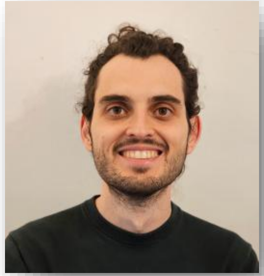
ÉTS ÉCOLE DE TECHNOLOGIE SUPÉRIEURE
Le génie pour l'industrie
Université du Québec

CentraleSupélec
université PARIS-SACLAY

IPMI 2025

Full Conformal Adaptation of Medical Vision-Language Models

Julio Silva-Rodríguez



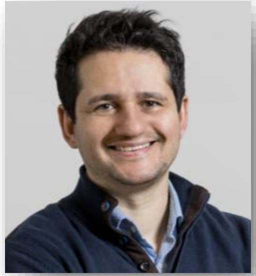
Leo Fillioux

Paul-Henry Cournède

Maria Vakalopoulou

Stergios Christodoulidis

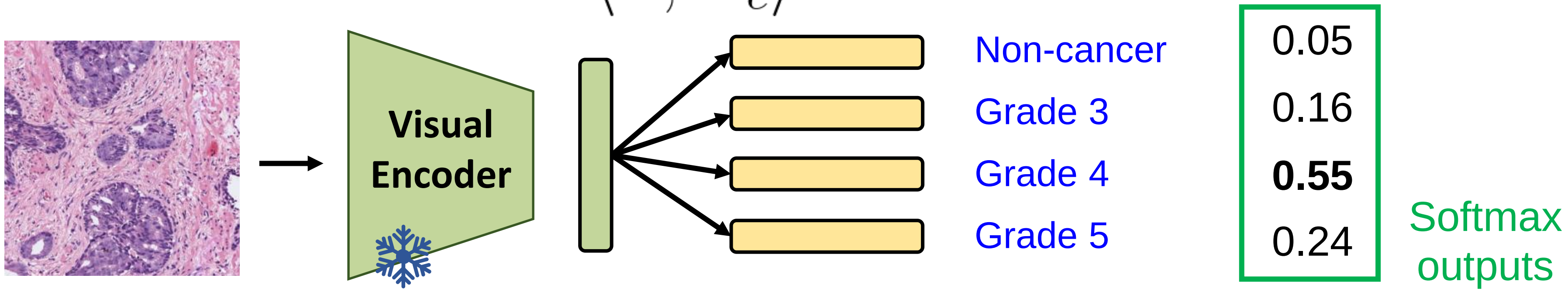
Ismail Ben Ayed



Jose Dolz

Vision-Language Models

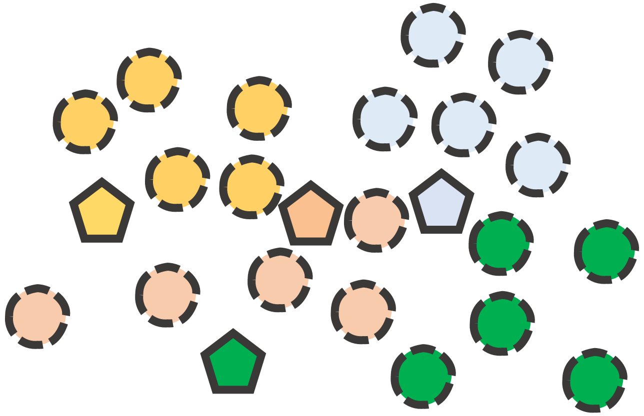
$$\langle \mathbf{v}, \mathbf{w}_c \rangle$$



$$f_\theta(\mathbf{x}) = \mathbf{v}$$

$$p_c(\mathbf{W}) = \frac{\exp(\mathbf{v}^\top \mathbf{w}_c / \tau)}{\sum_{j=1}^{C} \exp(\mathbf{v}^\top \mathbf{w}_j / \tau)}$$

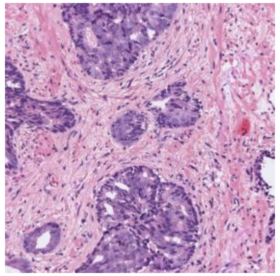
Vision-Language Models



"A prostate histology image of [CLS] tissue." → **Text Encoder**

$$f_\phi([\text{Description}]_{\mathbf{c}}) = \mathbf{t_c}$$

$$\langle \mathbf{v}, \mathbf{w}_c \rangle$$



→ **Visual Encoder**

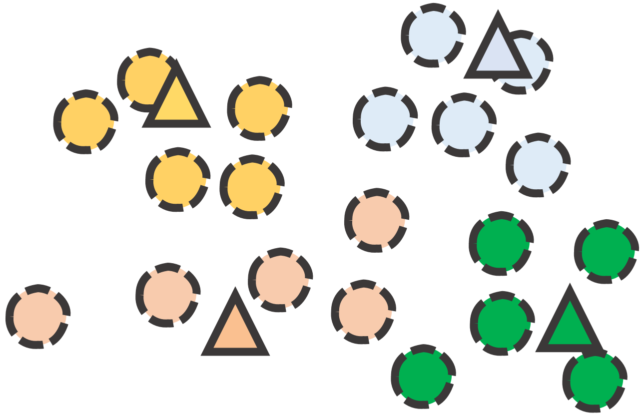
$$f_\theta(\mathbf{x}) = \mathbf{v}$$

Non-cancer	0.05
Grade 3	0.16
Grade 4	**0.55**
Grade 5	0.24

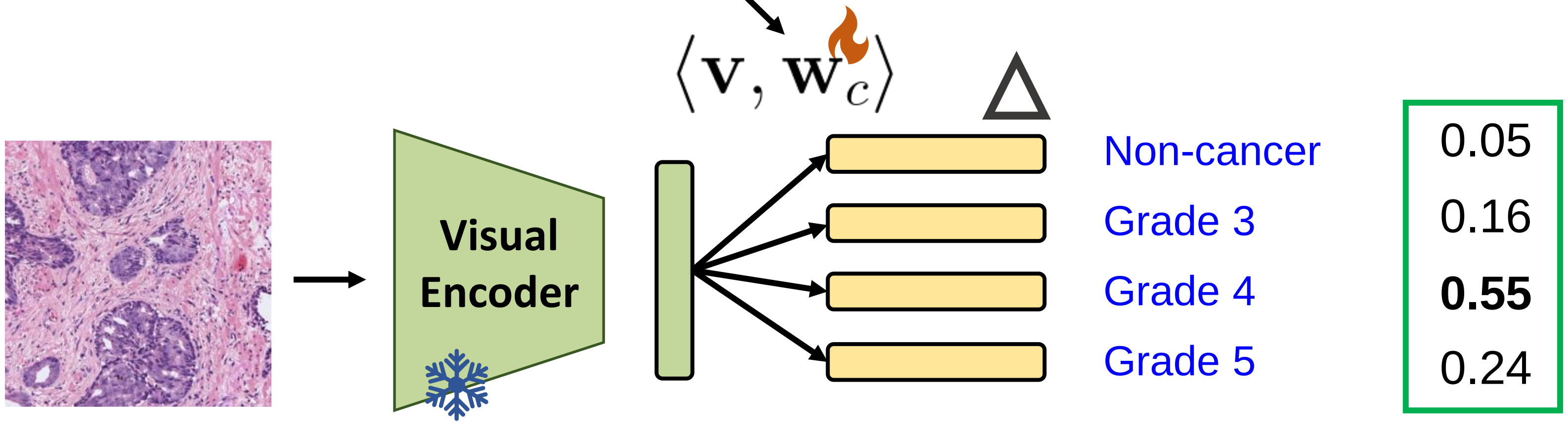
Softmax outputs

$$p_c(\mathbf{W}) = \frac{\exp(\mathbf{v}^\top \mathbf{w}_c/\tau)}{\sum_{j=1}^{C} \exp(\mathbf{v}^\top \mathbf{w}_j/\tau)}$$

Vision-Language Models



Support set
(labeled few-shot)
N = K (shots) x C (classes)

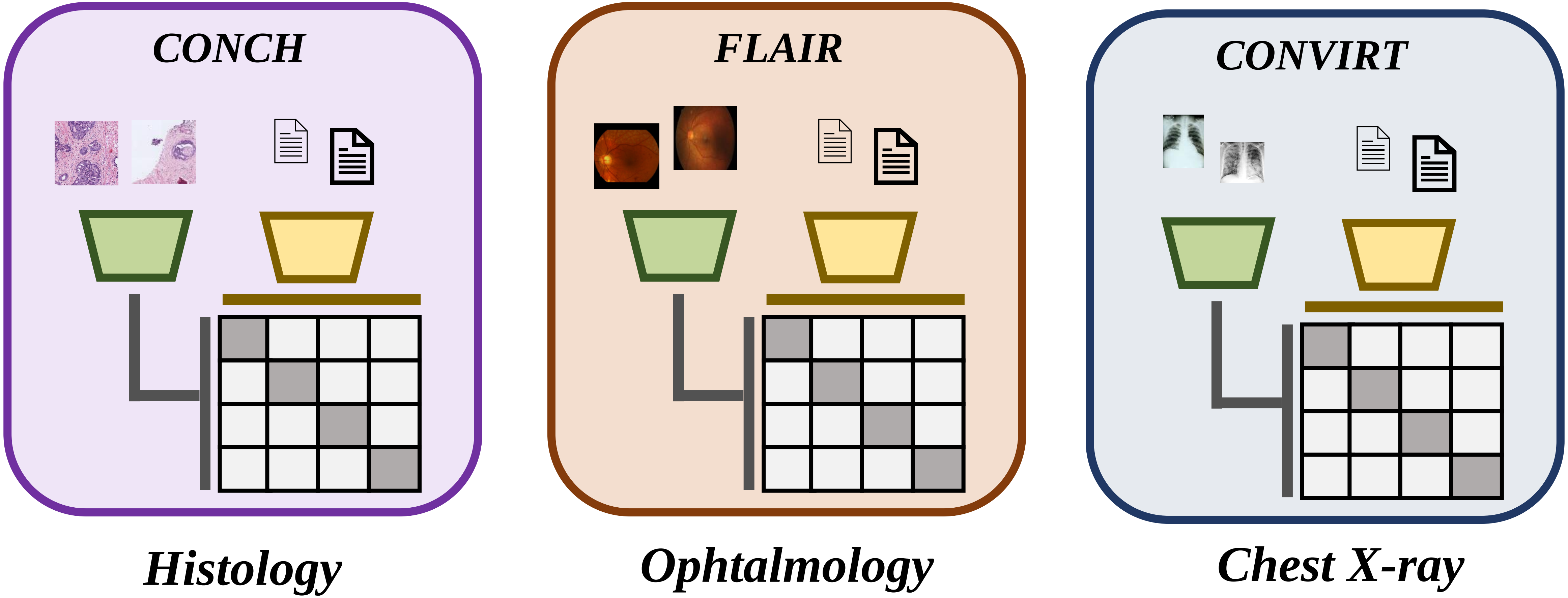


Softmax outputs

$$f_\theta(\mathbf{x}) = \mathbf{v}$$

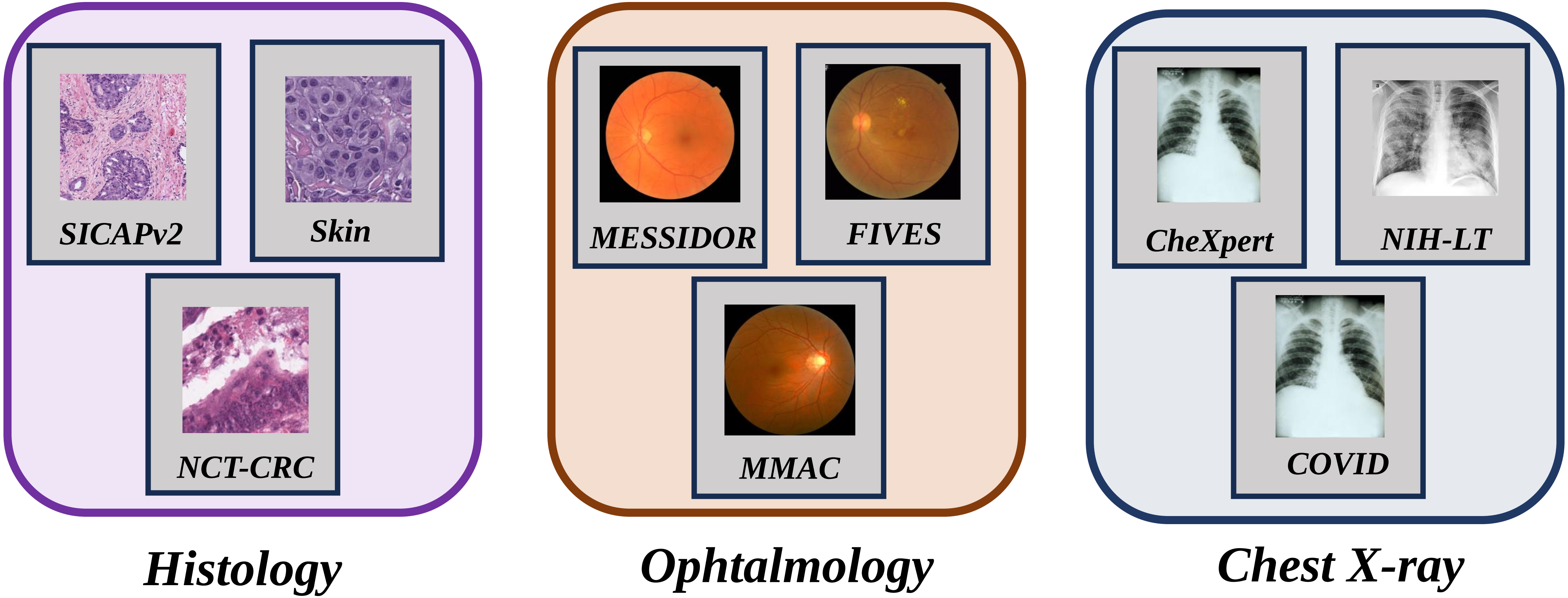
$$p_c(\mathbf{W}) = \frac{\exp(\mathbf{v}^\top \mathbf{w}_c / \tau)}{\sum_{j=1}^{C} \exp(\mathbf{v}^\top \mathbf{w}_j / \tau)}$$

Medical Vision-Language Models



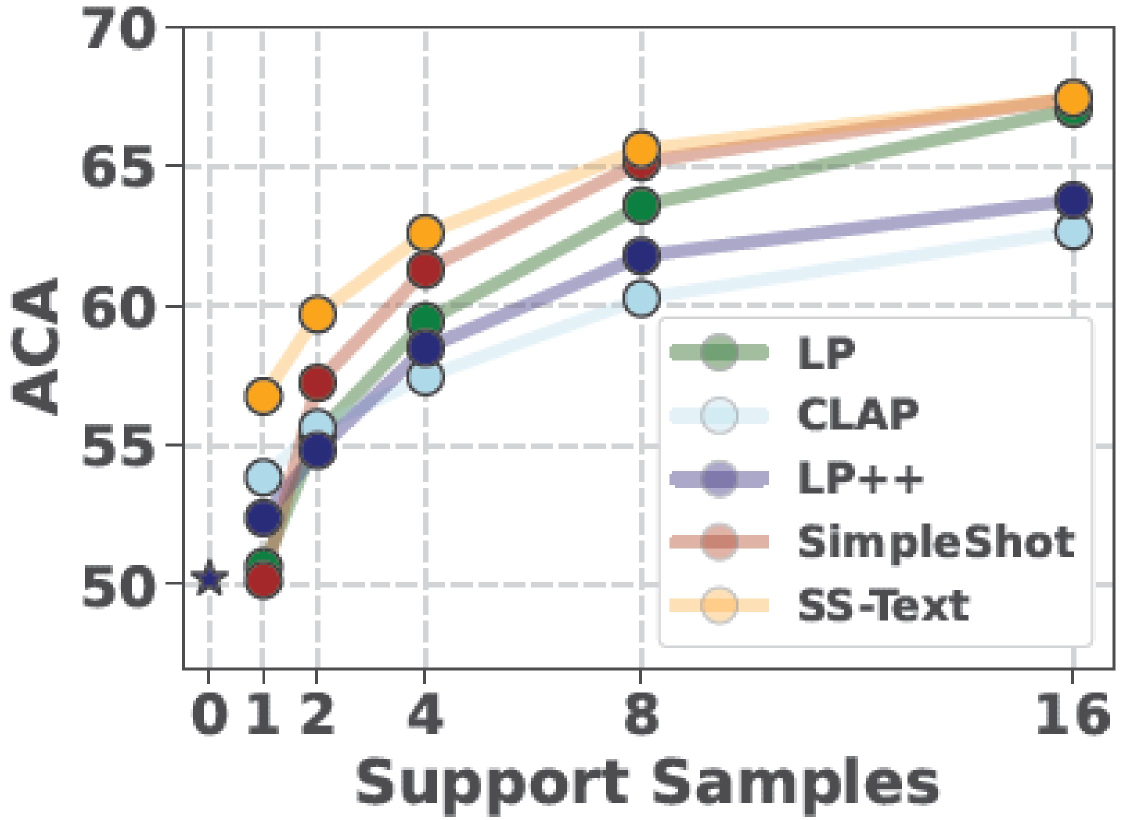
Medical Vision-Language Models

- **Downstream tasks**



Medical Vision-Language Models

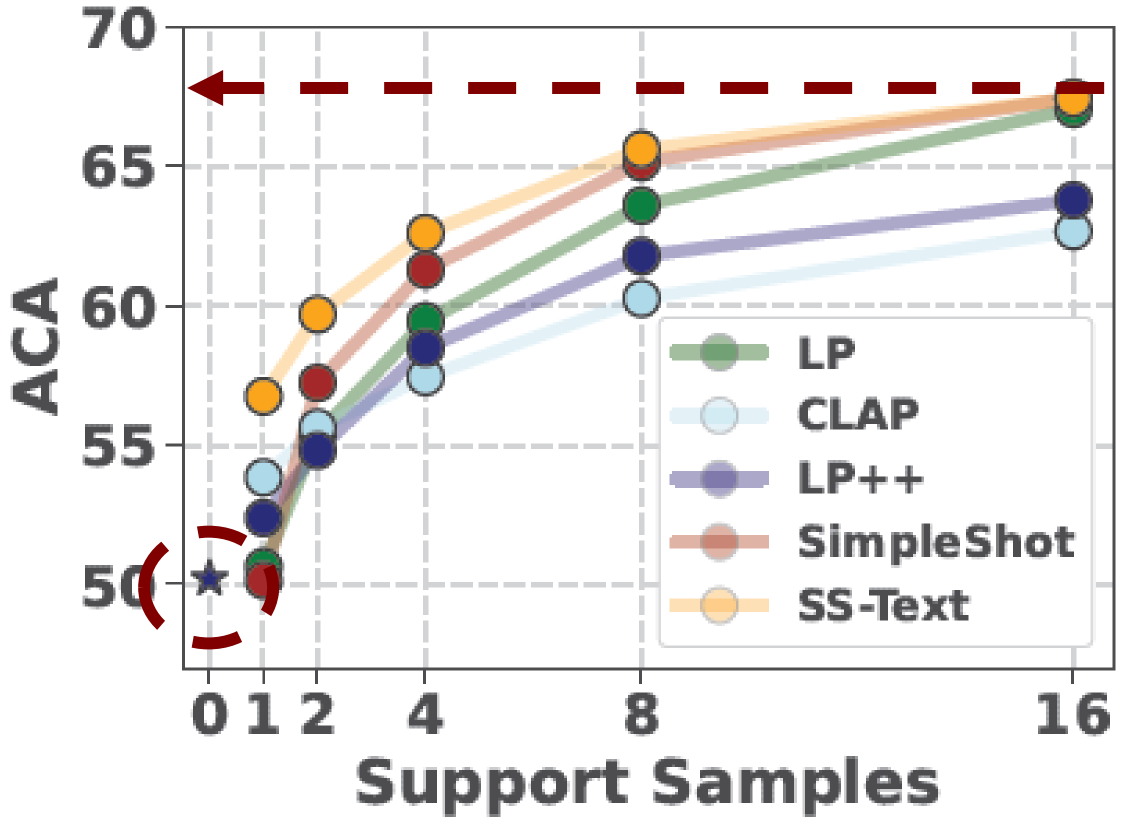
- **Efficient downstream adaptation**



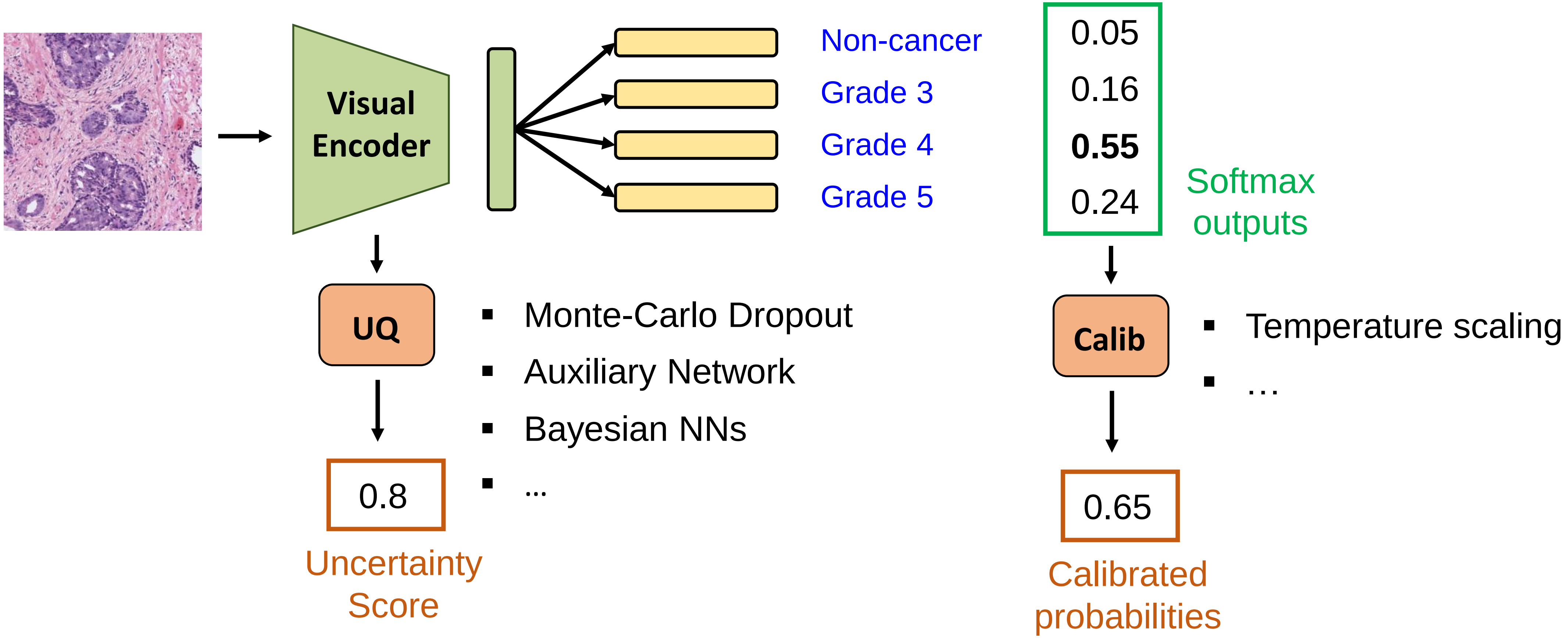
Medical Vision-Language Models

- **Efficient downstream adaptation**

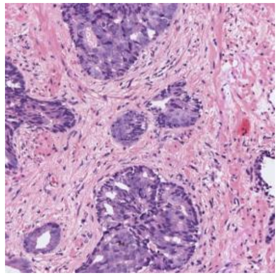
Machine learning models fail!



Uncertainty quantification trough sample rejection



Uncertainty quantification trough sample rejection



Visual Encoder

Non-cancer
Grade 3
Grade 4
Grade 5

0.05
0.16
0.55
0.24

Softmax outputs

UQ

- Monte-Carlo Dropout
- Auxiliary Network
- Bayesian NNs
- …

0.8

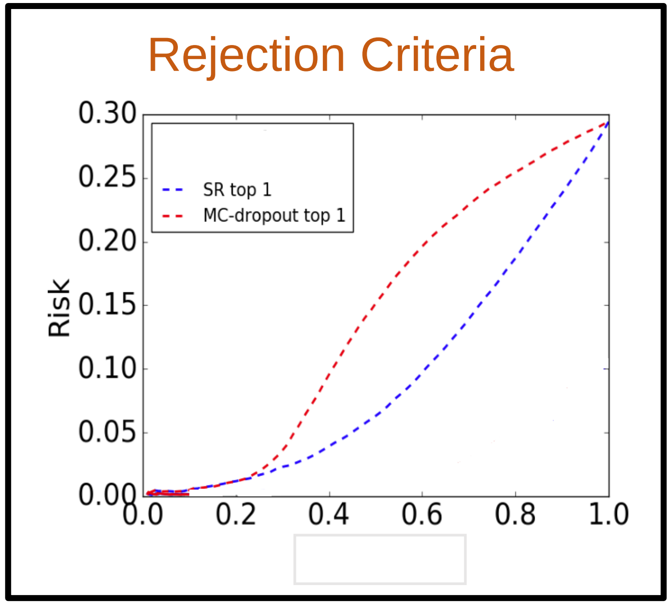
Uncertainty Score

Calib

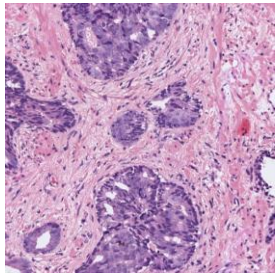
- Temperature scaling
- …

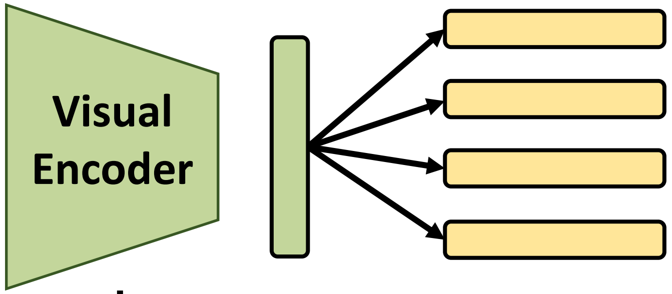
0.65

Calibrated probabilities



Uncertainty quantification trough sample rejection





Non-cancer
Grade 3
Grade 4
Grade 5

0.05
0.16
0.55
0.24

Softmax outputs

UQ

- Monte-Carlo Dropout
- Auxiliary Network
- Bayesian NNs
- …

0.8

Uncertainty Score

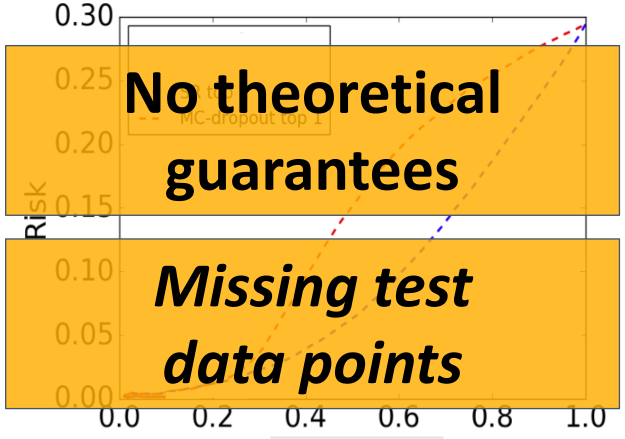
Calib

- Temperature scaling
- …

0.65

Calibrated probabilities

Rejection Criteria



Split Conformal Prediction

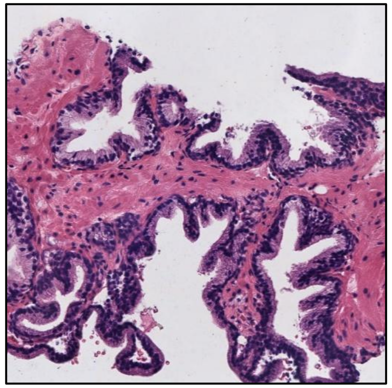
$$\mathcal{P}(Y \in C(\mathbf{x})) \geq 1 - \alpha$$

Split Conformal Prediction

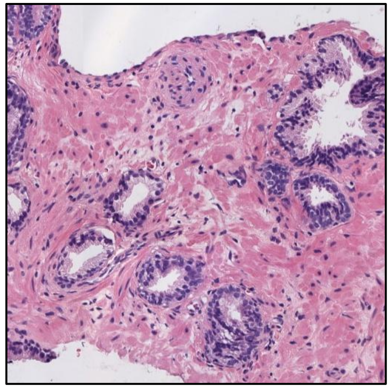
$$\mathcal{P}(Y \in C(\mathbf{x})) \geq 1 - \alpha$$

Conformal sets

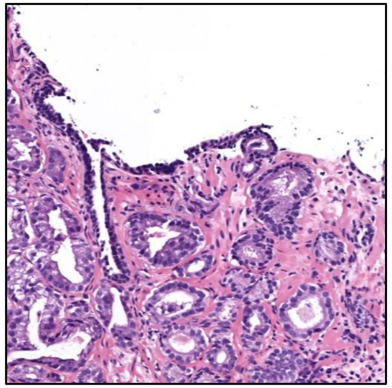
Error rate, e.g., 10%



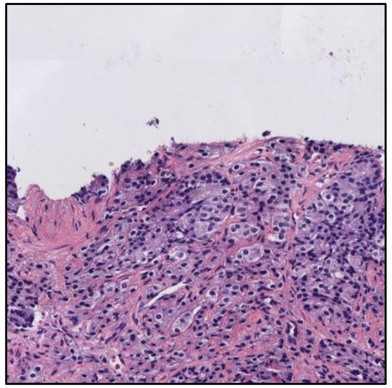
GT: NC
CP: [NC]



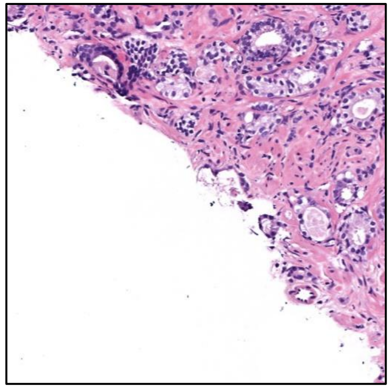
GT: G3
CP: [G3]



GT: G3
CP: [G3,G4]



GT: G5
CP: [G5]



GT: G5
CP: [G4,G5]

Split Conformal Prediction

$$\mathcal{P}(Y \in C(\mathbf{x})) \geq 1 - \alpha$$

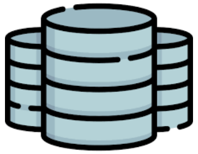
- **How to construct C(x)?**

$$\mathcal{S}(\mathbf{x}, y) = 1 - \hat{p}_{k=y}$$

non-conformity score

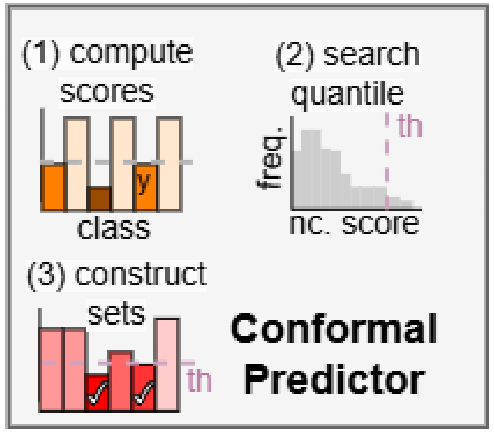
Calibration set
(labeled few-shot)

$$\hat{s} = \inf \left[s : \frac{|i \in \{1, \ldots, N\} : s_i \leq s|}{N} \geq \frac{\lceil (N+1)(1-\alpha) \rceil}{N} \right]$$



Test set
(unlabeled)

$$\mathcal{C}(\mathbf{x}) = \{y \in \mathcal{Y} : \mathcal{S}(\mathbf{x}, y) \leq \hat{s}\}$$

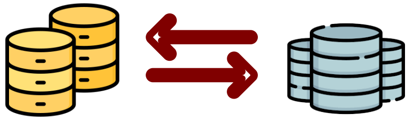


Split Conformal Prediction

Marginal over $\mathcal{X}\mathcal{Y}$

Under cal/test exchangeability

$$\mathcal{P}(Y \in C(\mathbf{x})) \geq 1 - \alpha$$



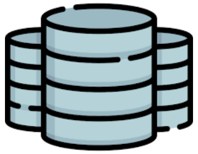
- **How to construct C(x)?**

$$\mathcal{S}(\mathbf{x}, y) = 1 - \hat{p}_{k=y}$$

non-conformity score

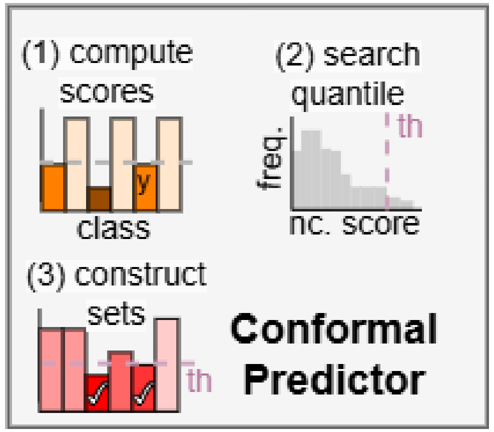
$$\hat{s} = \inf\left[s : \frac{|i \in \{1, ..., N\} : s_i \leq s|}{N} \geq \frac{\lceil (N+1)(1-\alpha) \rceil}{N} \right]$$

Calibration set
(labeled few-shot)



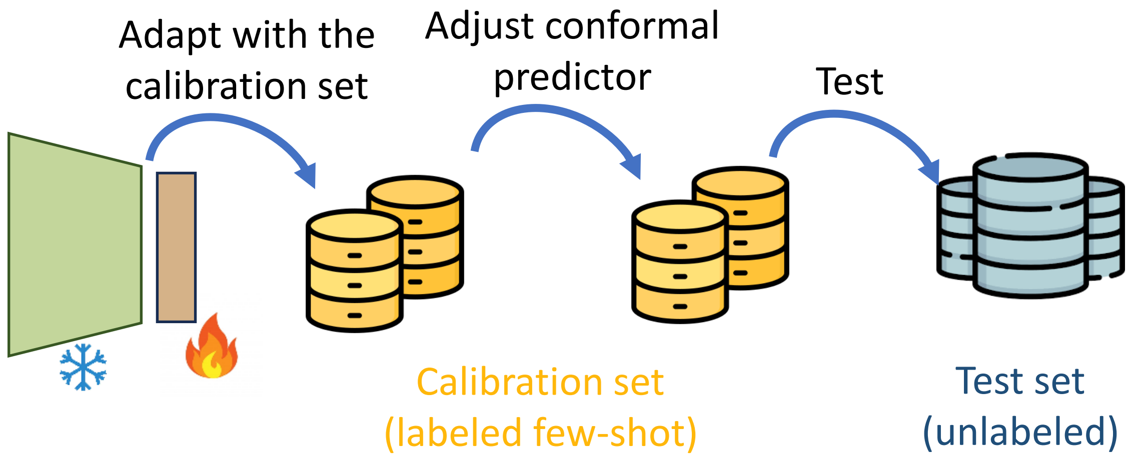
$$\mathcal{C}(\mathbf{x}) = \{ y \in \mathcal{Y} : \mathcal{S}(\mathbf{x}, y) \leq \hat{s} \}$$

Test set
(unlabeled)



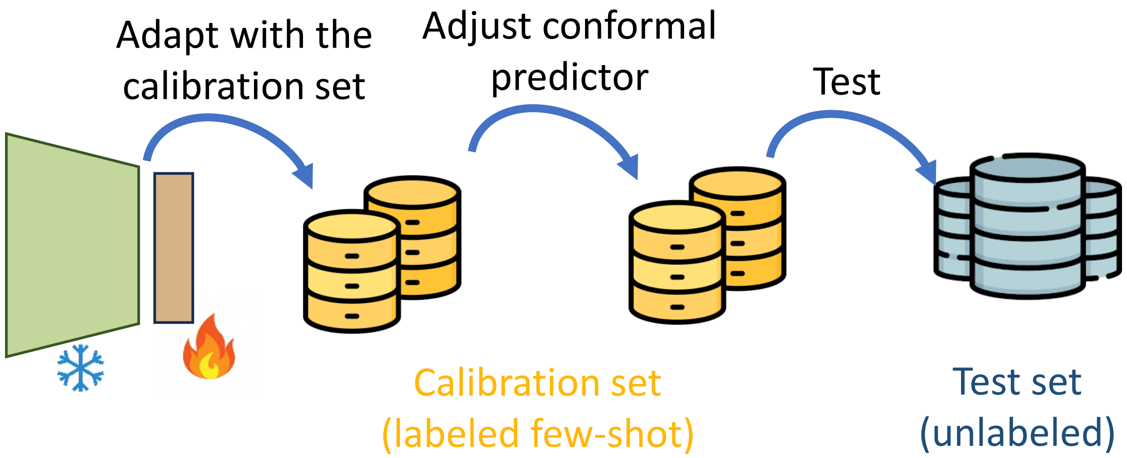
Pitfalls of conformal prediction and transfer learning

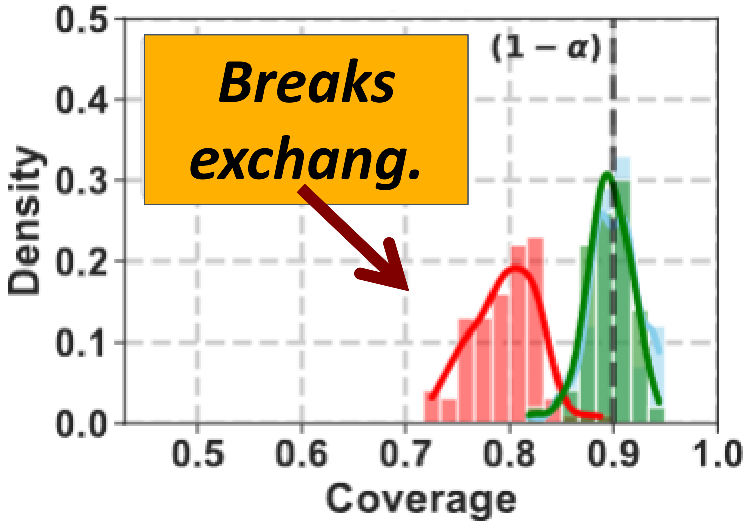
- **Can we adapt and conformalize with the <u>same calibration data</u>?**



Pitfalls of conformal prediction and transfer learning

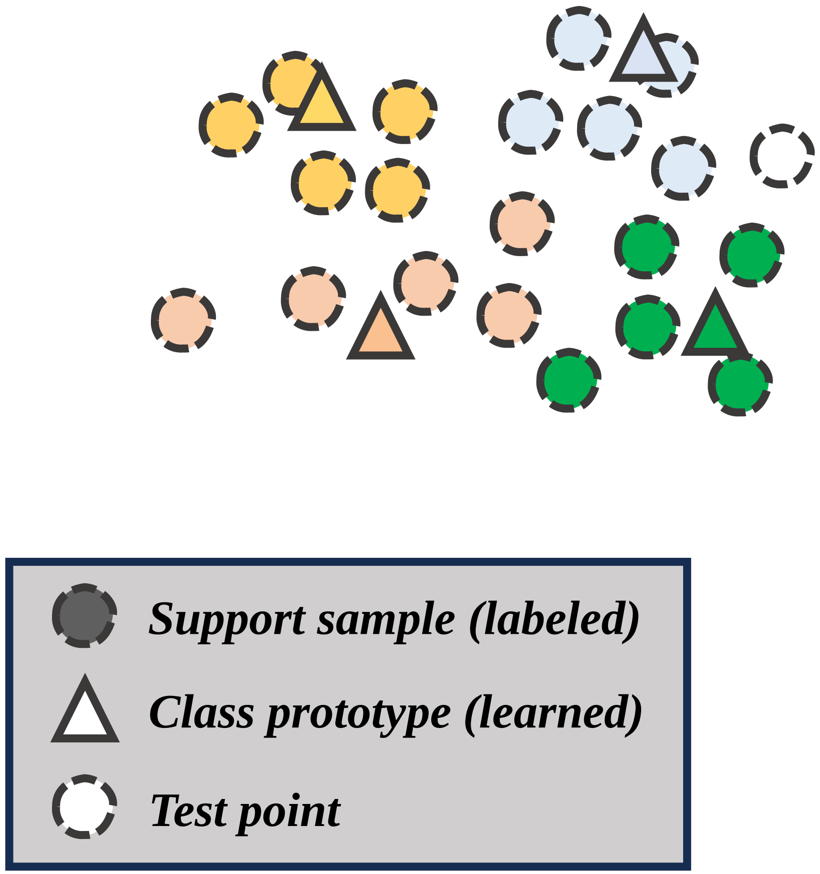
- **Can we adapt and conformalize with the <u>same calibration data</u>?**





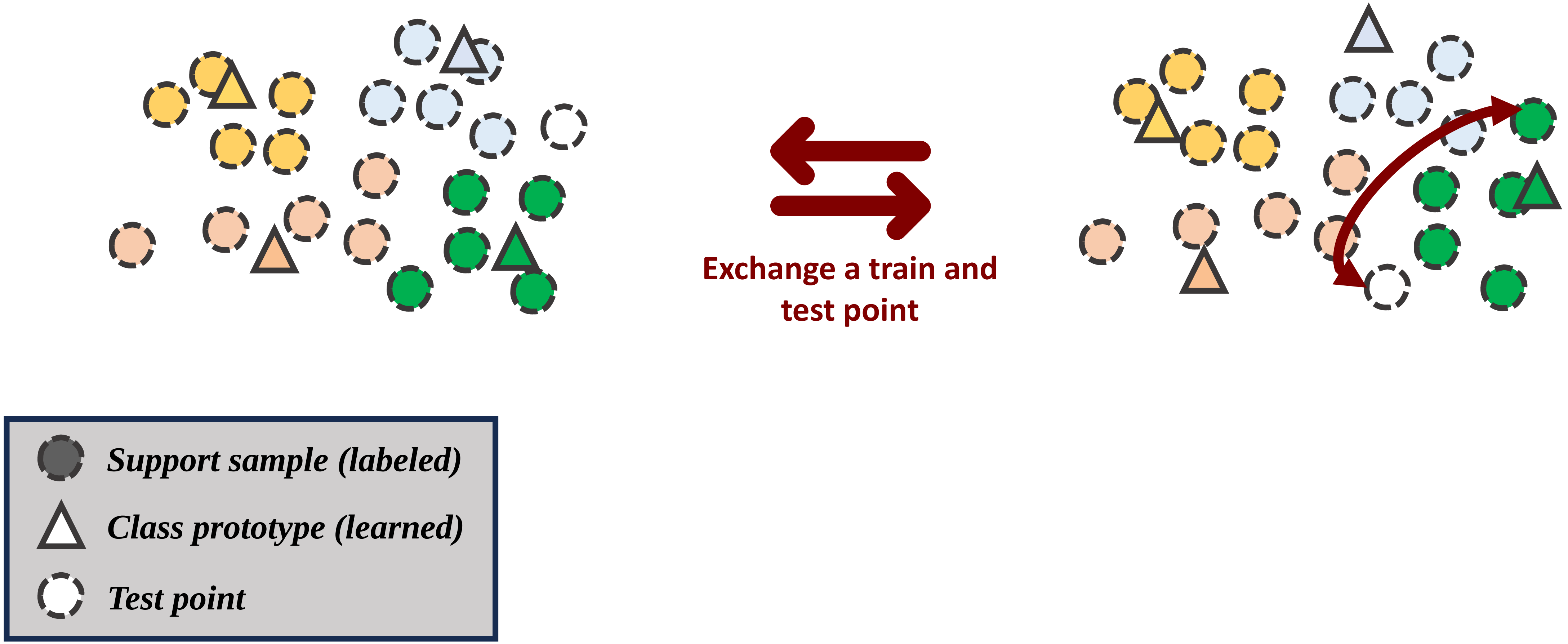
Pitfalls of conformal prediction and transfer learning

- **Why do we break the exchangeability assumption in the scores?**



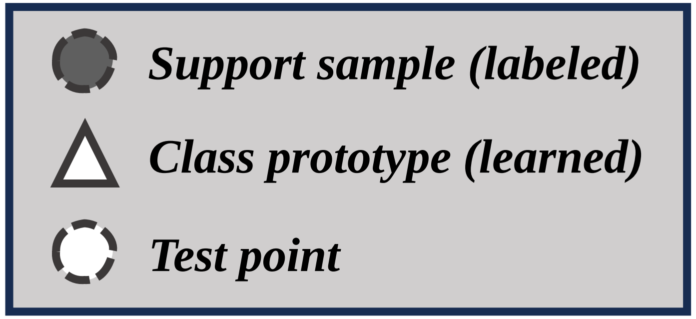
Pitfalls of conformal prediction and transfer learning

- **Why do we break the exchangeability assumption in the scores?**



Pitfalls of conformal prediction and transfer learning

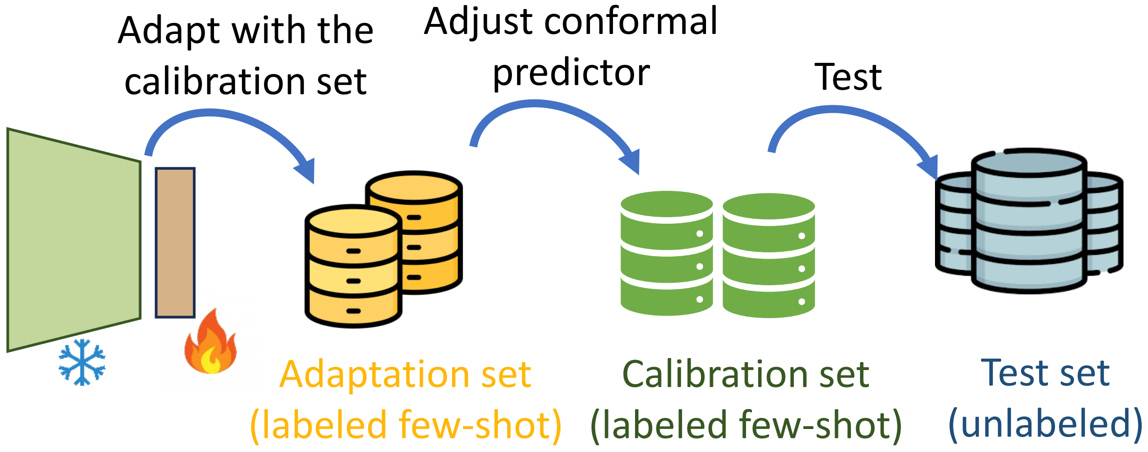
- **Why do we break the exchangeability assumption in the scores?**



Random permutations change the learned weights. Exchangeability between cal/test scores is lost.

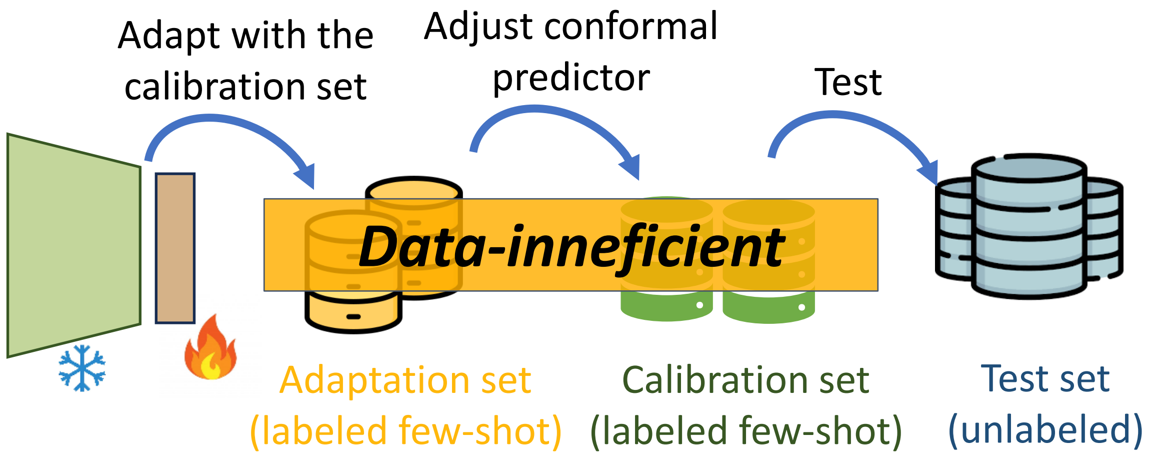
Pitfalls of conformal prediction and transfer learning

- **Alternative: adding additional data for calibration after adaptation.**



Pitfalls of conformal prediction and transfer learning

- **Alternative: adding additional data for calibration after adaptation.**



Our setting: full conformal adaptation

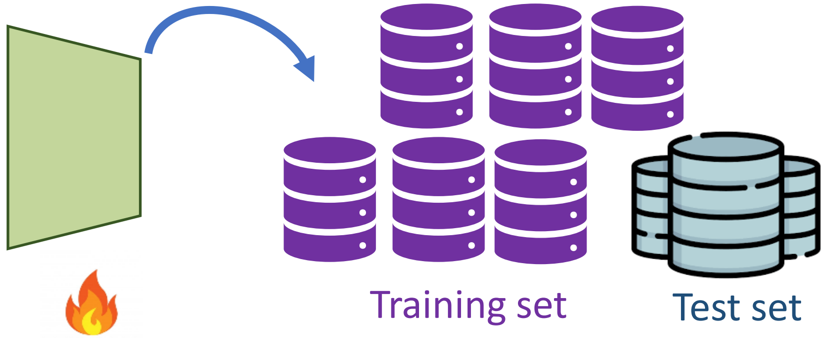
- **<u>Full conformal prediction</u> scenario (Vovk *et al*. 1998).**

Transduction with Confidence and Credibility

C. Saunders, A. Gam merman, V. Vovk
Royal Holloway, University of London
Egham, Surrey, England.
{craig,alex,vovk}@dcs.rhbnc.ac.nk

Algorithmic Learning in a Random World

Vladimir Vovk
University of London
Egham, United Kingdom



Our setting: full conformal adaptation

- **Full conformal prediction scenario (Vovk *et al*. 1998).**

Idea:

1) We know that the true label of a test point lies on the label space.

2) Let's fit the model wich each label assignment and check if the errors on the test point conform to the training observations.

Our setting: full conformal adaptation

- **<u>Full conformal prediction</u> scenario (Vovk *et al*. 1998).**

A: For each test data point…

B: For each label…

$$\mathcal{D}_{train} = \{(\mathbf{x}_1, y_1), ..., (\mathbf{x}_r, y_r), ...(\mathbf{x}_R, y_R), (\mathbf{x}_m, y)\}$$

Our setting: full conformal adaptation

- **<u>Full conformal prediction</u> scenario (Vovk *et al*. 1998).**

A: For each test data point…

B: For each label…

$$\mathcal{D}_{train} = \{(\mathbf{x}_1, y_1), ..., (\mathbf{x}_r, y_r), ...(\mathbf{x}_R, y_R), (\mathbf{x}_m, y)\}$$

1. Train model on joint dataset

$$\pi(\cdot)^y : y_m = y \in \mathcal{Y}$$

Our setting: full conformal adaptation

- **<u>Full conformal prediction</u> scenario (Vovk *et al*. 1998).**

A: For each test data point…

B: For each label…

$$\mathcal{D}_{train} = \{(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_r, y_r), \ldots(\mathbf{x}_R, y_R), (\mathbf{x}_m, y)\}$$

1. Train model on joint dataset

$$\pi(\cdot)^y : y_m = y \in \mathcal{Y}$$

2. Search quantile in training data

$$s_i^y = \mathcal{S}(\pi_i^y(\mathbf{x}), y_i)$$

3. Accept/Reject label

$$\mathcal{C}(\mathbf{x}) = \{y \in \mathcal{Y} : s^y \leq \hat{s}^y\}$$

Our setting: full conformal adaptation

- **Full conformal prediction scenario (Vovk *et al*. 1998).**

A: For each test data point…

B: For each label…

$$\mathcal{D}_{train} = \{(\mathbf{x}_1, y_1), ..., (\mathbf{x}_r, y_r), ...(\mathbf{x}_R, y_R), (\mathbf{x}_m, y)\}$$

1. Train model on joint dataset

$$\pi(\cdot)^y : y_m = y \in \mathcal{Y}$$

2. Search quantile in training data

$$s_i^y = \mathcal{S}(\pi_i^y(\mathbf{x}), y_i)$$

3. Accept/Reject label

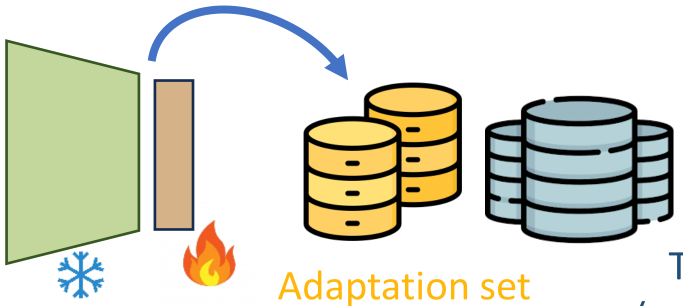
$$\mathcal{C}(\mathbf{x}) = \{y \in \mathcal{Y} : s^y \leq \hat{s}^y\}$$

Access to train data

Multiple (expensive) model fits per sample

Our setting: full conformal adaptation

- **Full conformal adaptation (FCA)**



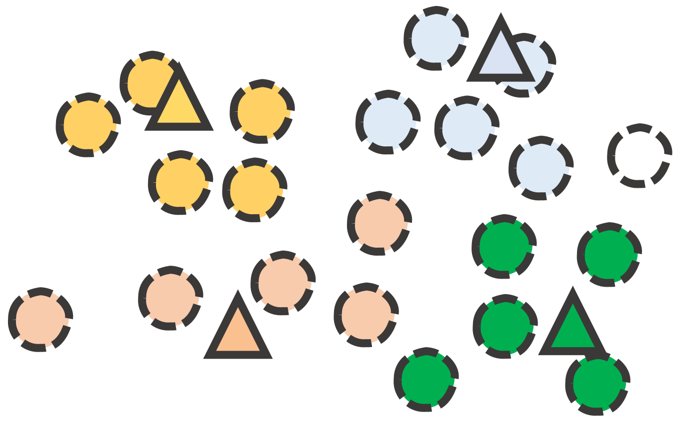
$$s_i^y = \mathcal{S}(\mathbf{p}_i(\mathbf{W}^y), y_i)$$

Access to (<u>small</u>) adaptation data

Múltiple (<u>light</u>) prototype fits per sample

Our setting: full conformal adaptation

- **Interpretation**



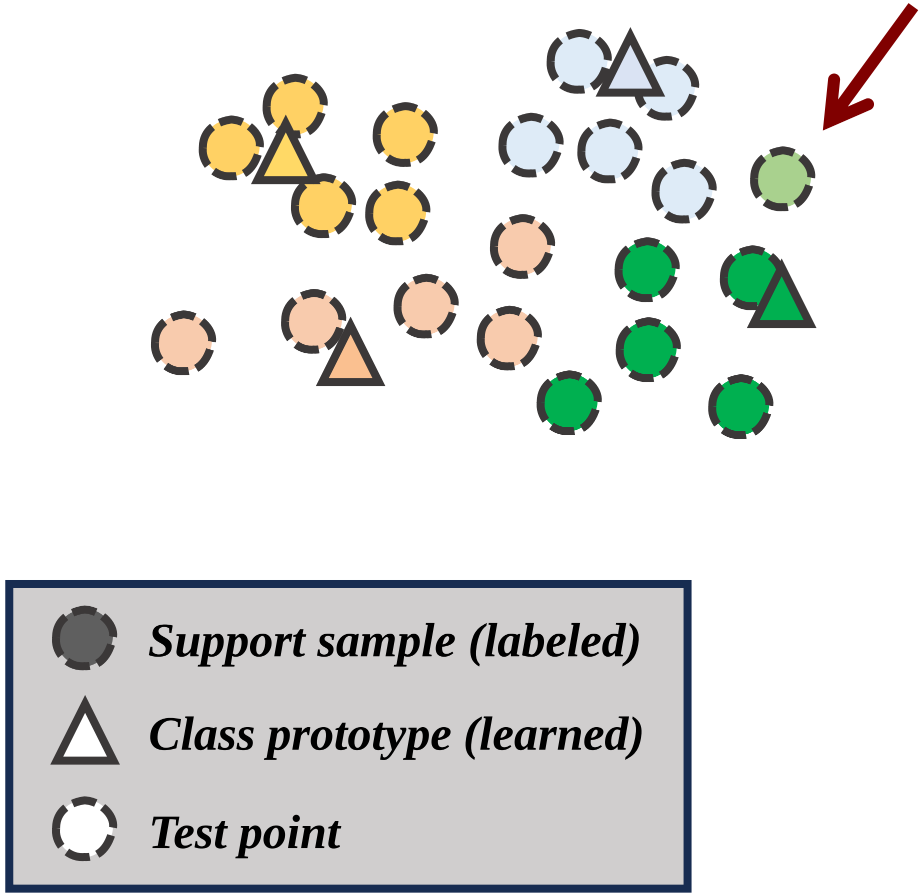
Support sample (labeled)

Class prototype (learned)

Test point

Our setting: full conformal adaptation

- **Interpretation**

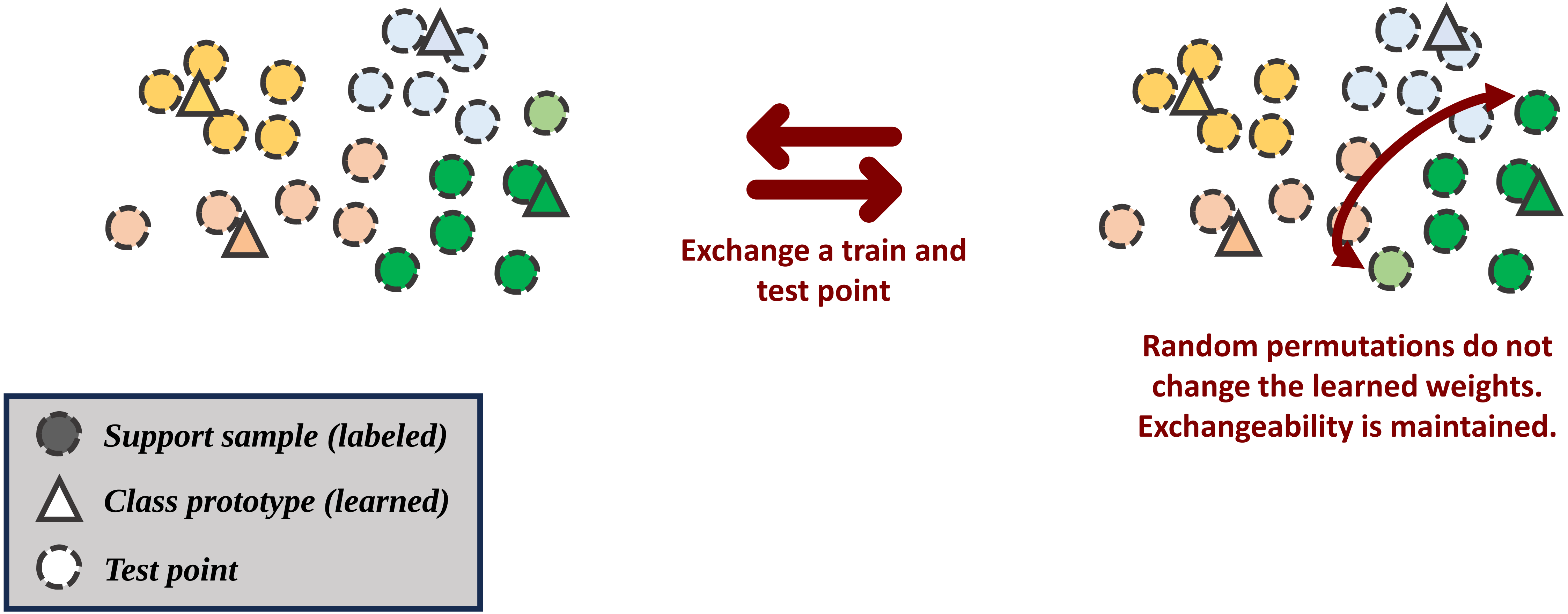


Assume the true label is assigned. In full conformal settings, the test point is used for training (transductive)

$$\mathcal{D}_{train} = \{(\mathbf{x}_1, y_1), ..., (\mathbf{x}_r, y_r), ...(\mathbf{x}_R, y_R), (\mathbf{x}_m, y)\}$$

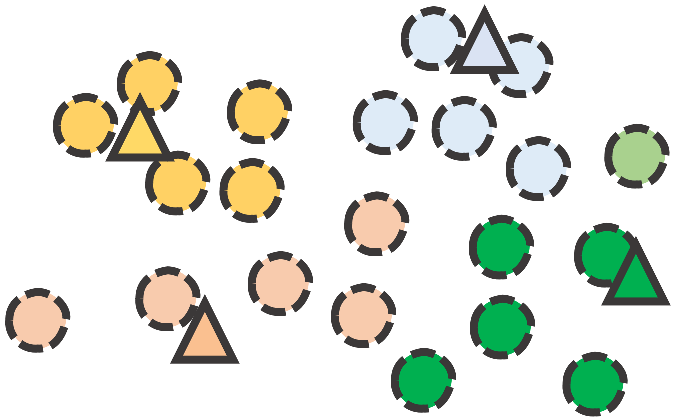
Our setting: full conformal adaptation

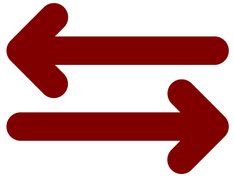
- **Interpretation**



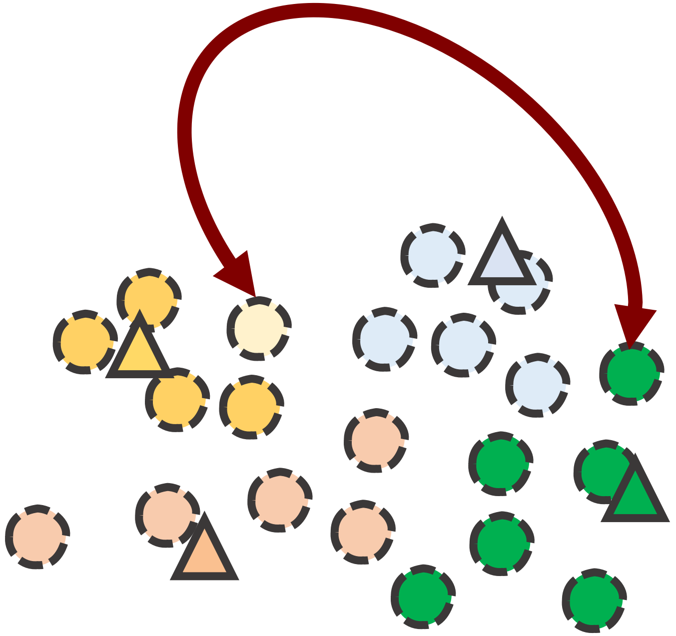
Our setting: full conformal adaptation

- **Interpretation**

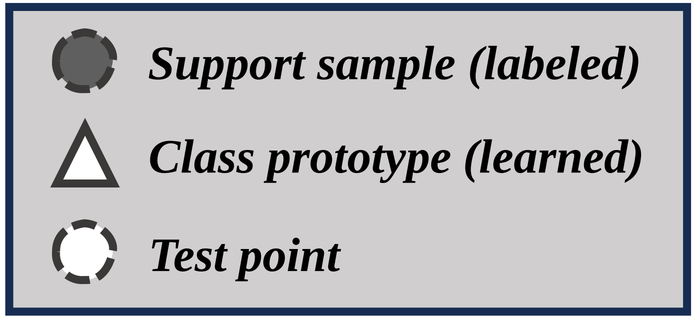




Exchange a train and test point

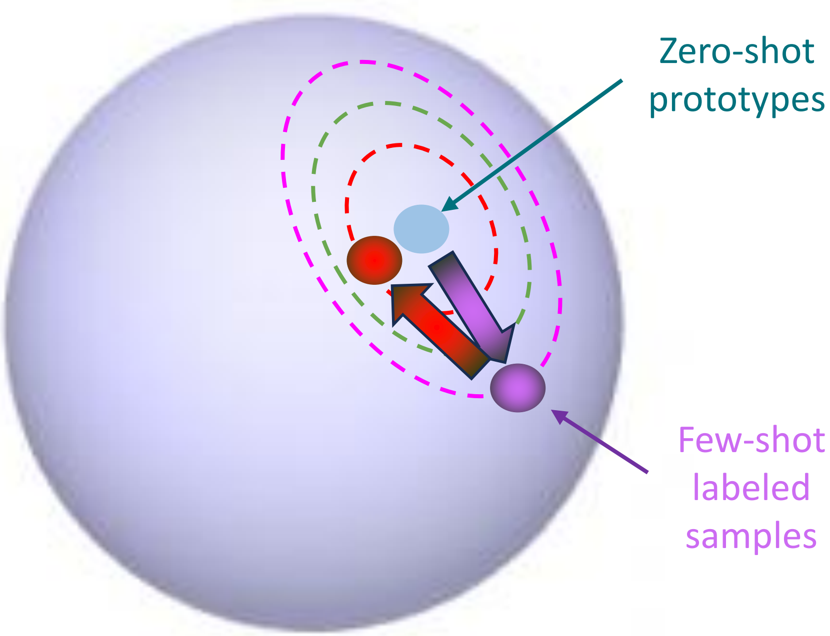


Random permutations do not change the learned weights. Exchangeability is maintained.



Training-free approximation for constrained linear probing

$$\min_{\mathbf{W}} \ \mathcal{L}(\mathbf{W}) = \underbrace{-\frac{1}{N}\sum_{i=1}^{N}\sum_{c=1}^{C} y_{ic}\ln(p_{ic}(\mathbf{W}))}_{\text{Cross-entropy on few-shots}} + \underbrace{\frac{\lambda}{2}\sum_{c=1}^{C}||\mathbf{w}_c - \mathbf{t}_c||_2^2}_{\text{Penalty if the solution deviates from the zero-shot prototypes}}.$$



$$p_c(\mathbf{W}) = \frac{\exp(\mathbf{v}^\top \mathbf{w}_c/\tau)}{\sum_{j=1}^{C}\exp(\mathbf{v}^\top \mathbf{w}_j/\tau)}$$

Training-free approximation for constrained linear probing

$$\min_{\mathbf{W}} \mathcal{L}(\mathbf{W}) = -\frac{1}{N}\sum_{i=1}^{N}\sum_{c=1}^{C} y_{ic} \ln(p_{ic}(\mathbf{W})) + \frac{\lambda}{2}\sum_{c=1}^{C} ||\mathbf{w}_c - \mathbf{t}_c||_2^2.$$

$$\mathcal{L} = g_1 + g_2$$

$$g_1 = -\frac{1}{N}\sum_{i=1}^{N}\sum_{c=1}^{C} y_{ic} \left(\mathbf{v}^\top \mathbf{w}_c/\tau\right) + \frac{\lambda}{2}\sum_{c=1}^{C} ||\mathbf{w}_c - \mathbf{t}_c||_2^2$$

$$g_2 = \frac{1}{N}\sum_{i=1}^{N} \ln\left(\sum_{j=1}^{C} \exp(\mathbf{v}^\top \mathbf{w}_j/\tau)\right)$$

$$p_c(\mathbf{W}) = \frac{\exp(\mathbf{v}^\top \mathbf{w}_c/\tau)}{\sum_{j=1}^{C} \exp(\mathbf{v}^\top \mathbf{w}_j/\tau)}$$

Training-free approximation for constrained linear probing

$$\min_{\mathbf{W}} \mathcal{L}(\mathbf{W}) = -\frac{1}{N}\sum_{i=1}^{N}\sum_{c=1}^{C} y_{ic} \ln(p_{ic}(\mathbf{W})) + \frac{\lambda}{2}\sum_{c=1}^{C} ||\mathbf{w}_c - \mathbf{t}_c||_2^2.$$

$$\mathcal{L} = g_1 + g_2$$

$$g_1 = -\frac{1}{N}\sum_{i=1}^{N}\sum_{c=1}^{C} y_{ic}\left(\mathbf{v}^\top \mathbf{w}_c/\tau\right) + \frac{\lambda}{2}\sum_{c=1}^{C} ||\mathbf{w}_c - \mathbf{t}_c||_2^2$$

$$\frac{\partial g_1}{\partial \mathbf{w}_c} = -\frac{1}{N}\sum_{i=1}^{N}(y_{ic}\,\mathbf{v}/\tau) + \lambda(\mathbf{w}_c - \mathbf{t}_c)$$

$$\mathbf{w}_c^* = \arg\min_{\mathbf{w}_c} \frac{\partial g_1}{\partial \mathbf{w}_c} = \underbrace{\frac{1}{\lambda N\tau}\sum_{i=1}^{N} y_{ic}\mathbf{v}}_{\text{Visual prototypes}} + \underbrace{\mathbf{t}_c}_{\text{Textual}}$$

Training-free approximation for constrained linear probing

SS-Text solver:

$$\mathbf{w}_c^* = \frac{1}{\lambda N \tau} \sum_{i=1}^{N} y_{ic} \mathbf{v} + \mathbf{t}_c$$

Training-free approximation for constrained linear probing

SS-Text solver:

$$\mathbf{w}_c^* = \frac{1}{\lambda N \tau} \sum_{i=1}^{N} y_{ic}\mathbf{v} + \mathbf{t}_c$$

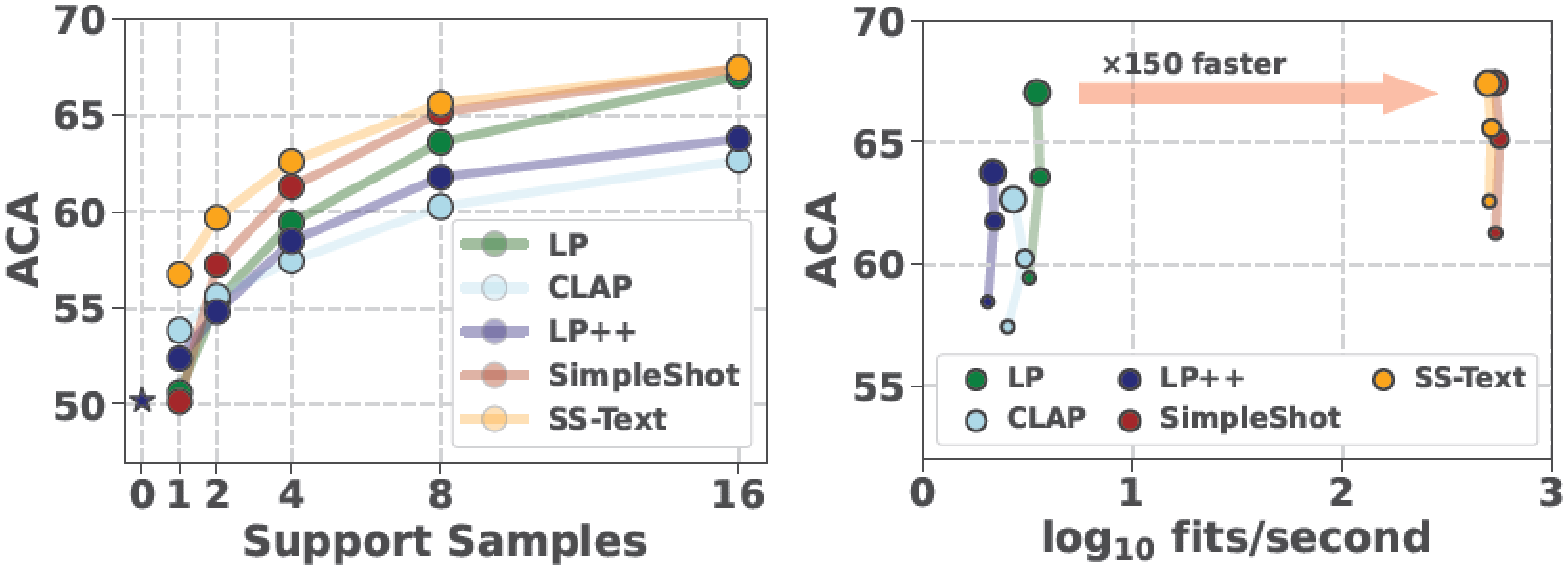
$$\lambda = 1/(N\tau)$$

$$\min_{\mathbf{W}} \; \mathcal{L}(\mathbf{W}) = -\frac{1}{N}\sum_{i=1}^{N}\sum_{c=1}^{C} y_{ic} \ln(p_{ic}(\mathbf{W})) + \frac{\lambda}{2}\sum_{c=1}^{C} ||\mathbf{w}_c - \mathbf{t}_c||_2^2.$$

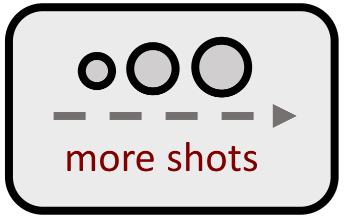
Training-free approximation for constrained linear probing

SS-Text solver:

$$\mathbf{w}_c^* = \frac{1}{\lambda N \tau} \sum_{i=1}^{N} y_{ic} \mathbf{v} + \mathbf{t}_c$$



(a) Few-shot adaptation (b) FCA Efficiency



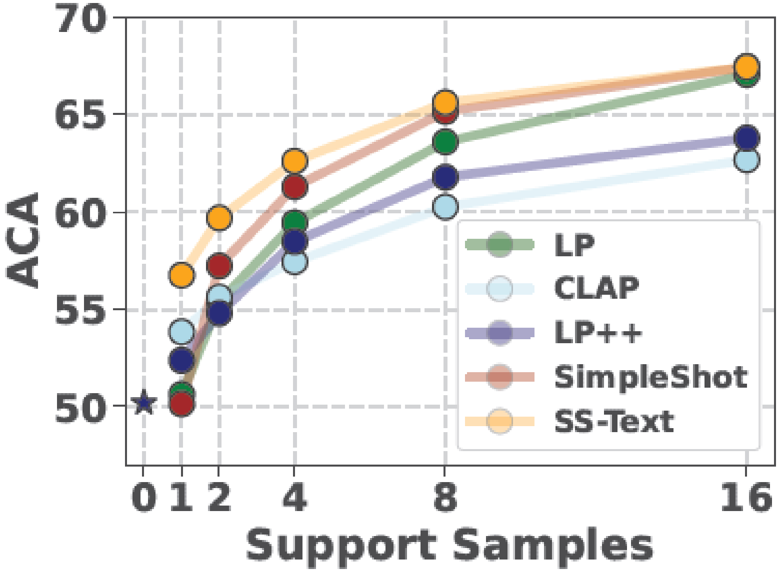
Training-free approximation for constrained linear probing

SS-Text solver:

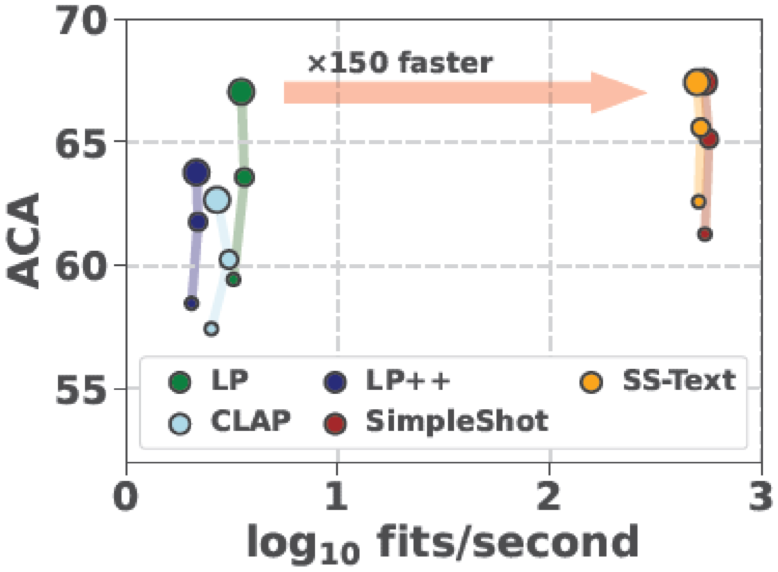
$$\mathbf{w}_c^* = \frac{1}{\lambda N \tau} \sum_{i=1}^{N} y_{ic}\mathbf{v} + \mathbf{t}_c$$

(in a particular use-case)

Skin (histology)
M=30.000, C=16
- **LP: 23h**
- **SS-Text: 15min**



(a) Few-shot adaptation



(b) FCA Efficiency

Conformal Prediction Results

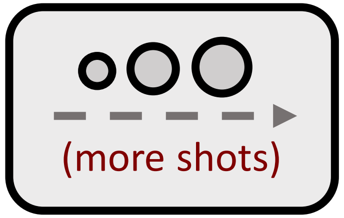
- **Valid (finite-sample) coverage and strong discriminative performance**

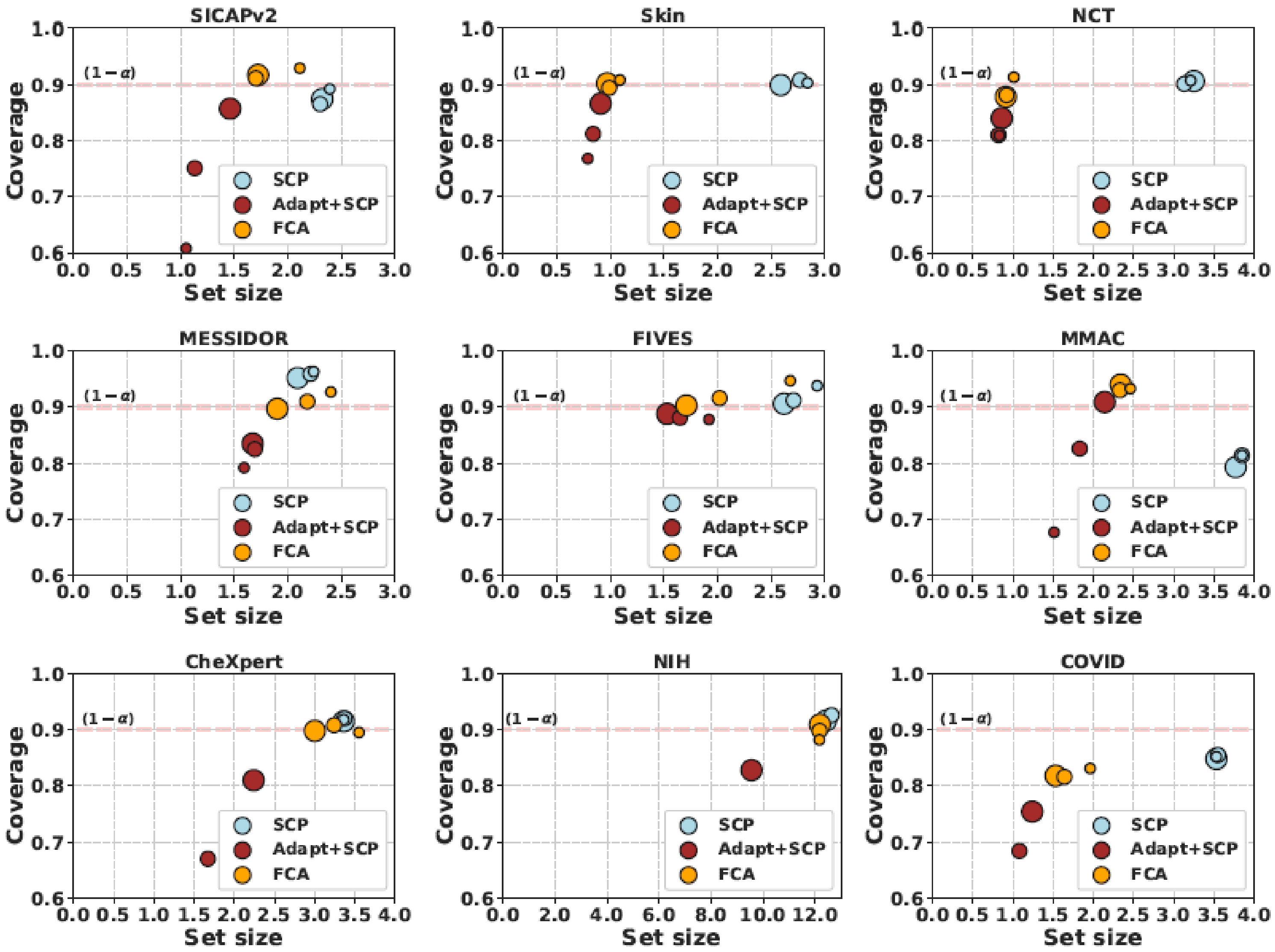
	Method	$\alpha = 0.10$			
		ACA↑	Cov.	Size↓	CCV↓
LAC	SCP	50.2	0.890	3.99	9.96
	Adapt+SCP	**67.1**$_{+16.9}$	0.842	**2.40**$_{-1.59}$	11.17$_{+1.21}$
	FCA (*Ours*)	**67.1**$_{+16.9}$	0.896	2.91$_{-1.08}$	**8.38**$_{-1.58}$
APS	SCP	50.2	0.900	4.05	9.59
	Adapt+SCP	**67.1**$_{+16.9}$	0.858	**2.56**$_{-1.49}$	8.57$_{-1.02}$
	FCA (*Ours*)	**67.1**$_{+16.9}$	0.898	3.06$_{-0.99}$	**6.12**$_{-3.47}$
RAPS	SCP	50.2	0.901	4.16	9.55
	Adapt+SCP	**67.1**$_{+16.9}$	0.856	**2.55**$_{-1.61}$	8.64$_{-0.91}$
	FCA (*Ours*)	**67.1**$_{+16.9}$	0.898	3.05$_{-1.11}$	**6.21**$_{-3.34}$

(16 labeled shots per class for adaptation)

Conformal Prediction Results

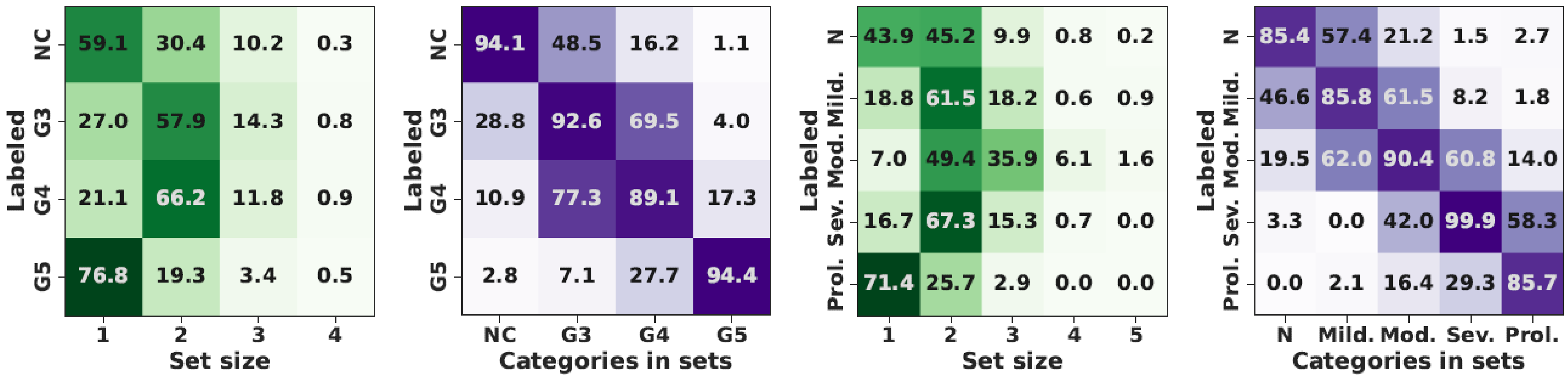
- **Per-dataset analysis**





Conformal Prediction Results

- **Qualitative assessment**



(a) SICAPv2 (b) MESSIDOR

Some take-home messages

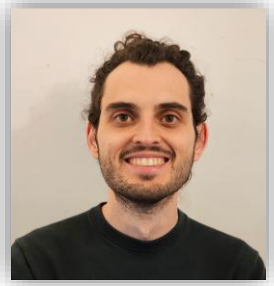
- Medical **VLMs are strong black-box embedding models**. Its transfer potential allows us for data-efficient deployment scenarios.
- Conformal prediction is a promising ML framework for providing practitioners with **user-controlled, real-time guarantees**.
- **From focusing on "accuracy" to "efficiency"/"usefulness"**.
- Many aspects to explore yet, e.g., is **exchangeability** a realistic assumption in medical image analysis?

ÉCOLE DE TECHNOLOGIE SUPÉRIEURE
Le génie pour l'industrie
Université du Québec
CentraleSupélec
université PARIS-SACLAY

2025

Full Conformal Adaptation of Medical Vision-Language Models

Julio Silva-Rodríguez



Leo Fillioux

Paul-Henry Cournède

Maria Vakalopoulou

Stergios Christodoulidis

Ismail Ben Ayed

Jose Dolz

Few-shot adaptation results

- **Ablation study on hyper-parameters for SS-text**

Method	Setting	$K=1$	$K=2$	$K=4$	$K=8$	$K=16$
Zero-shot [33]	(only text)	50.2	50.2	50.2	50.2	50.2
SimpleShot [47]	(only vision)	50.1	57.2	61.3	65.1	**67.4**
TIP-Adapter [51]	(training-free)	55.5	55.5	60.2	62.2	63.2
LP++ [14]	(training-free)	50.7	51.0	51.4	52.1	53.2
SS-Text	Fixed $\lambda = 0.1/\tau$	55.1	59.0	61.5	63.6	64.3
SS-Text	Fixed $\lambda = 1.0/\tau$	53.5	54.1	54.5	54.7	54.6
SS-Text	Fixed $\lambda = 10/\tau$	51.2	51.2	51.2	51.1	51.2
SS-Text	$\lambda_c \simeq$ zero-shot perf. [39]	51.4	58.4	62.6	**65.7**	**67.4**
SS-Text (*Ours*)	$\lambda = 1/(N\tau)$	**56.7**	**59.7**	**62.6**	65.6	**67.4**

Conformal Prediction Results

- **More error rates**

	Method	ACA↑	$\alpha=0.10$ Cov.	$\alpha=0.10$ Size↓	$\alpha=0.10$ CCV↓	$\alpha=0.05$ Cov.	$\alpha=0.05$ Size↓	$\alpha=0.05$ CCV↓
LAC	SCP	50.2	0.890	3.99	9.96	0.951	4.88	5.68
LAC	Adapt+SCP	**67.1**$_{+16.9}$	0.842	**2.40**$_{-1.59}$	11.17$_{+1.21}$	0.921	**3.07**$_{-1.81}$	6.87$_{+1.19}$
LAC	FCA (*Ours*)	**67.1**$_{+16.9}$	0.896	2.91$_{-1.08}$	**8.38**$_{-1.58}$	0.952	3.56$_{-1.32}$	**5.02**$_{-0.66}$
APS	SCP	50.2	0.900	4.05	9.59	0.952	4.88	5.54
APS	Adapt+SCP	**67.1**$_{+16.9}$	0.858	**2.56**$_{-1.49}$	8.57$_{-1.02}$	0.924	**3.19**$_{-1.69}$	6.08$_{+0.54}$
APS	FCA (*Ours*)	**67.1**$_{+16.9}$	0.898	3.06$_{-0.99}$	**6.12**$_{-3.47}$	0.949	3.67$_{-1.21}$	**4.24**$_{-1.30}$
RAPS	SCP	50.2	0.901	4.16	9.55	0.952	5.12	5.57
RAPS	Adapt+SCP	**67.1**$_{+16.9}$	0.856	**2.55**$_{-1.61}$	8.64$_{-0.91}$	0.923	**3.17**$_{-1.95}$	6.12$_{+0.55}$
RAPS	FCA (*Ours*)	**67.1**$_{+16.9}$	0.898	3.05$_{-1.11}$	**6.21**$_{-3.34}$	0.951	3.66$_{-1.46}$	**4.23**$_{-1.34}$

(16 labeled shots per class for adaptation)