$ **Efficient** sets

Angelopoulos & Bates. A Gentle Introduction to Conformal Prediction and Distribution-Free Uncertainty Quantification, arXiv:2107.07511, 2021

3. Conformal Predictions: Algorithm

Let's get back to our FDA-approved diagnostic model $\hat{\mathcal{M}}$

3. Conformal Predictions: Algorithm

Let's get back to our FDA-approved diagnostic model $\hat{\mathcal{M}}$

No retraining allowed

3. Conformal Predictions: Algorithm

Let's get back to our FDA-approved diagnostic model $\hat{\mathcal{M}}$

No retraining allowed, but you have some fresh data available:

$$(\mathrm{bMRI}_1, y_1), \ldots, (\mathrm{bMRI}_\mathrm{N}, y_i) = \{(x_i, y_i)\}_{i=1}^{\mathrm{N}} \sim \mathbb{P} \; i.i.d.$$

We call this Calibration Set (sorry about that)

MICCAI2025 Daejeon REPUBLIC OF KOREA

3. Conformal Predictions: Algorithm

Let's get back to our FDA-approved diagnostic model $\hat{\mathcal{M}}$

No retraining allowed, but you have some fresh data available:

$$(\mathrm{bMRI}_1, y_1), \ldots, (\mathrm{bMRI}_\mathrm{N}, y_i) = \{(x_i, y_i)\}_{i=1}^{\mathrm{N}} \sim \mathbb{P} \; i.i.d.$$

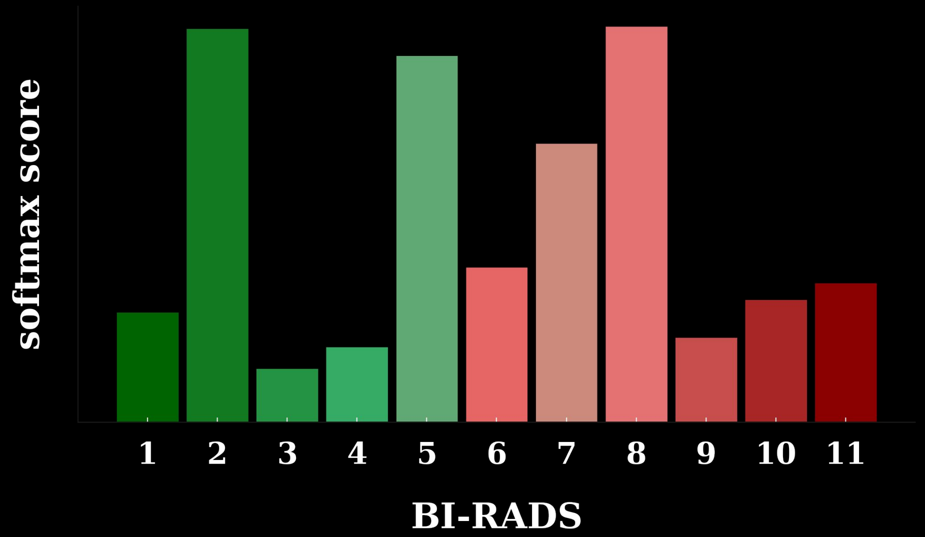
We call this Calibration Set (sorry about that)

How do we build these Conformal Prediction Sets?

MICCAI2025 Daejeon REPUBLIC OF KOREA

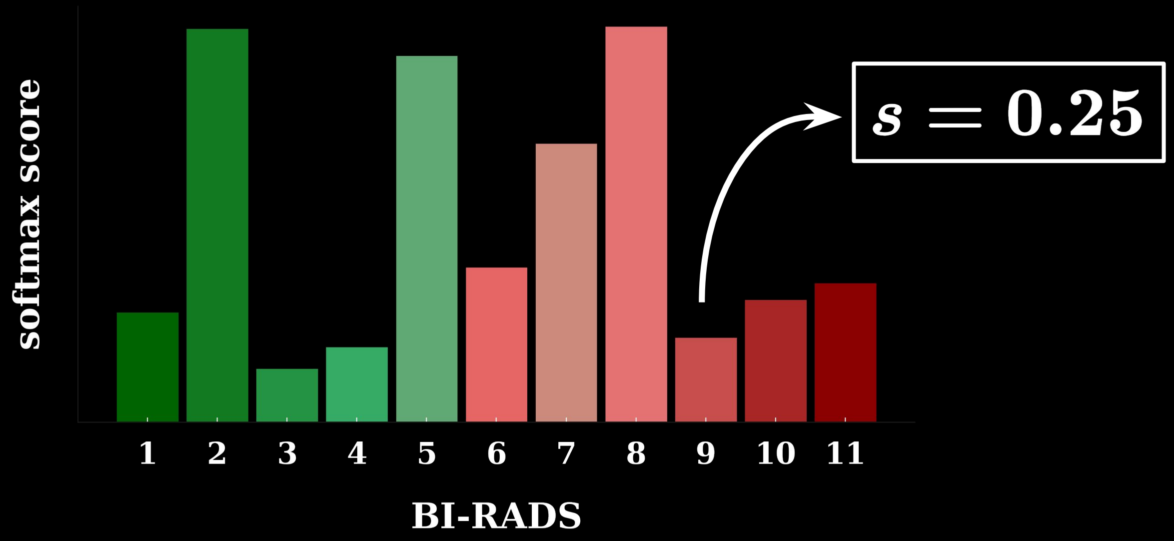
3. Conformal Predictions: Algorithm

1. Collect scores of correct classes



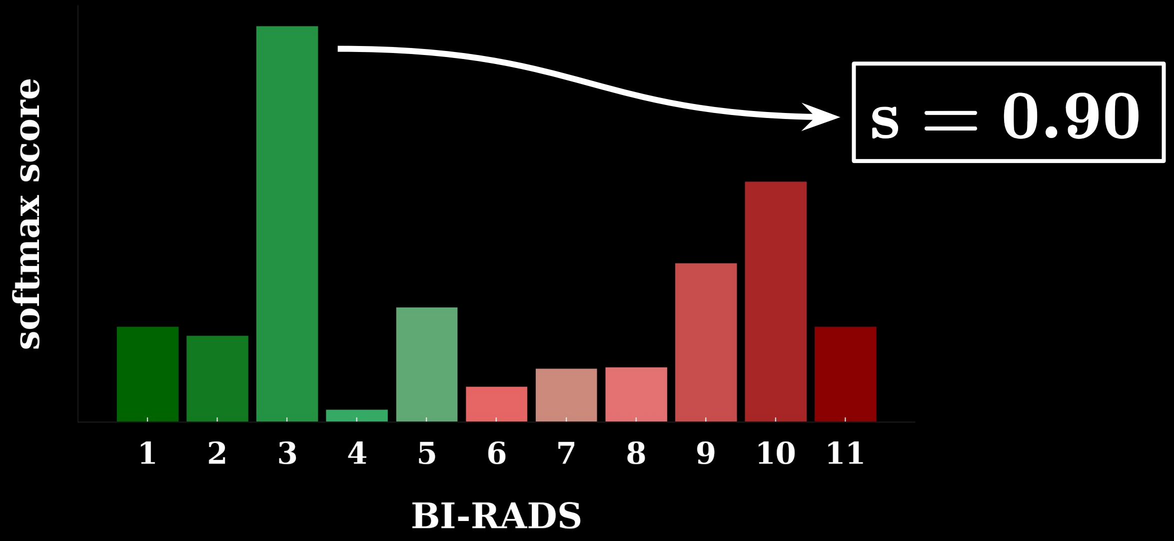
3. Conformal Predictions: Algorithm

1. Collect scores of correct classes $\mathbf{E}=\{0.25,\ \}$



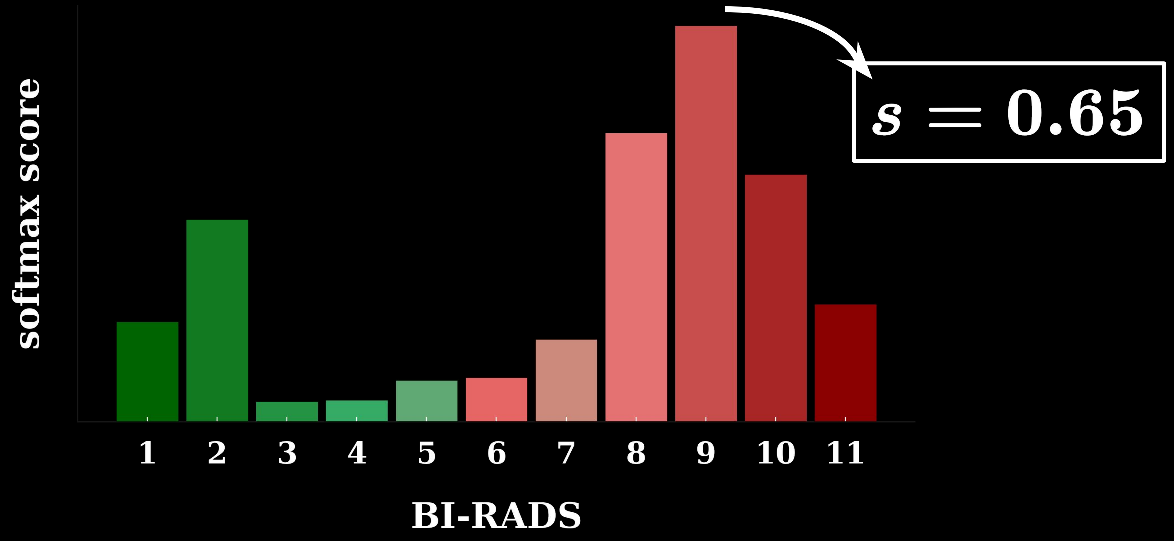
3. Conformal Predictions: Algorithm

1. Collect scores of correct classes $\mathbf{E}=\{0.25,\ 0.90,\}$



3. Conformal Predictions: Algorithm

1. Collect scores of correct classes $\mathbf{E}=\{0.25,\ 0.90,\ \ldots,\ 0.65\}$



MICCAI2025 Daejeon REPUBLIC OF KOREA

3. Conformal Predictions: Algorithm

1. Collect scores of correct classes $\mathbf{E}=\{\mathbf{0.25},\ \mathbf{0.90},\ \ldots,\ \mathbf{0.65}\}$
2. For a desired coverage of α, find a value $\hat{q_\alpha}$ such that you keep $1\text{-}\alpha$ of the scores in $\mathbb{E}$:

$$\boxed{\hat{q_\alpha} = \texttt{np.quantile}([\mathtt{E}_1, \mathtt{E}_2, \ldots, \mathtt{E}_\mathrm{N}],\ \alpha)}$$

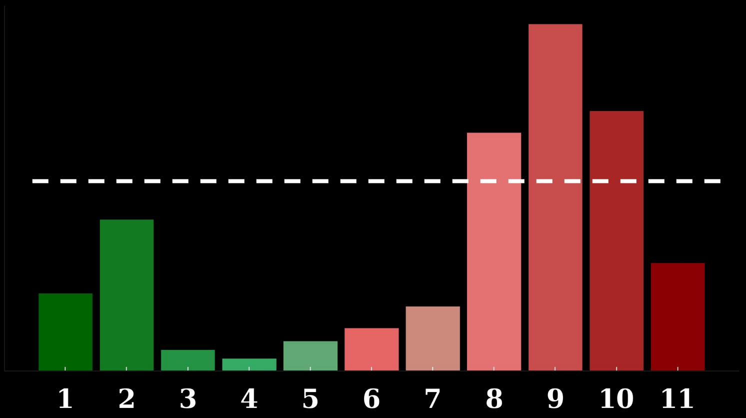
MICCAI2025 Daejeon REPUBLIC OF KOREA

3. Conformal Predictions: Algorithm

1. Collect scores of correct classes $\mathbf{E}=\{0.25,\ 0.90,\ \ldots,\ 0.65\}$
2. For a desired coverage of α, find a value $\hat{q}_\alpha$ such that you keep $1\text{-}\alpha$ of the scores in $\mathbb{E}$:

$$\boxed{\hat{q}_\alpha = \texttt{np.quantile}([\mathrm{E}_1, \mathrm{E}_2, \ldots, \mathrm{E}_N],\ \alpha)}$$

What happens if we use $\hat{q}_\alpha$ to build prediction sets in the calibration dataset?



3. Conformal Predictions: Algorithm

1. Collect scores of correct classes $\mathbf{E}=\{0.25,\ 0.90,\ \ldots,\ 0.65\}$
2. For a desired coverage of α, find a value $\hat{q_\alpha}$ such that you keep $1\text{-}\alpha$ of the scores in $\mathbb{E}$:

$$\boxed{\hat{q_\alpha} = \texttt{np.quantile}\left([\mathtt{E}_1, \mathtt{E}_2, \ldots, \mathtt{E}_\mathtt{N}],\ \alpha\right)}$$

3. If we use $\hat{q_\alpha}$ to build prediction sets on test data, we have <u>theoretically guaranteed coverage</u>

3. Conformal Predictions: Algorithm

1. Collect scores of correct classes $\mathbf{E} = \{0.25,\ 0.90,\ \ldots,\ 0.65\}$
2. For a desired coverage of α, find a value $\hat{q_\alpha}$ such that you keep $1\text{-}\alpha$ of the scores in $\mathbb{E}$:

$$\boxed{\hat{q_\alpha} = \texttt{np.quantile}\left([\mathrm{E}_1,\ \mathrm{E}_2,\ \ldots,\ \mathrm{E}_\mathrm{N}],\ \alpha\right)}$$

3. If we use $\hat{q_\alpha}$ to build prediction sets on test data, we have <u>theoretically guaranteed coverage</u>, if data is <u>interchangeable</u>.

$$\boxed{\mathbf{1}\text{-}\alpha \leq \mathbb{P}\left(\mathbf{y}^* \in \mathcal{T}_{x^*}^{\hat{q_\alpha}}\right) \leq (\mathbf{1}\text{-}\alpha) + \frac{\mathbf{1}}{\mathbf{N}+\mathbf{1}}}$$

3. Conformal Predictions: Algorithm

1. Collect scores of correct classes $\mathbf{E}=\{\mathbf{0.25},\ \mathbf{0.90},\ \ldots,\ \mathbf{0.65}\}$

2. For a desired coverage of α, find a value $\hat{q_\alpha}$ such that you keep $1\text{-}\alpha$ of the scores in $\mathbb{E}$:

$$\boxed{\hat{q_\alpha} = \texttt{np.quantile}\left(\left[\mathrm{E}_1, \mathrm{E}_2, \ldots, \mathrm{E}_\mathrm{N}\right], \alpha\right)}$$

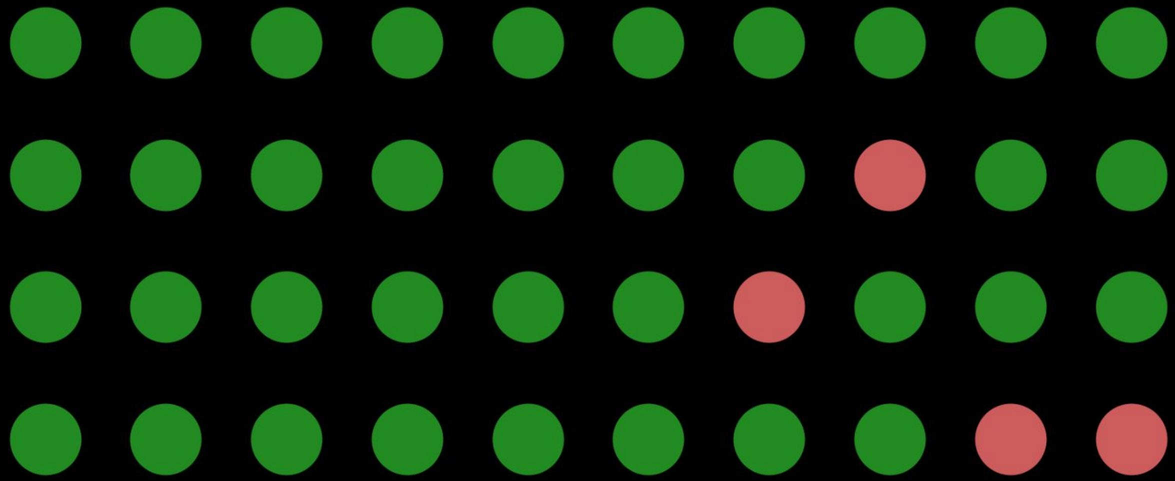
3. If we use $\hat{q_\alpha}$ to build prediction sets on test data, we have <u>theoretically guaranteed coverage</u>. If data is <u>interchangeable</u>:

$$\boxed{1\text{-}\alpha \leq \mathbb{P}\left(\mathbf{y}^* \in \mathcal{T}_{x^*}^{\hat{q_\alpha}}\right) \leq (1\text{-}\alpha) + \frac{1}{\mathrm{N}+1}}$$

MICCAI2025 Daejeon REPUBLIC OF KOREA

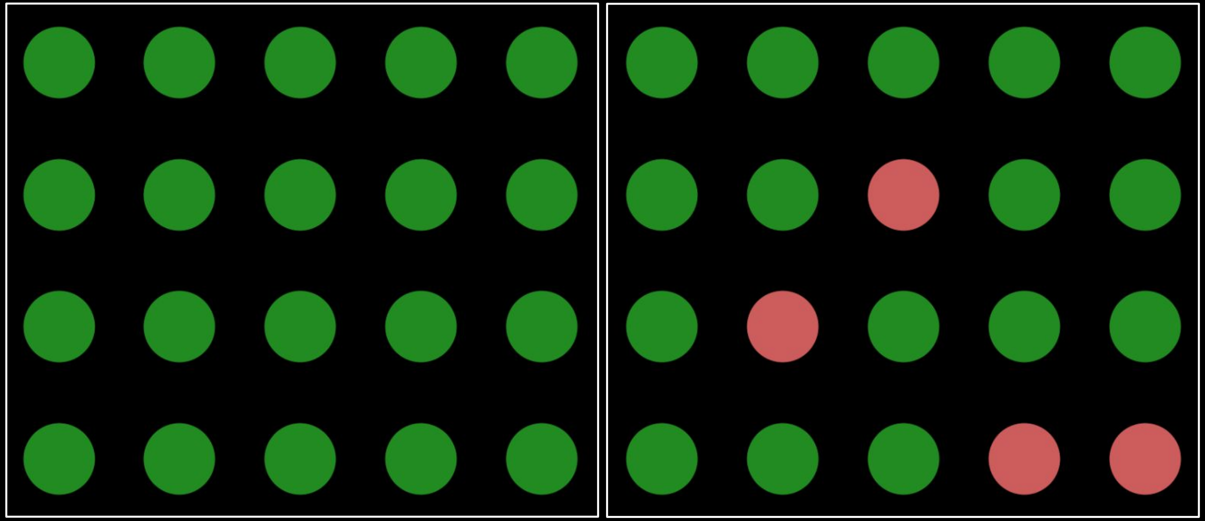
3.5 Beyond Coverage

Marginal Coverage of 90%



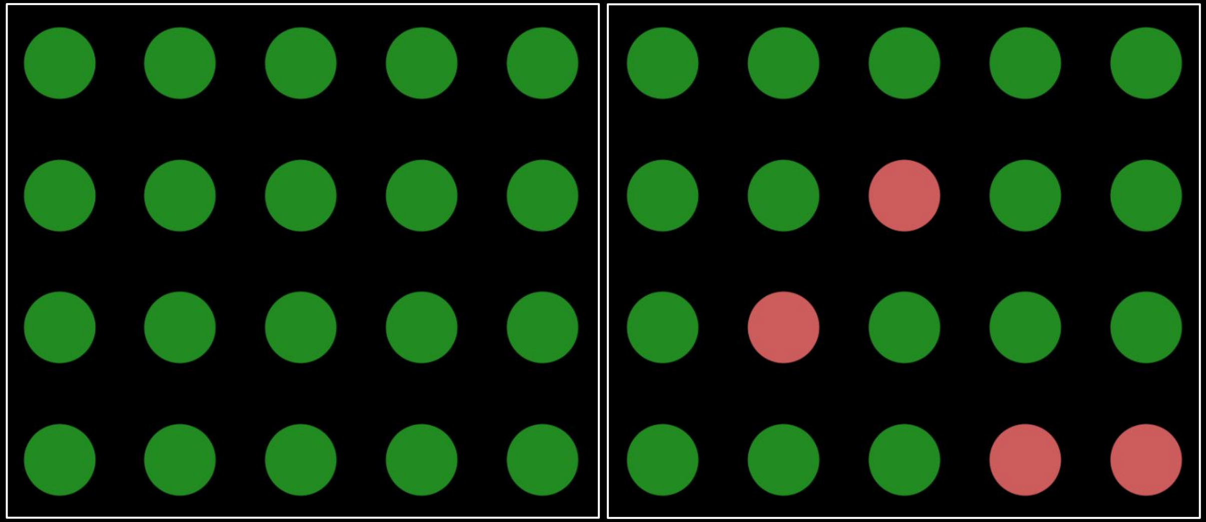
3.5 Beyond Coverage

Marginal Coverage of 90% | **No Conditional Coverage**



3.5 Beyond Coverage

Marginal Coverage of 90% **No Conditional Coverage**



Check out also Conformal Risk Control

MICCAI2025 Daejeon REPUBLIC OF KOREA

4. Hands-On

Github repository:

https://github.com/agaldran/uqinmia-miccai