

# Conformal Prediction for Reliable AI Uncertainty Quantification

**Jungeum Kim**

**March 18, 2026**



**NC STATE**  
UNIVERSITY



***“I feel very itchy  
but do not know what it is”***

***6/15/2018, at Purdue***

**Spreading, red, itchy, dry rash on my  
neck with some swelling or hives**

***6/15/2018, 12:36 pm, Purdue***

# Can ChatGPT diagnose me?



6/15/2018, 12:36 pm, Purdue

ChatGPT 4o ▾

Share

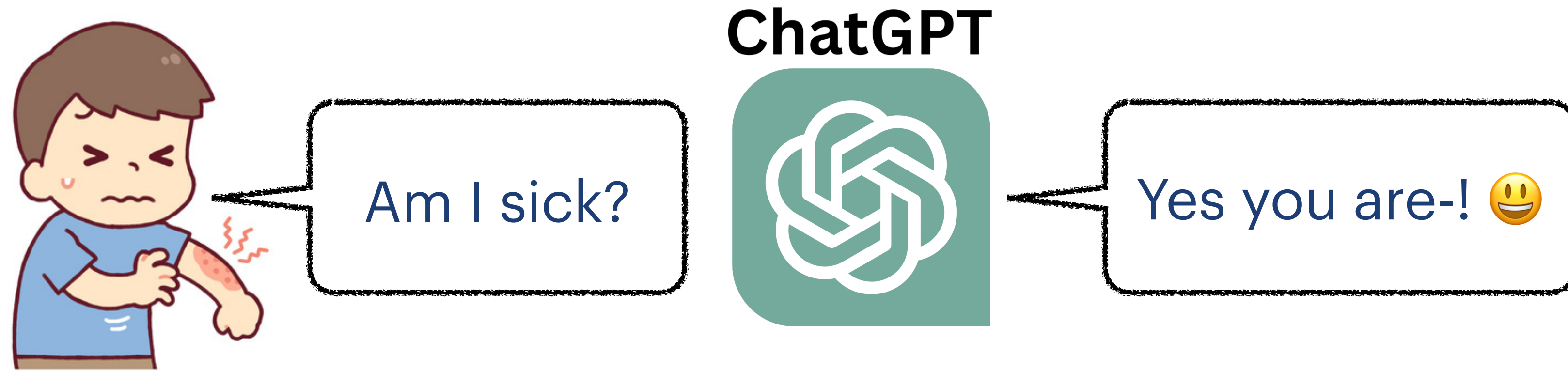


I have a spreading, red, itchy, dry rash on my neck with some swelling or hives. What's the most likely diagnosis?



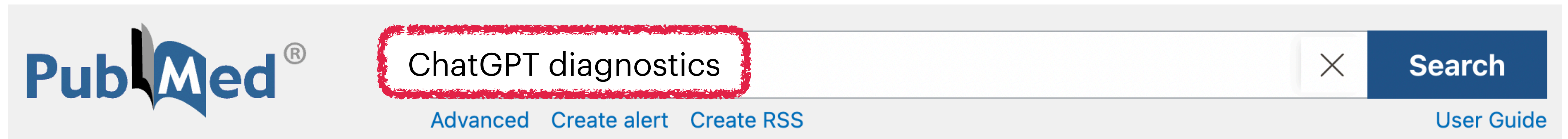
1. Contact dermatitis
2. Eczema
3. Urticaria
4. Psoriasis
5. Drug reaction
6. Fungal infection
7. Heat rash
8. Lupus
9. Rosacea
10. Lichen planus

# Controversial use of ChatGPT in healthcare?



Medical research interest in diagnostics through ChatGPT is increasing

-> Dentistry, eye diseases, and cancer,...



1209 results; Up to 2022: zero\*, 2023: 393, 2024: 901, 2025: 2042 +

\* ChatGPT was launched on November 30, 2022

## Ex) Skin disease diagnosis

Garg et al, 2023. Exploring the role of ChatGPT in patient care (diagnosis and treatment) and medical research: A systematic review...

Kuroiwa et al, 2023. The Potential of ChatGPT as a Self-Diagnostic Tool in Common Orthopedic Diseases: Exploratory Study....

Shieh et al, 2024. Assessing ChatGPT 4.0's test performance and clinical diagnostic accuracy on USMLE STEP 2 CK...

# Is AI's internal UQ trustworthy?

AI  
UQ



6/15/2018, 12:36 pm, Purdue

ChatGPT 4o

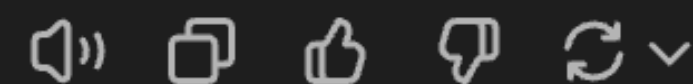
Share

Give me probability on those sum to 1



Here are the probabilities summing to 1:

1. Contact dermatitis: 0.30
2. Eczema: 0.20
3. Urticaria: 0.15
4. Psoriasis: 0.10
5. Drug reaction: 0.10
6. Fungal infection: 0.05
7. Heat rash: 0.05
8. Lupus: 0.02
9. Rosacea: 0.02
10. Lichen planus: 0.01



**Why Conformal Prediction?**

**... and Why *Not*?!**

# Split Conformal Prediction Vovk et al. (2005), Lei et al. (2015)

**Black-box Model:**  $\hat{\mu}(x)$ , **Test Data:**  $\mathcal{D}_{\text{test}} = \{(X_{\text{test}}, Y_{\text{test}})\}$

**Goal:**  $\mathbb{P} \left( Y_{\text{test}} \in \hat{C}_n(X_{\text{test}}) \right) \geq 1 - \alpha$

$\hat{C}_n$ : Based on the **model performance** on the calibration dataset  $\mathcal{D}_{\text{calib}} = \{(X_1, Y_1), \dots, (X_n, Y_n)\}$

**Conformal Score:** **negative model performance**  $\{S_i = S(\hat{\mu}(X_i), Y_i)\}_{i=1}^n$

# Split Conformal Prediction Vovk et al. (2005), Lei et al. (2015)

**Black-box Model:**  $\hat{\mu}(x)$ , **Test Data:**  $\mathcal{D}_{\text{test}} = \{(X_{\text{test}}, Y_{\text{test}})\}$

**Goal:**  $\mathbb{P} \left( Y_{\text{test}} \in \hat{C}_n(X_{\text{test}}) \right) \geq 1 - \alpha$

$\hat{C}_n$ : Based on the **model performance** on the calibration dataset  $\mathcal{D}_{\text{calib}} = \{(X_1, Y_1), \dots, (X_n, Y_n)\}$

**Conformal Score:** **negative model performance**  $\{S_i = S(\hat{\mu}(X_i), Y_i)\}_{i=1}^n$

**Coverage Guarantee:** For  $C = \lceil (1 - \alpha)(n + 1) \rceil$ -th smallest value in  $\{S_i\}_{i=1}^n$ ,

$$1 - \alpha \leq \mathbb{P} \left( S_{\text{test}} \leq C \right) = \frac{\lceil (1 - \alpha)(n + 1) \rceil}{n + 1} \leq 1 - \alpha + \frac{1}{n + 1}$$

Equality and upper bound: when there is no tie the conformity scores

# Split Conformal Prediction Vovk et al. (2005), Lei et al. (2015)

Black-box Model:  $\hat{\mu}(x)$ , Test Data:  $\mathcal{D}_{\text{test}} = \{(X_{\text{test}}, Y_{\text{test}})\}$

Goal:  $\mathbb{P} \left( Y_{\text{test}} \in \hat{C}_n(X_{\text{test}}) \right) \geq 1 - \alpha$

$\hat{C}_n$ : Based on the model performance on the calibration dataset  $\mathcal{D}_{\text{calib}} = \{(X_1, Y_1), \dots, (X_n, Y_n)\}$

Conformal Score: **negative model performance**  $\{S_i = S(\hat{\mu}(X_i), Y_i)\}_{i=1}^n$

Coverage Guarantee: For  $C = \lceil (1 - \alpha)(n + 1) \rceil$ -th smallest value in  $\{S_i\}_{i=1}^n$ ,

$$1 - \alpha \leq \mathbb{P} \left( S_{\text{test}} \leq C \right) = \frac{\lceil (1 - \alpha)(n + 1) \rceil}{n + 1} \leq 1 - \alpha + \frac{1}{n + 1}$$

Equality and upper bound: when there is no tie the conformity scores

No assumptions except **exchangeability assumption**:

$$(X_1, Y_1), \dots, (X_n, Y_n), (X_{\text{test}}, Y_{\text{test}}) \stackrel{d}{=} (X_{\sigma(1)}, Y_{\sigma(1)}), \dots, (X_{\sigma(n)}, Y_{\sigma(n)}), (Y_{\sigma(\text{test})}, X_{\sigma(\text{test})})$$

$\Rightarrow$  Rank = uniform-!!  $\text{Rank}(S_{\text{test}}) \sim \text{Unif}(1, \dots, n + 1)$ , where  $S_{\text{test}} = S(X_{\text{test}}, Y_{\text{test}})$

# Split Conformal Prediction Vovk et al. (2005), Lei et al. (2015)

**Black-box Model:**  $\hat{\mu}(x)$ , **Test Data:**  $\mathcal{D}_{\text{test}} = \{(X_{\text{test}}, Y_{\text{test}})\}$

**Goal:**  $\mathbb{P} \left( Y_{\text{test}} \in \hat{C}_n(X_{\text{test}}) \right) \geq 1 - \alpha$

$\hat{C}_n$ : Based on the **model performance** on the calibration dataset  $\mathcal{D}_{\text{calib}} = \{(X_1, Y_1), \dots, (X_n, Y_n)\}$

**Conformal Score:** **negative model performance**  $\{S_i = S(\hat{\mu}(X_i), Y_i)\}_{i=1}^n$

**Coverage Guarantee:** For  $C = \lceil (1 - \alpha)(n + 1) \rceil$ -th smallest value in  $\{S_i\}_{i=1}^n$ ,

$$1 - \alpha \leq \mathbb{P} \left( S_{\text{test}} \leq C \right) = \frac{\lceil (1 - \alpha)(n + 1) \rceil}{n + 1} \leq 1 - \alpha + \frac{1}{n + 1}$$

Equality and upper bound: when there is no tie the conformity scores

**Regression:**  $S(x, y) = |\hat{\mu}(x) - y| \rightarrow \hat{C}_n = [\hat{\mu}(X_{\text{test}}) - C, \hat{\mu}(X_{\text{test}}) + C]$

**Classification:**  $S(x, y) = 1 - \hat{\mu}_y(x) \rightarrow \hat{C}_n = \{k \mid 1 - \hat{\mu}_k(X_{\text{test}}) \leq C\}$ ,  
where  $\sum_{k=1}^{\mathcal{K}} \hat{\mu}_k(x) = 1$  ( $\mathcal{K}$  : number of classes).

# Split Conformal Prediction Vovk et al. (2005), Lei et al. (2015)

Black-box Model:  $\hat{\mu}(x)$ , Test Data:  $\mathcal{D}_{\text{test}} = \{(X_{\text{test}}, Y_{\text{test}})\}$

Goal:  $\mathbb{P} \left( Y_{\text{test}} \in \hat{C}_n(X_{\text{test}}) \right) \geq 1 - \alpha$

$\hat{C}_n$ : Based on the model performance on the calibration dataset  $\mathcal{D}_{\text{calib}} = \{(X_1, Y_1), \dots, (X_n, Y_n)\}$

Conformal Score: **negative** model performance  $\{S_i = S(\hat{\mu}(X_i), Y_i)\}_{i=1}^n$

Pros:

**No distributional assumptions**

Only need **exchangeability assumption** between the calibration and test sets

And, independence between (cal, test) vs (training) sets.

# Split Conformal Prediction Vovk et al. (2005), Lei et al. (2015)

Black-box Model:  $\hat{\mu}(x)$ , Test Data:  $\mathcal{D}_{\text{test}} = \{(X_{\text{test}}, Y_{\text{test}})\}$

Goal:  $\mathbb{P} \left( Y_{\text{test}} \in \hat{C}_n(X_{\text{test}}) \right) \geq 1 - \alpha$

$\hat{C}_n$ : Based on the model performance on the calibration dataset  $\mathcal{D}_{\text{calib}} = \{(X_1, Y_1), \dots, (X_n, Y_n)\}$

Conformal Score: negative model performance  $\{S_i = S(\hat{\mu}(X_i), Y_i)\}_{i=1}^n$

Cons: Too large prediction set 

Too simple settings

# Split Conformal Prediction Vovk et al. (2005), Lei et al. (2015)

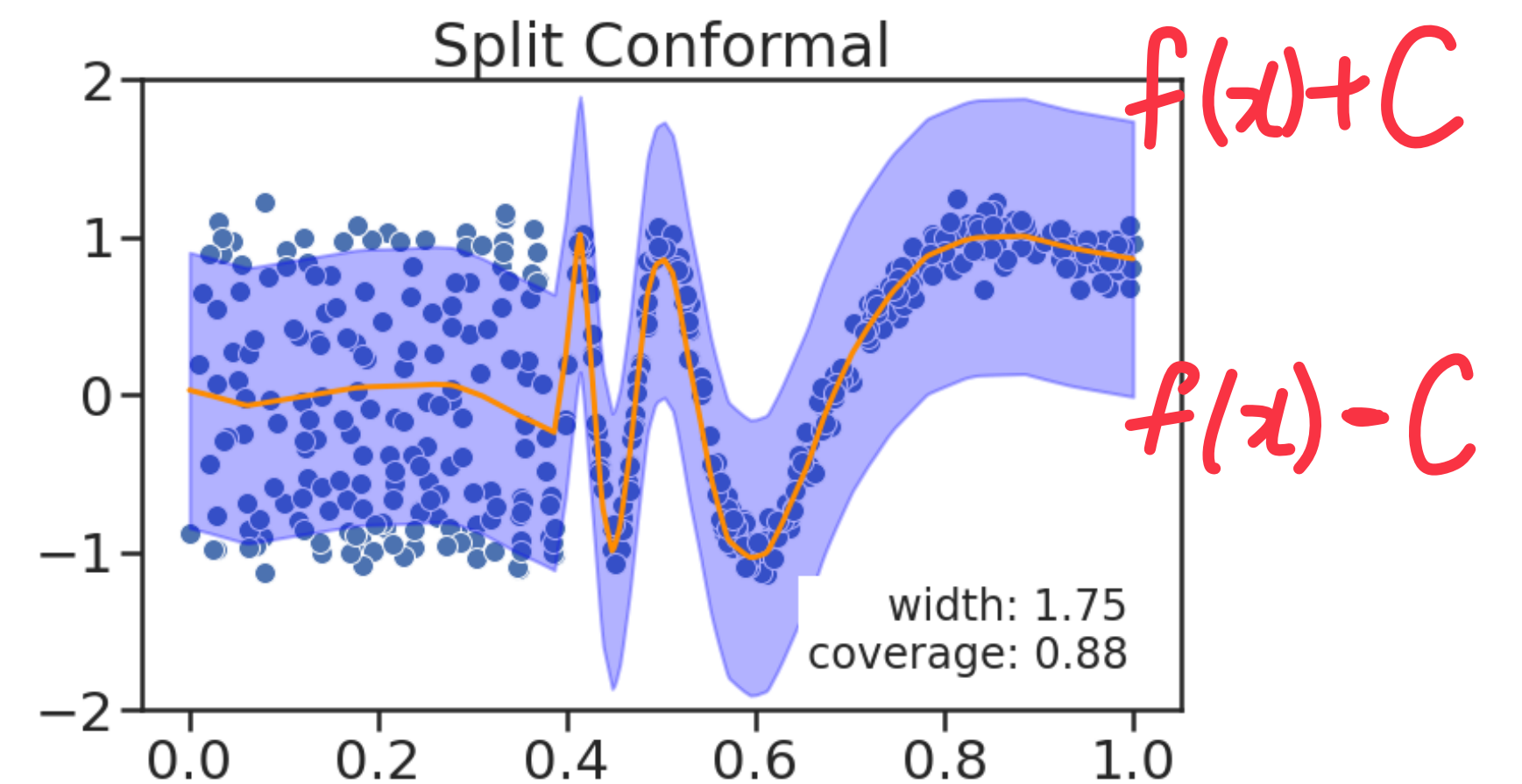
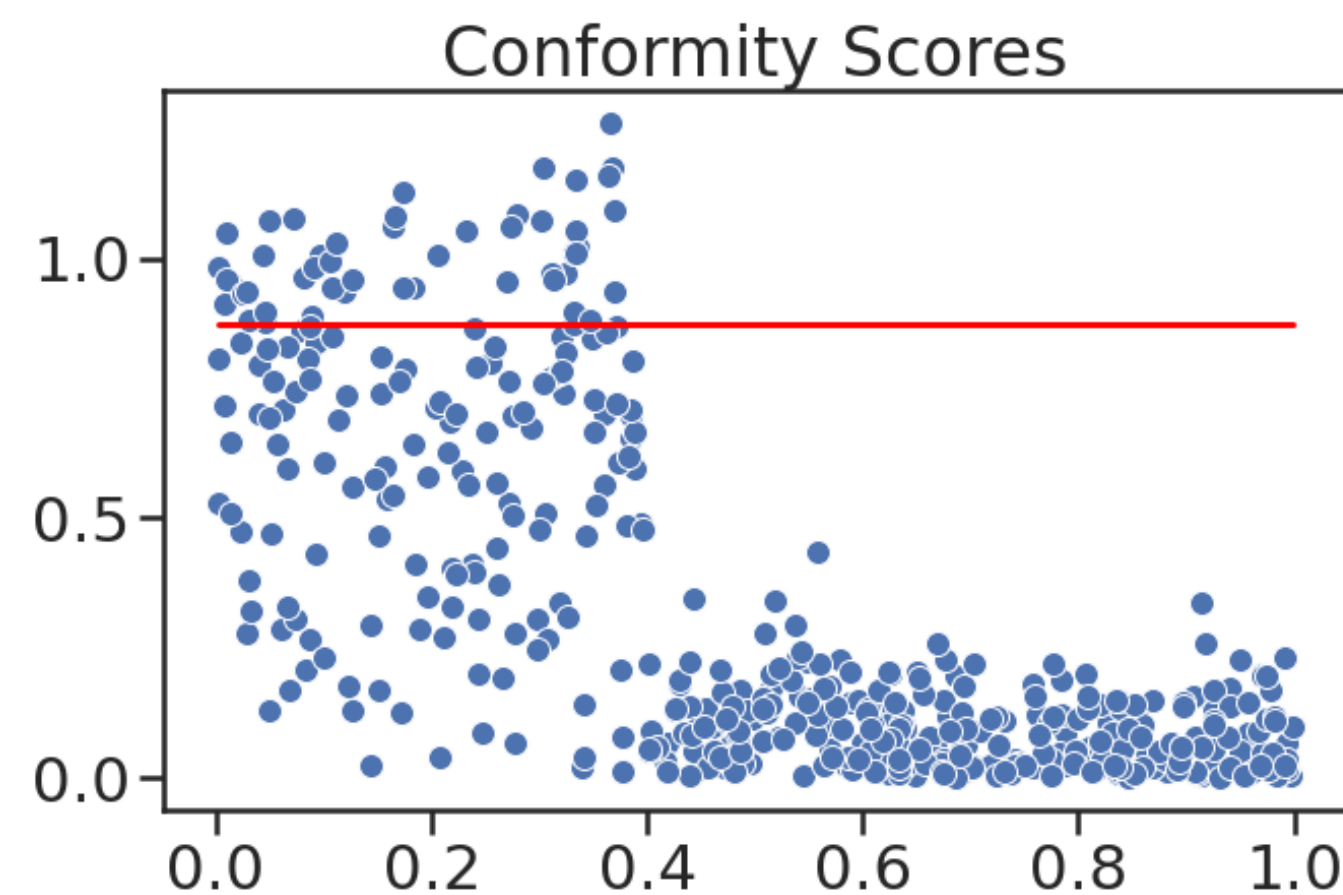
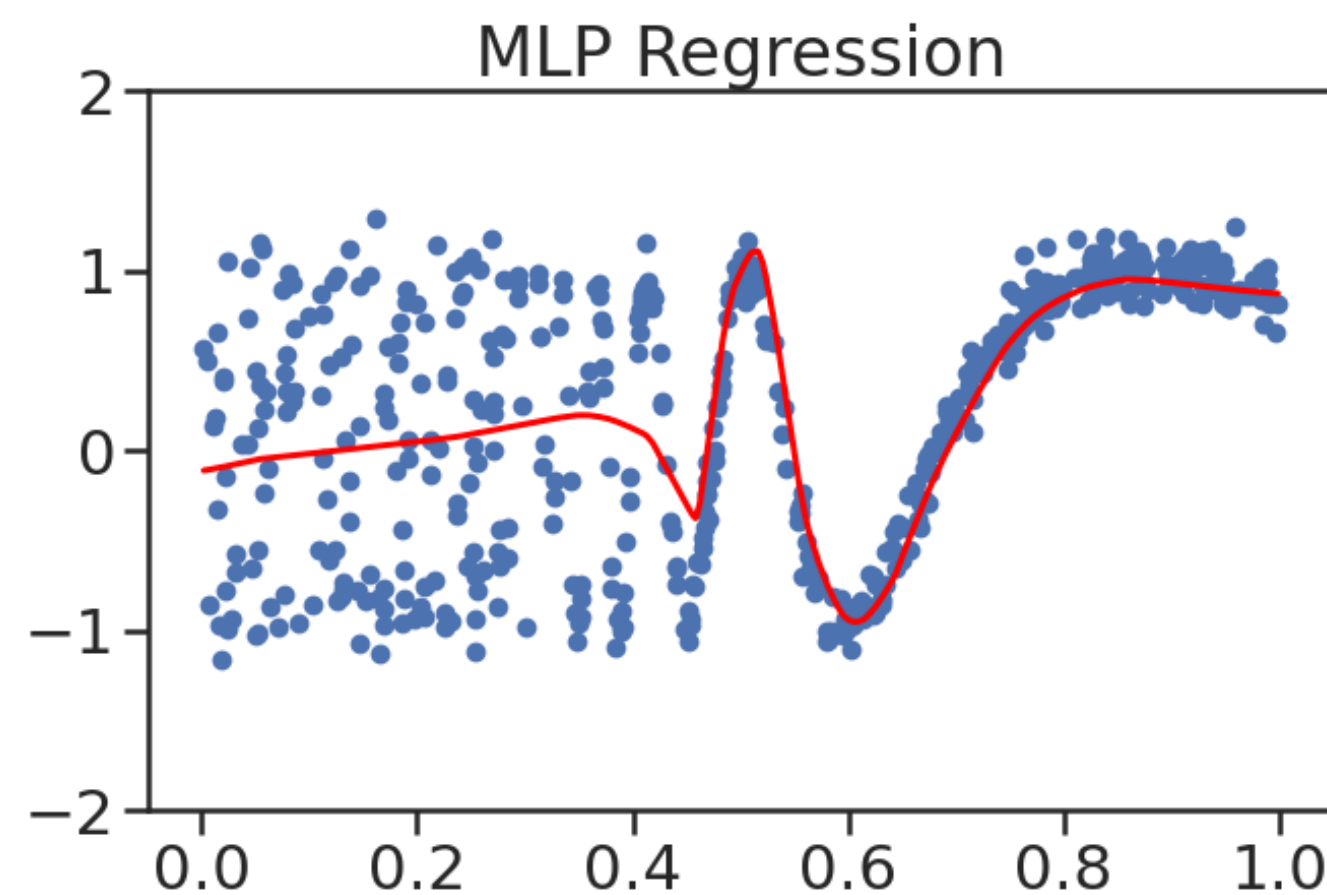
Black-box Model:  $\hat{\mu}(x)$ , Test Data:  $\mathcal{D}_{\text{test}} = \{(X_{\text{test}}, Y_{\text{test}})\}$

Goal:  $\mathbb{P} \left( Y_{\text{test}} \in \hat{C}_n(X_{\text{test}}) \right) \geq 1 - \alpha$

$\hat{C}_n$ : Based on the model performance on the calibration dataset  $\mathcal{D}_{\text{calib}} = \{(X_1, Y_1), \dots, (X_n, Y_n)\}$

Conformal Score: **negative** model performance  $\{S_i = S(\hat{\mu}(X_i), Y_i)\}_{i=1}^n$

Cons: **Too large prediction set** 



Regression:  $S(x, y) = |\hat{\mu}(x) - y| \rightarrow \hat{C}_n = [\hat{\mu}(X_{\text{test}}) - C, \hat{\mu}(X_{\text{test}}) + C]$

# Split Conformal Prediction Vovk et al. (2005), Lei et al. (2015)

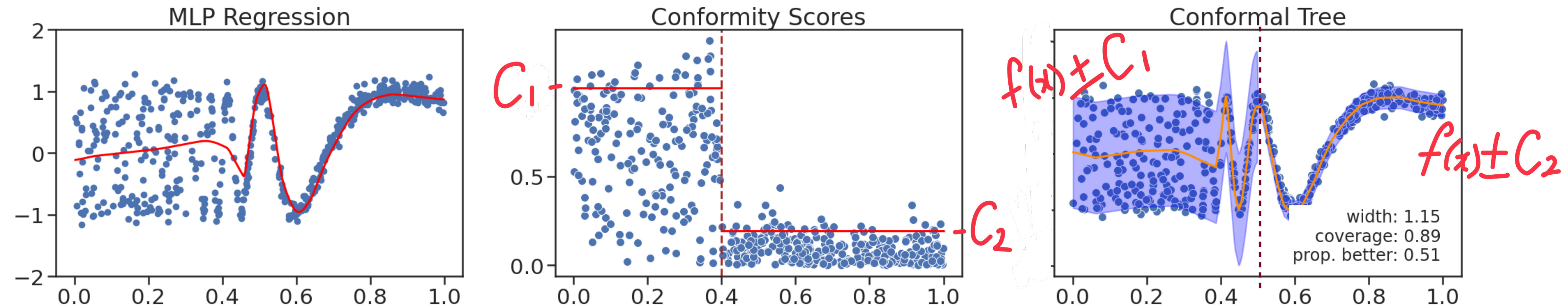
Black-box Model:  $\hat{\mu}(x)$ , Test Data:  $\mathcal{D}_{\text{test}} = \{(X_{\text{test}}, Y_{\text{test}})\}$

Goal:  $\mathbb{P} \left( Y_{\text{test}} \in \hat{C}_n(X_{\text{test}}) \right) \geq 1 - \alpha$

$\hat{C}_n$ : Based on the model performance on the calibration dataset  $\mathcal{D}_{\text{calib}} = \{(X_1, Y_1), \dots, (X_n, Y_n)\}$

Conformal Score: **negative model performance**  $\{S_i = S(\hat{\mu}(X_i), Y_i)\}_{i=1}^n$

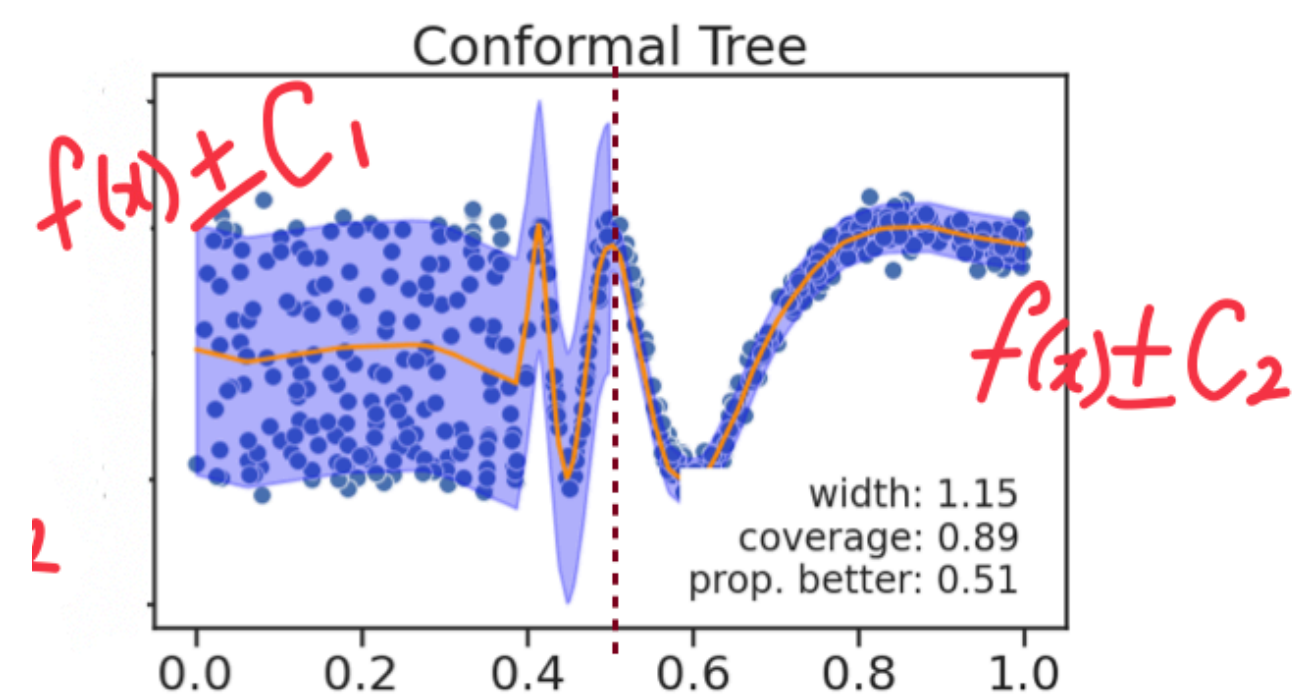
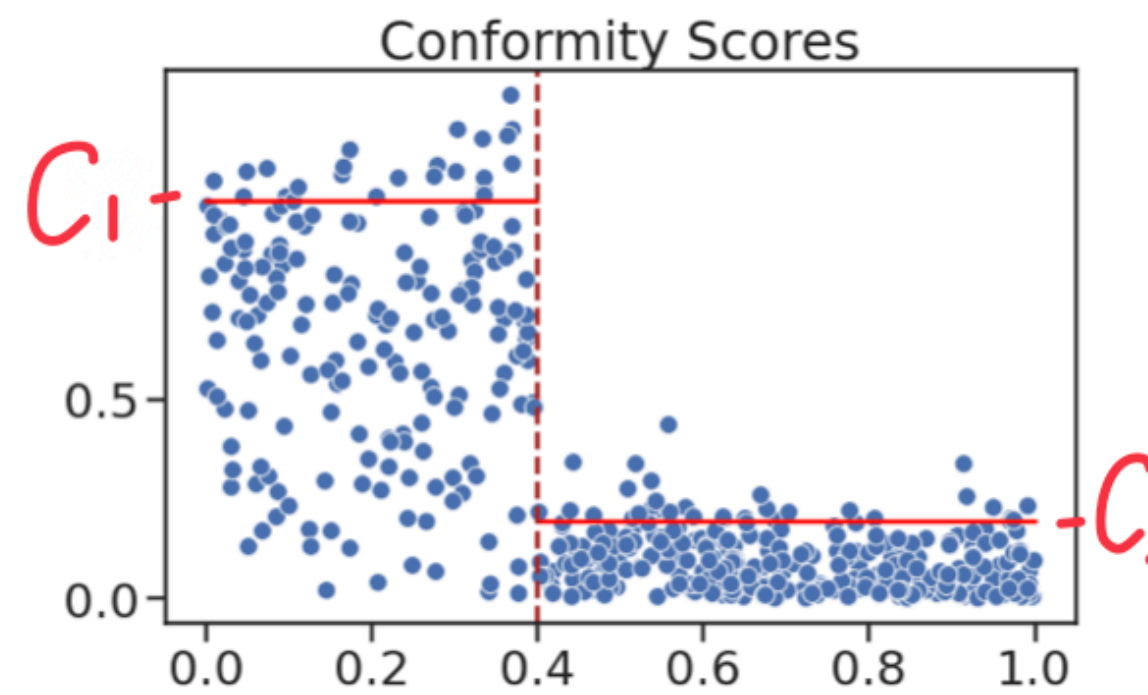
Cons: **Too large prediction set**



**Regression:**  $S(x, y) = |\hat{\mu}(x) - y| \rightarrow \hat{C}_n = [\hat{\mu}(X_{\text{test}}) - C, \hat{\mu}(X_{\text{test}}) + C]$

# Three Approaches to Computing Local Thresholds (Adaptivity)

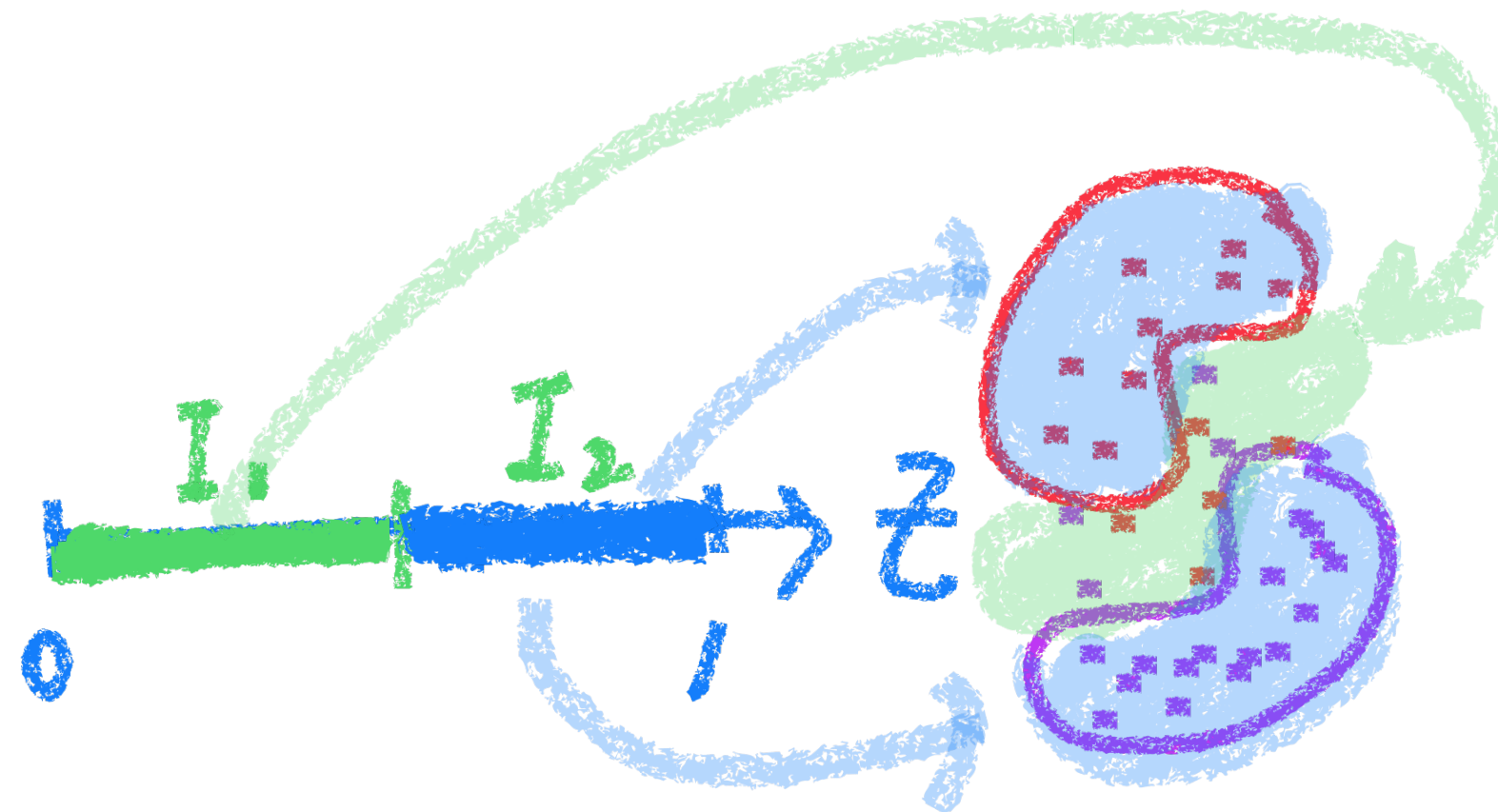
Input space partition:  
homogeneous group



Kim, O'hagan, and Rockova (2025)

Confidence proxy space (1-dim)  
partition: homogeneous group

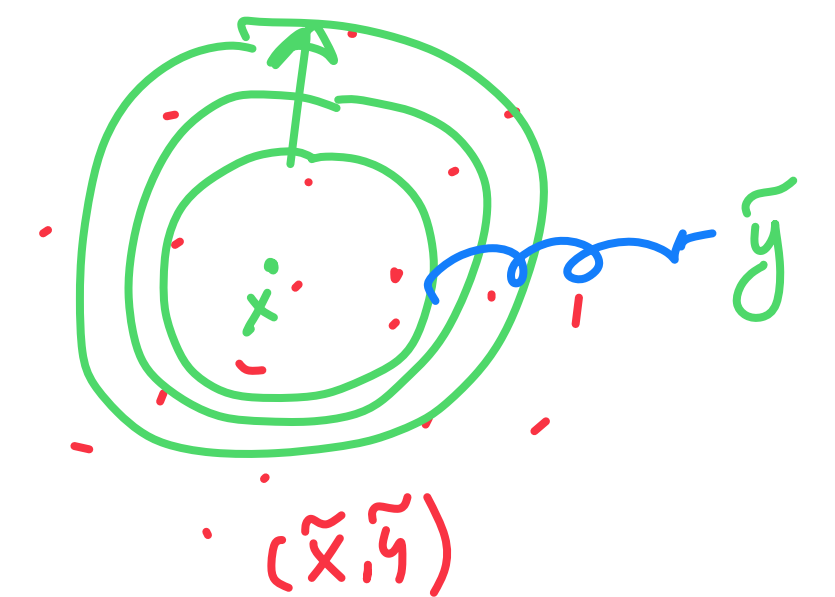
$$Z(x) = \min_y \hat{p}(y|x)$$



Kim, Zhao, Zhu, Zhang, and Yang (2026)

KNN local grouping:  
Embedding space partition:  
homogeneous group

$$\hat{C}_K(X) = \text{unique}\{\tilde{y} \text{ of } (\tilde{x}, \tilde{y}) \text{ in } D_{ref} \text{ s.t. } \tilde{x} \in N_K(X)\}$$

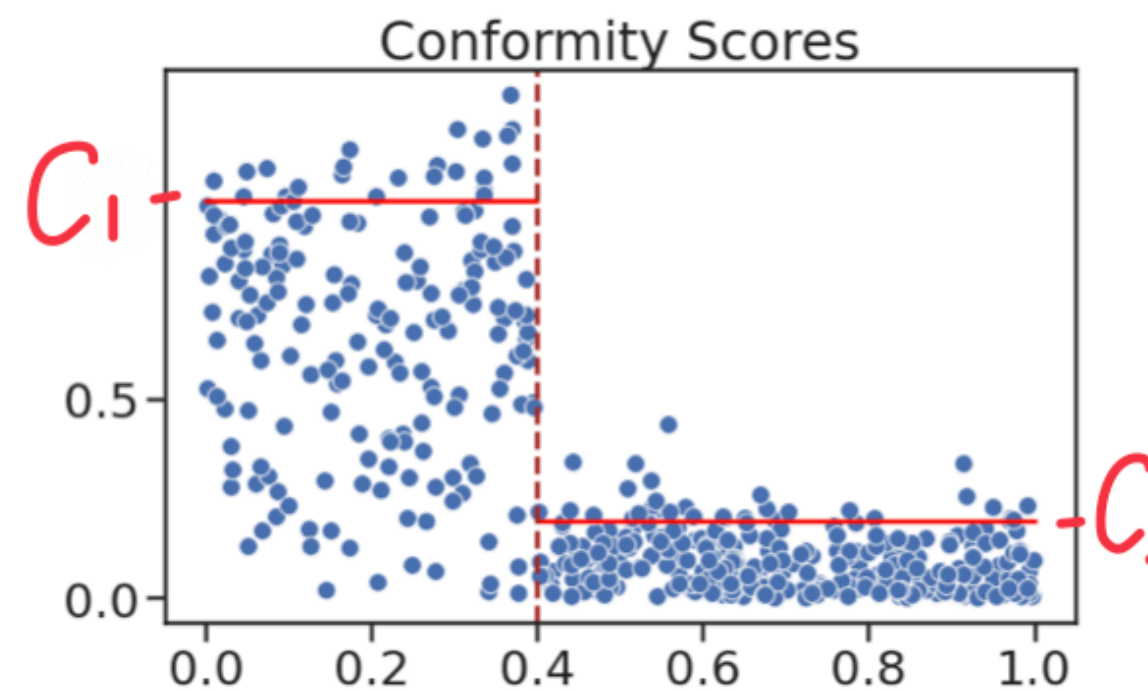


$$S(x, y) = \operatorname{argmin}_{K \in \mathbb{N}} \{y \in \hat{C}_K(X)\}$$

Feng, Reich, ...,and Kim (2026)

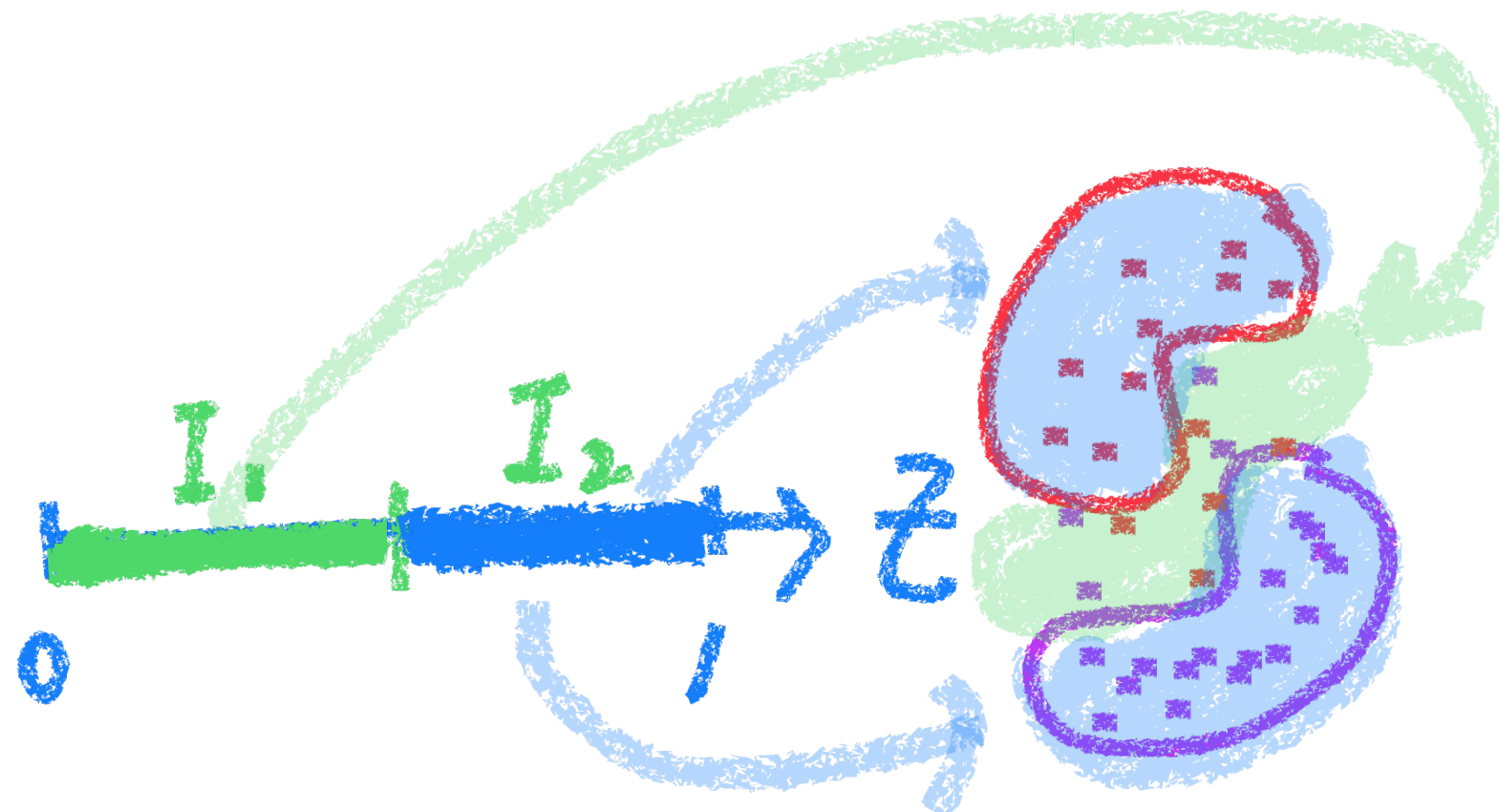
# Three Approaches to Computing Local Thresholds (Adaptivity)

Input space partition:  
homogeneous group



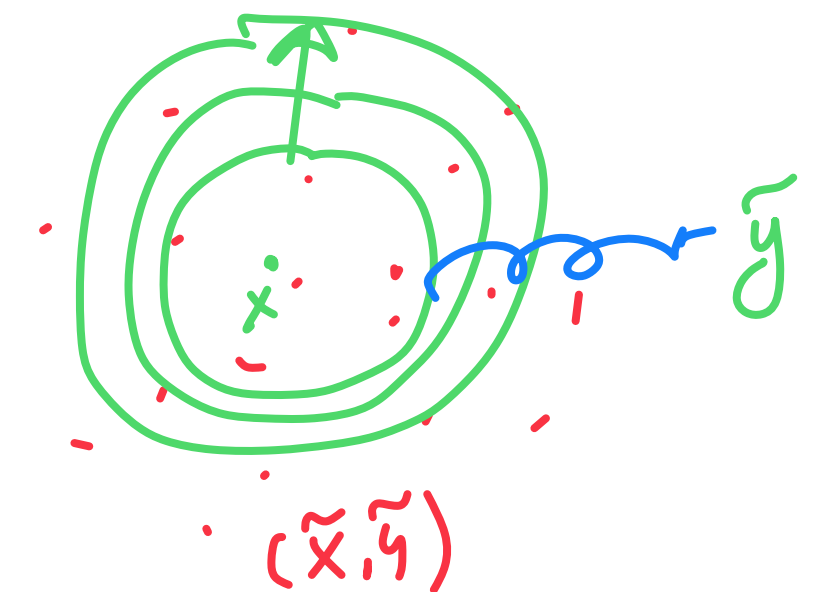
Confidence proxy space (1-dim)  
partition: homogeneous group

$$Z(x) = \min_y \hat{p}(y|x)$$



KNN local grouping:  
Embedding space partition:  
homogeneous group

$$\hat{C}_K(X) = \text{unique}\{\tilde{y} \text{ of } (\tilde{x}, \tilde{y}) \text{ in } D_{ref} \text{ s.t. } \tilde{x} \in N_K(X)\}$$



$$S(x, y) = \operatorname{argmin}_{K \in \mathbb{N}} \{y \in \hat{C}_K(X)\}$$

# Dataset Economy in ML

Training Dataset:  
(Unavailable)

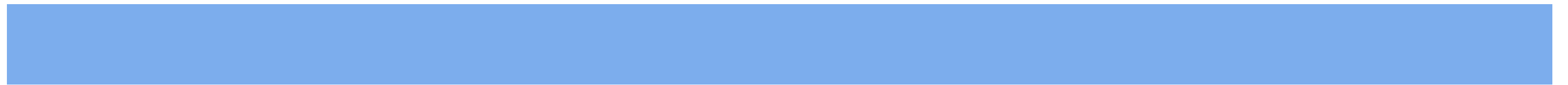


Test Dataset :



# Dataset Economy in Conformal Prediction

Training Dataset:  
(Unavailable)



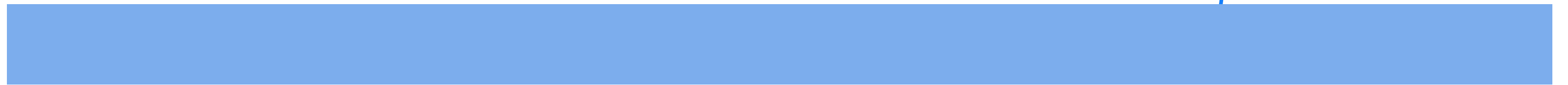
Test Dataset

Calibration Dataset



# Dataset Economy for Local Adaptivity: Pain

Training Dataset:  
(Unavailable)



Test Dataset

Calibration Dataset



Adaptivity Training Dataset

# Conformal Tree

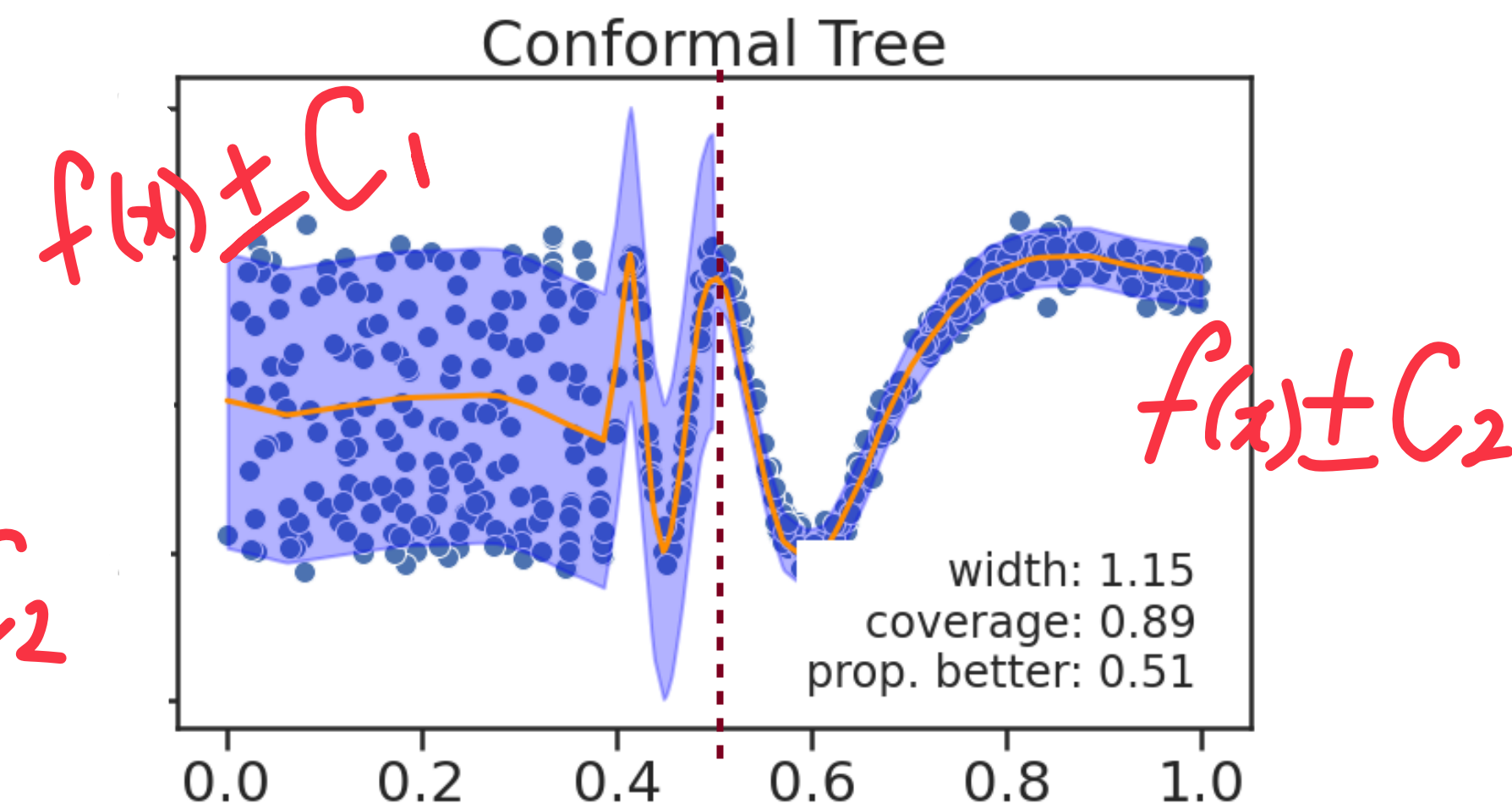
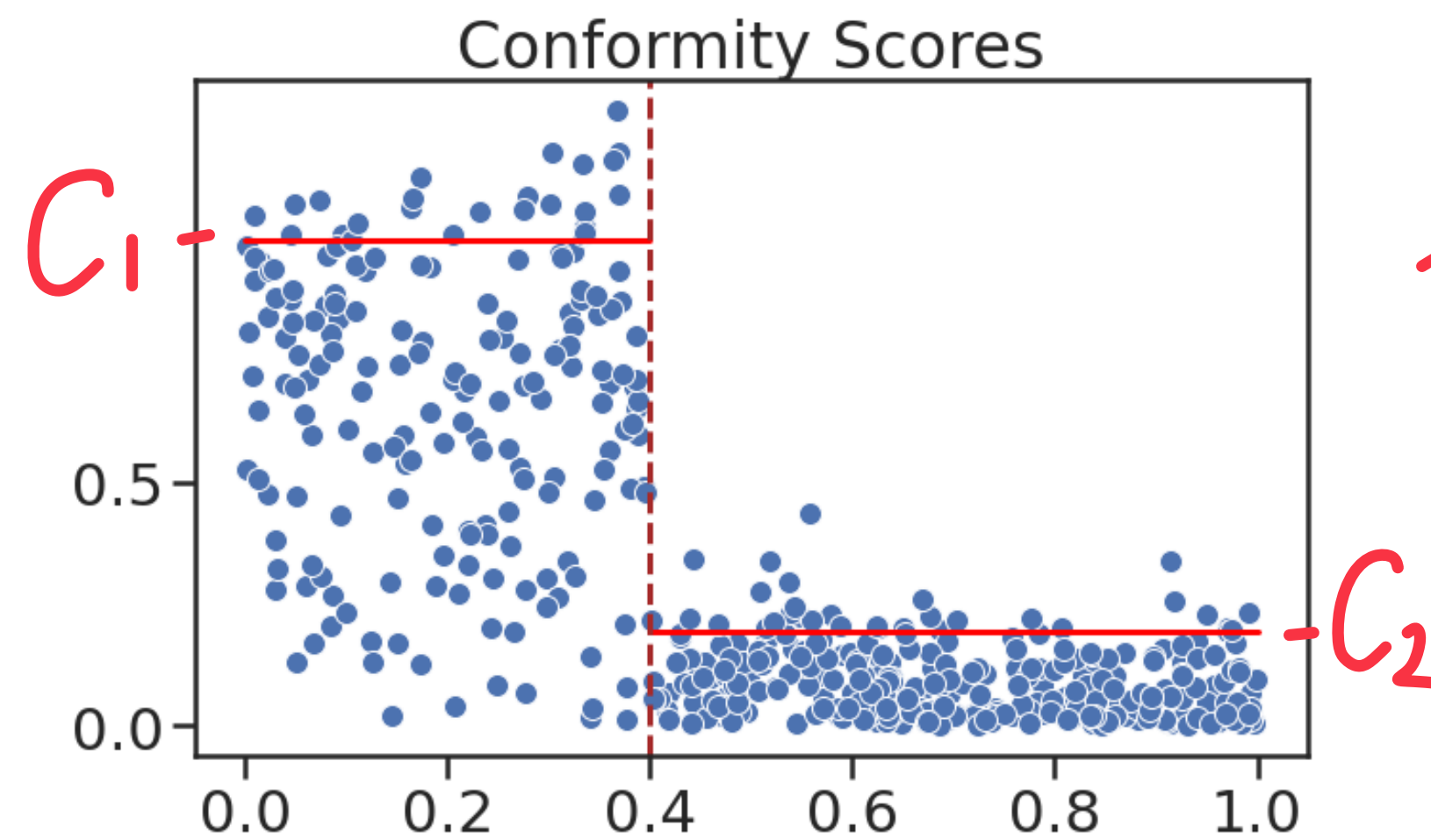
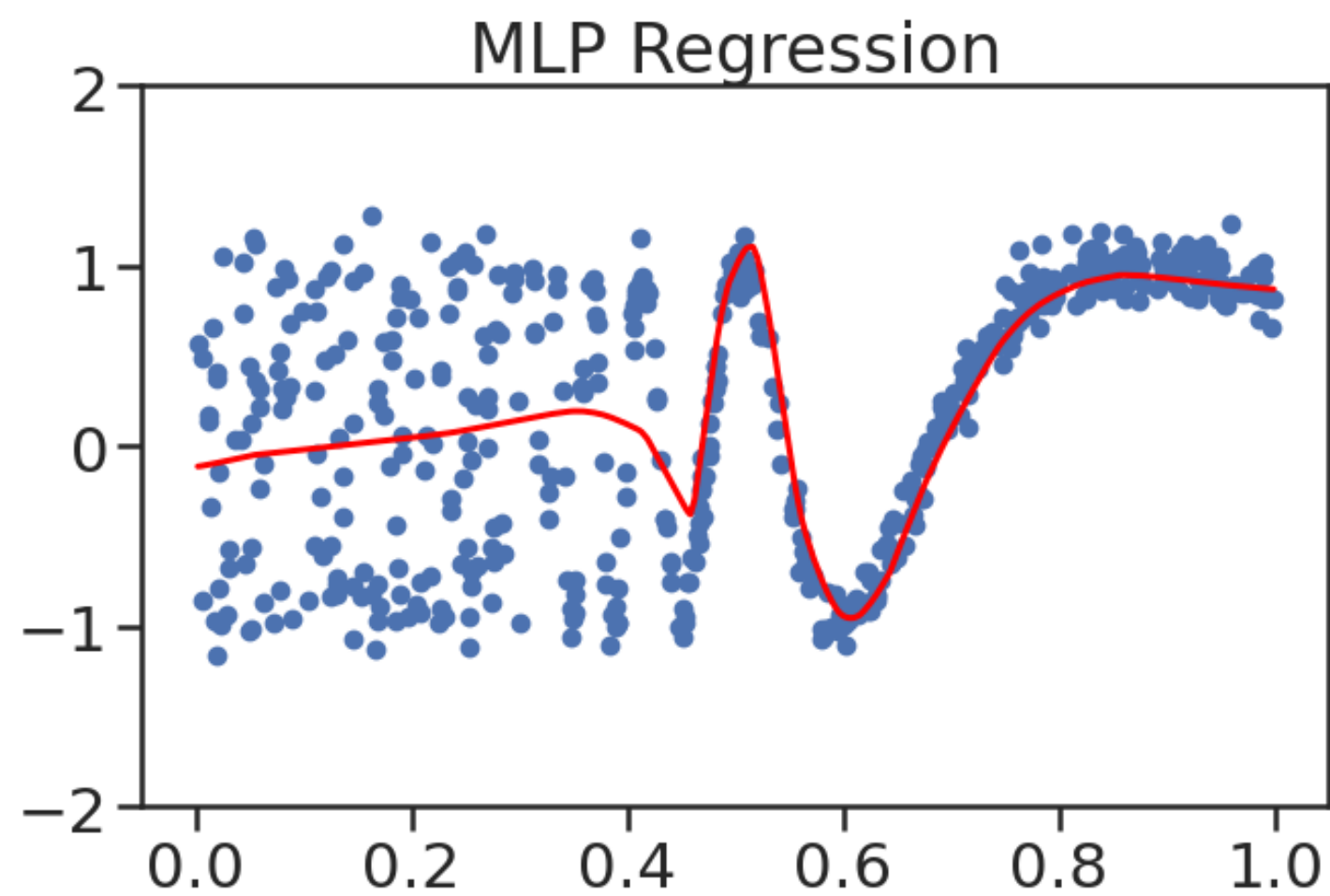
**Avoiding Split of Precious Calibration Datasets (Adaptivity)**



Kim\*, O'hagan\*, and Rockova (2025)

# Tree Regression is useful for Adaptive UQ

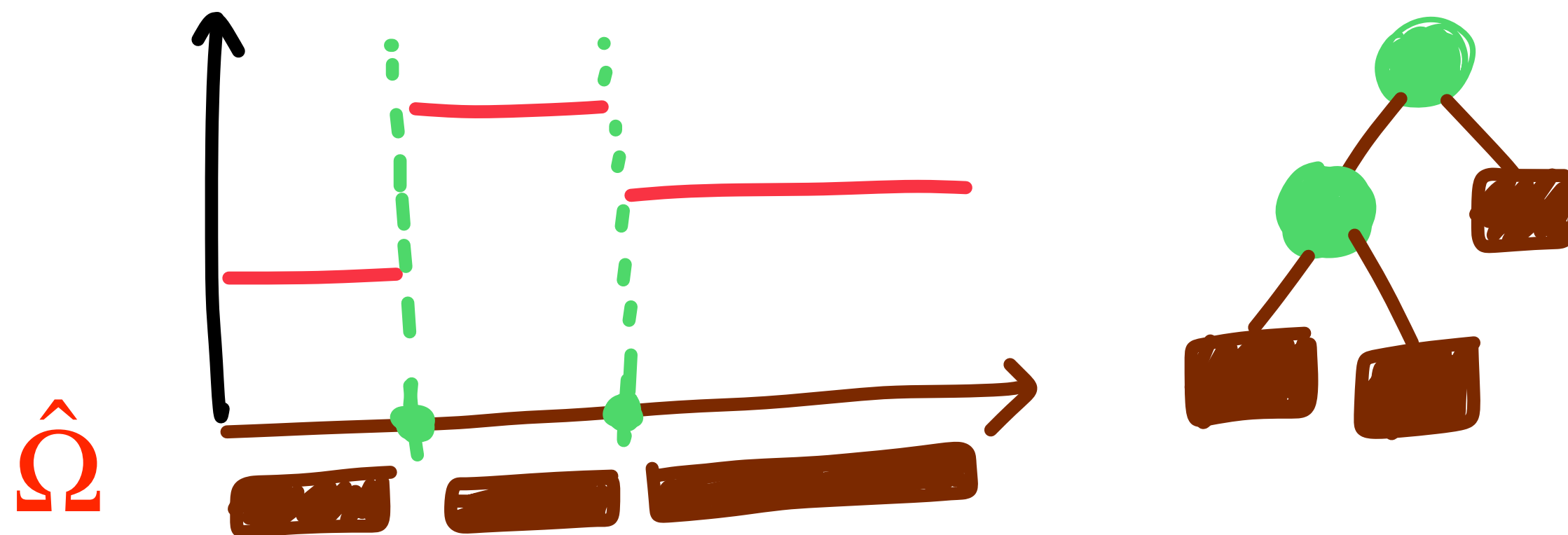
Idea: Automatic Grouping through Tree Regression 



Step 1) We get a partition  $\hat{\Omega}(\mathcal{D}_{\text{calib}})$  by applying a tree regression on conformity scores of  $\mathcal{D}_{\text{calib}}$

Step 2) Then we apply conformal prediction on each bin

WHY Tree? Tree regression = partitioning tool for capturing shape



# Challenge of exchangeability

Conditional Coverage Guarantee?:

For  $C = \lceil (1 - \alpha)(m + 1) \rceil$ -th smallest value of  $\{S_i\}_{i: X_i \in G}$

$$1 - \alpha \stackrel{??}{\leq} \mathbb{P} \left( S_{\text{test}} \leq C \mid X_{\text{test}} \in G \right) \stackrel{??}{\leq} 1 - \alpha + \frac{1}{m+1}, \quad m = \min_G |G|, \quad G \in \hat{\Omega}(\mathcal{D}_{\text{cal}})$$

# Challenge of exchangeability

Conditional Coverage Guarantee **X**:

For  $C = \lceil (1 - \alpha)(m + 1) \rceil$ -th smallest value of  $\{S_i\}_{i: X_i \in G}$

$$1 - \alpha \not\leq \mathbb{P}(S_{\text{test}} \leq C | X_{\text{test}} \in G) \not\leq 1 - \alpha + \frac{1}{m+1}, \quad m = \min_G |G|, \quad G \in \hat{\Omega}(\mathcal{D}_{\text{cal}})$$

Our group depends on the calibration dataset-!!

# Challenge of exchangeability

For  $C = \lceil (1 - \alpha)(m + 1) \rceil$ -th smallest value of  $\{S_i\}_{i: X_i \in G}$

$$1 - \alpha \not\leq \mathbb{P}(S_{\text{test}} \leq C | X_{\text{test}} \in G) \not\leq 1 - \alpha + \frac{1}{m+1}, \quad m = \min_G |G|, \quad G \in \hat{\Omega}(\mathcal{D}_{\text{cal}})$$

**Challenge:** exchangeability broken down

$$X_1, \dots, X_n, X_{\text{test}} | \hat{\Omega}(\mathcal{D}_{\text{calib}}) \stackrel{d}{\neq} X_{\sigma(1)}, \dots, X_{\sigma(n)}, X_{\sigma(\text{test})} | \hat{\Omega}(\mathcal{D}_{\text{calib}}),$$

where  $\hat{\Omega}(\mathcal{D}_{\text{calib}})$  is the partition obtained on the calibration data

# Challenge of exchangeability

For  $C = \lceil (1 - \alpha)(m + 1) \rceil$ -th smallest value of  $\{S_i\}_{i: X_i \in G}$

$$1 - \alpha \not\leq \mathbb{P}(S_{\text{test}} \leq C | X_{\text{test}} \in G) \not\leq 1 - \alpha + \frac{1}{m+1}, \quad m = \min_G |G|, \quad G \in \hat{\Omega}(\mathcal{D}_{\text{cal}})$$

**Challenge:** exchangeability broken down

$$X_1, \dots, X_n, X_{\text{test}} | \hat{\Omega}(\mathcal{D}_{\text{calib}}) \stackrel{d}{\neq} X_{\sigma(1)}, \dots, X_{\sigma(n)}, X_{\sigma(\text{test})} | \hat{\Omega}(\mathcal{D}_{\text{calib}}),$$

where  $\hat{\Omega}(\mathcal{D}_{\text{calib}})$  is the partition obtained on the calibration data

$$X_1, \dots, X_n, X_{\text{test}} | \hat{\Omega}(\mathcal{D}_{\text{calib}+\text{test}}) \stackrel{d}{=} X_{\sigma(1)}, \dots, X_{\sigma(n)}, X_{\sigma(\text{test})} | \hat{\Omega}(\mathcal{D}_{\text{calib}+\text{test}}),$$

# Challenge of exchangeability

For  $C = \lceil (1 - \alpha)(m + 1) \rceil$ -th smallest value of  $\{S_i\}_{i: X_i \in G}$

$$1 - \alpha \leq \mathbb{P}(S_{\text{test}} \leq C | X_{\text{test}} \in G) \leq 1 - \alpha + \frac{1}{m+1}, \quad m = \min_G |G|, \quad G \in \hat{\Omega}(\mathcal{D}_{\text{cal}})$$

**Challenge:** exchangeability broken down

$$X_1, \dots, X_n, X_{\text{test}} | \hat{\Omega}(\mathcal{D}_{\text{calib}}) \stackrel{d}{\neq} X_{\sigma(1)}, \dots, X_{\sigma(n)}, X_{\sigma(\text{test})} | \hat{\Omega}(\mathcal{D}_{\text{calib}}),$$

where  $\hat{\Omega}(\mathcal{D}_{\text{calib}})$  is the partition obtained on the calibration data

$$X_1, \dots, X_n, X_{\text{test}} | \hat{\Omega}(\mathcal{D}_{\text{calib}+\text{test}}) \stackrel{d}{=} X_{\sigma(1)}, \dots, X_{\sigma(n)}, X_{\sigma(\text{test})} | \hat{\Omega}(\mathcal{D}_{\text{calib}+\text{test}}),$$

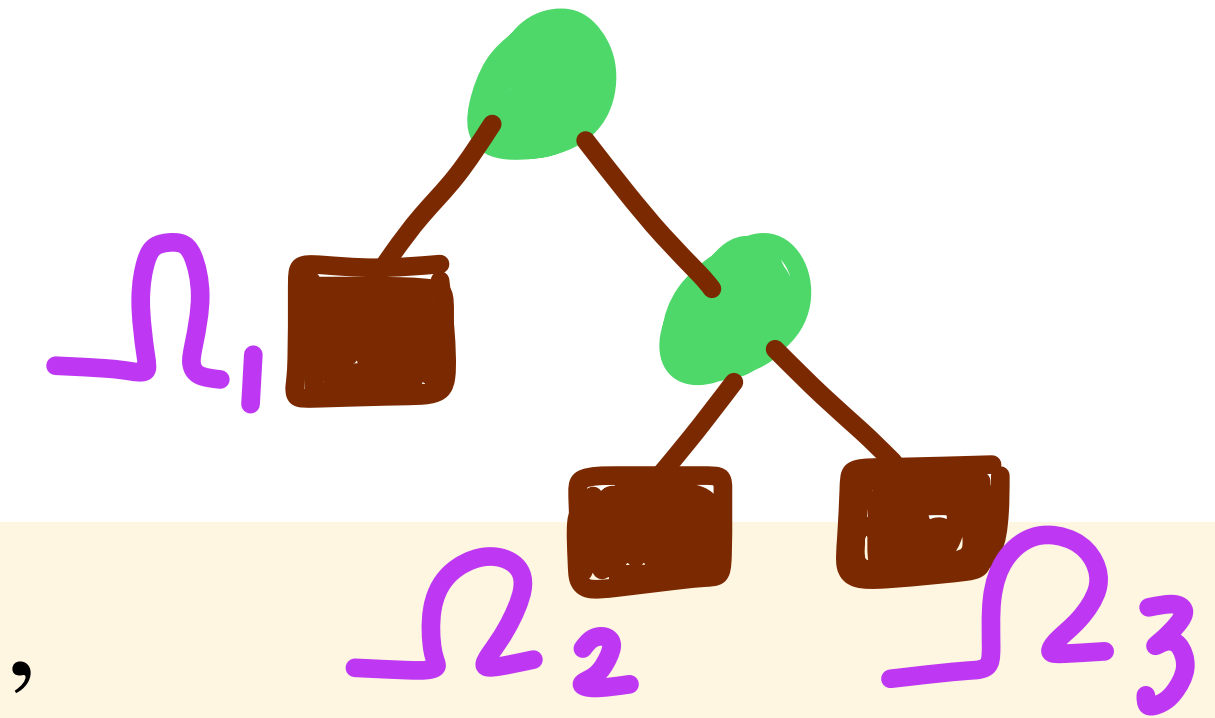
**Want: Unchangeability**  $\hat{\Omega}(\mathcal{D}_{\text{calib}}) = \hat{\Omega}(\mathcal{D}_{\text{calib}+\text{test}})$

(Unchangeable after adding or removing one observation with large probability),

# Unchangeable Tree

We consider dyadic trees

New split criterion: reduction of range



**Range:**  $R(\Omega_k) = \max\{S_i : X_i \in \Omega_k\} - \min\{S_i : X_i \in \Omega_k\}$ ,

**Reduction of Range:**  $RR(\Omega_k, j) = R(\Omega_k) - (R^{j, \text{left}}(\Omega_k) + R^{j, \text{right}}(\Omega_k))/2$

$j \in [1, \dots, d]$  : Split direction,  $\Omega_k$  : A leaf node (k-th bin),  $S_i$  : The conformity score of data point  $X_i$

Candidates:

Leaf nodes having more than  $m$  calibration points (and, all directions)

Split:

$$(j^*, \Omega_k^*) = \arg \max_{j, \Omega_k \in \text{Candidates}} RR(\Omega_k, j)$$

Stopping Criteria:

Candidates empty

# How unchangeability is obtained?

When removing  $(X_{\text{test}}, S_{\text{test}})$

**Unchangeability**

$$\hat{\Omega}(\mathcal{D}_{\text{calib}}) = \hat{\Omega}(\mathcal{D}_{\text{calib} + \text{test}})$$

**Splitting Candidacy**  
**Unchanged**

**Criterion Values**  
**Unchanged**

**Dyadic Tree**

**Range Reduction**

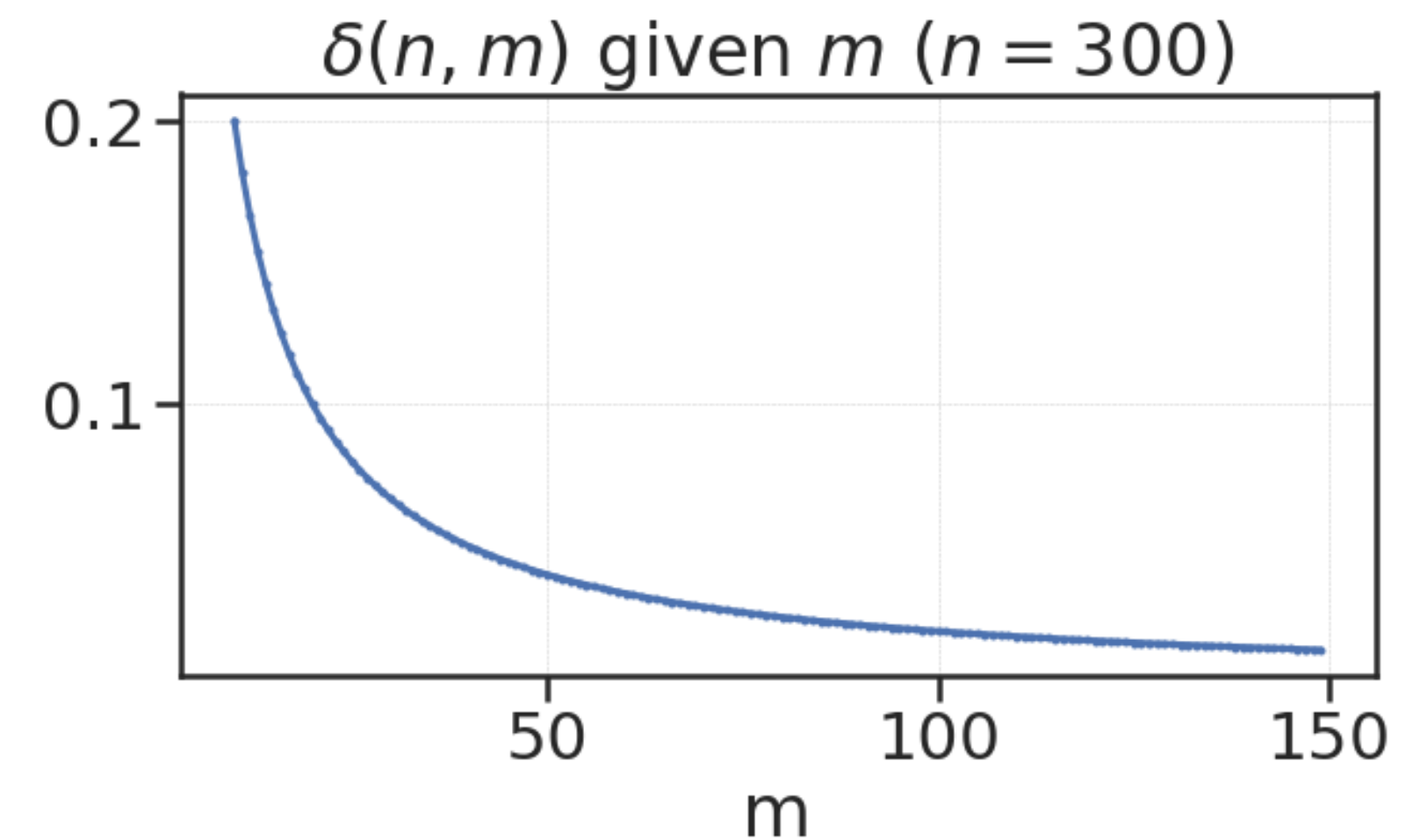
High Probability

# Unchangeability Theorem (Dyadic Tree)

If the data points in  $\mathcal{D}_{\text{calib}+\text{test}}$  are **i.i.d.**, when applying our robust tree algorithm,

$$\mathbb{P}\{\hat{\Omega}(\mathcal{D}_{\text{calib}}) = \hat{\Omega}(\mathcal{D}_{\text{calib}+\text{test}})\} \geq 1 - \delta(n, m),$$

where  $\delta(n, m) := \frac{2}{m} + e^{-(2m-1)\left(1 + \frac{1}{n}\right)}$



# Unchangeability Theorem (Dyadic Tree)

If the data points in  $\mathcal{D}_{\text{calib}+\text{test}}$  are **i.i.d.**, when applying our robust tree algorithm,

$$\mathbb{P}\{\hat{\Omega}(\mathcal{D}_{\text{calib}}) = \hat{\Omega}(\mathcal{D}_{\text{calib}+\text{test}})\} \geq 1 - \delta(n, m),$$

where  $\delta(n, m) := \frac{2}{m} + e^{-(2m-1)\left(1 + \frac{1}{n}\right)}$

Groupwise:

$$1 - \alpha - \delta(n, m) \leq \mathbb{P}\{S_{\text{test}} \leq C \mid X_{\text{test}} \in \hat{\Omega}_k, \hat{\Omega}(\mathcal{D}_{\text{calib}})\} \leq 1 - \alpha + \frac{1}{1 + m_k} + \delta(n, m_k)$$

for  $C = \lceil (1 - \alpha)(m_k - 2) \rceil + 1$ -th smallest value of  $\{S_i\}_{i: X_i \in \hat{\Omega}_k}$ , where

$$\hat{\Omega}_k \in \hat{\Omega}(\mathcal{D}_{\text{calib}}) \text{ and } m_k = \left| \{i \mid X_i \in \hat{\Omega}_k\}_{i=1}^n \right|$$

$\delta(n, m)$ : A small cost for the advantage of **reusing** the calibration data.

(Upper bounds: if the conformity scores are continuous)

# Unchangeability Theorem (Dyadic Tree)

If the data points in  $\mathcal{D}_{\text{calib}+\text{test}}$  are **i.i.d.**, when applying our robust tree algorithm,

$$\mathbb{P}\{\hat{\Omega}(\mathcal{D}_{\text{calib}}) = \hat{\Omega}(\mathcal{D}_{\text{calib}+\text{test}})\} \geq 1 - \delta(n, m),$$

where  $\delta(n, m) := \frac{2}{m} + e^{-(2m-1)\left(1 + \frac{1}{n}\right)}$

Marginal:

$$1 - \alpha - \delta(n, m) \leq \mathbb{P}\{S_{\text{test}} \leq C\} \leq 1 - \alpha + \frac{1}{1 + m} + \delta(n, m)$$

for  $C = \lceil (1 - \alpha)(m_k - 2) \rceil + 1$ -th smallest value of  $\{S_i\}_{i: X_i \in \hat{\Omega}_k}$ , where

$$\hat{\Omega}_k \in \hat{\Omega}(\mathcal{D}_{\text{calib}}) \text{ and } m_k = \left| \{i \mid X_i \in \hat{\Omega}_k\}_{i=1}^n \right|$$

$\delta(n, m)$ : A small cost for the advantage of **reusing** the calibration data.

(Upper bounds: if the conformity scores are continuous)

# Our method gives tighter prediction set

**Data Source:** Dermatology (1997) . UCI Machine Learning Repository

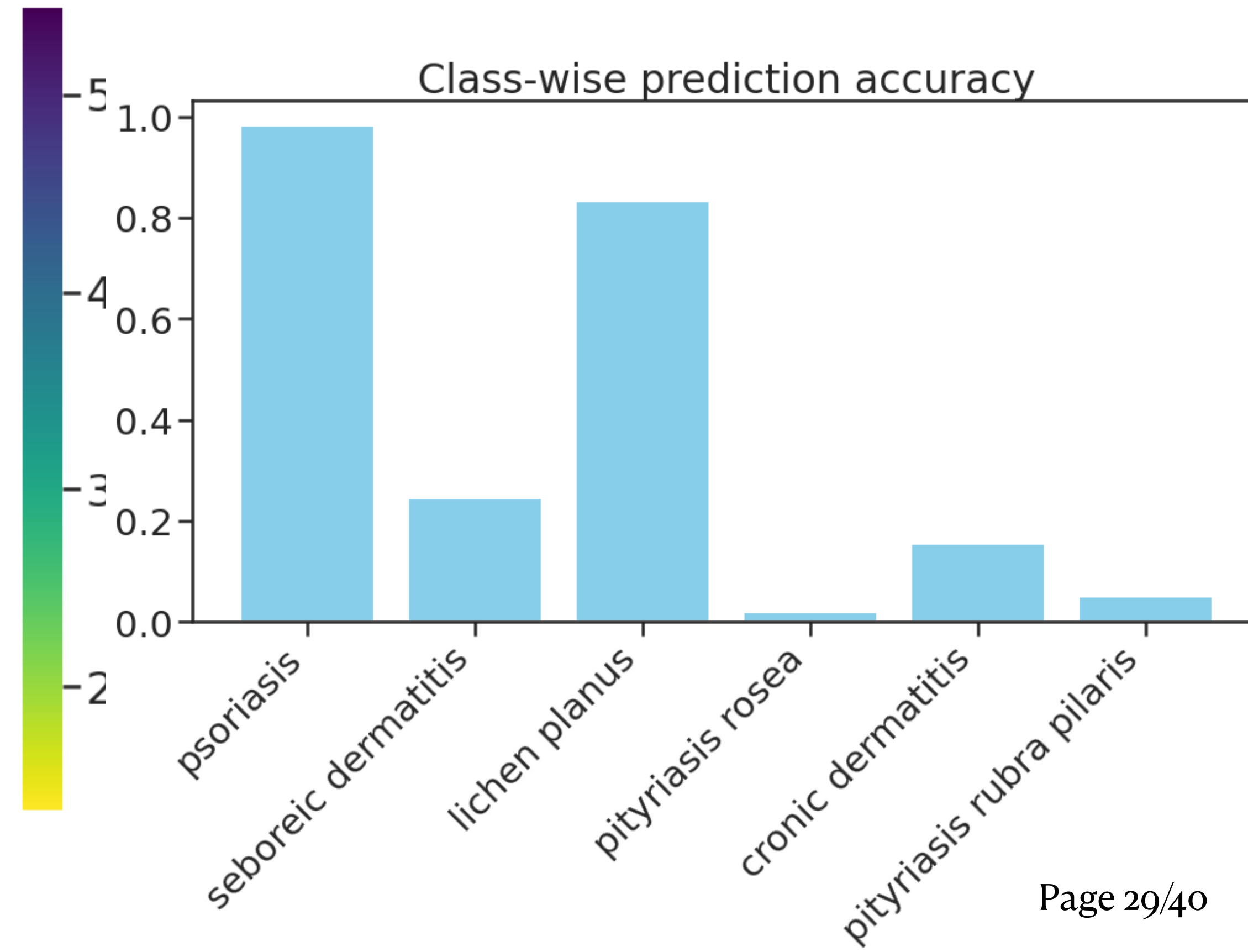
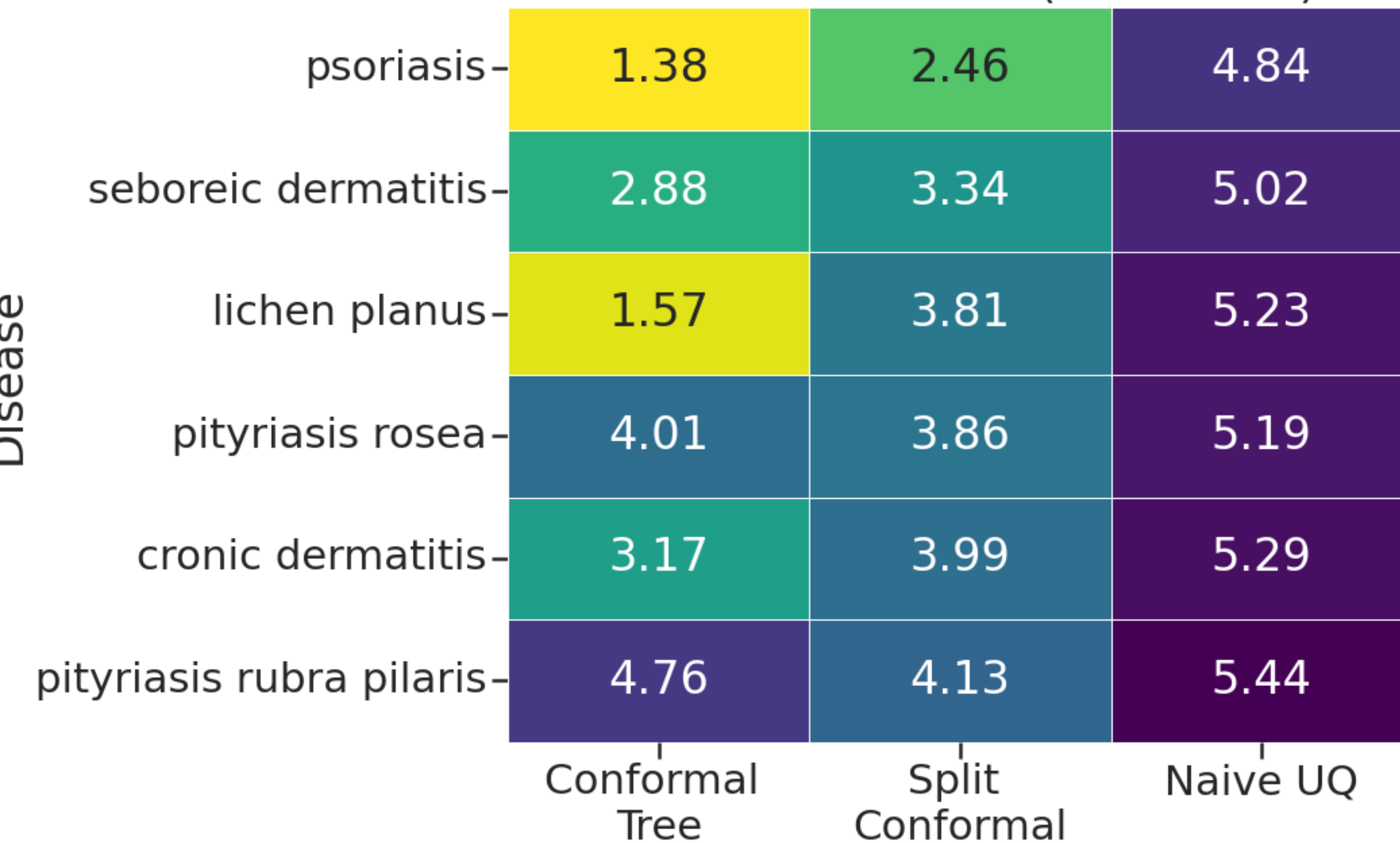
**Covariate: 12 features    Diagnosis: 6 categories**

**Calibration Set: n = 219    Test Set: n = 147**

**Conformal Tree tuning parameter: m=50**

**Multiple Splits (30)**

Prediction Set size ( $1 - \alpha = 0.9$ )



# Our method gives tighter prediction set

**Data Source:** Dermatology (1997) . UCI Machine Learning Repository

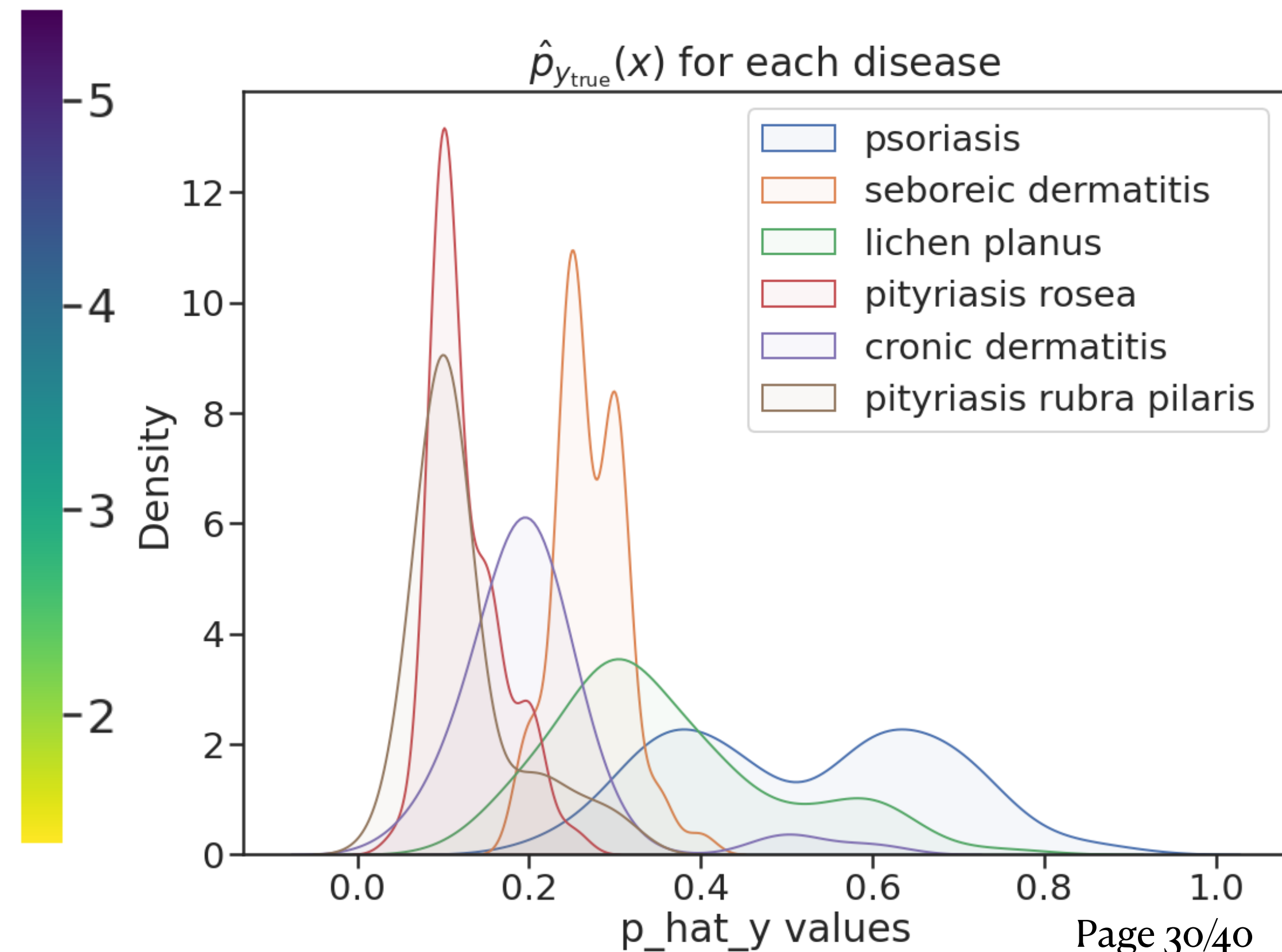
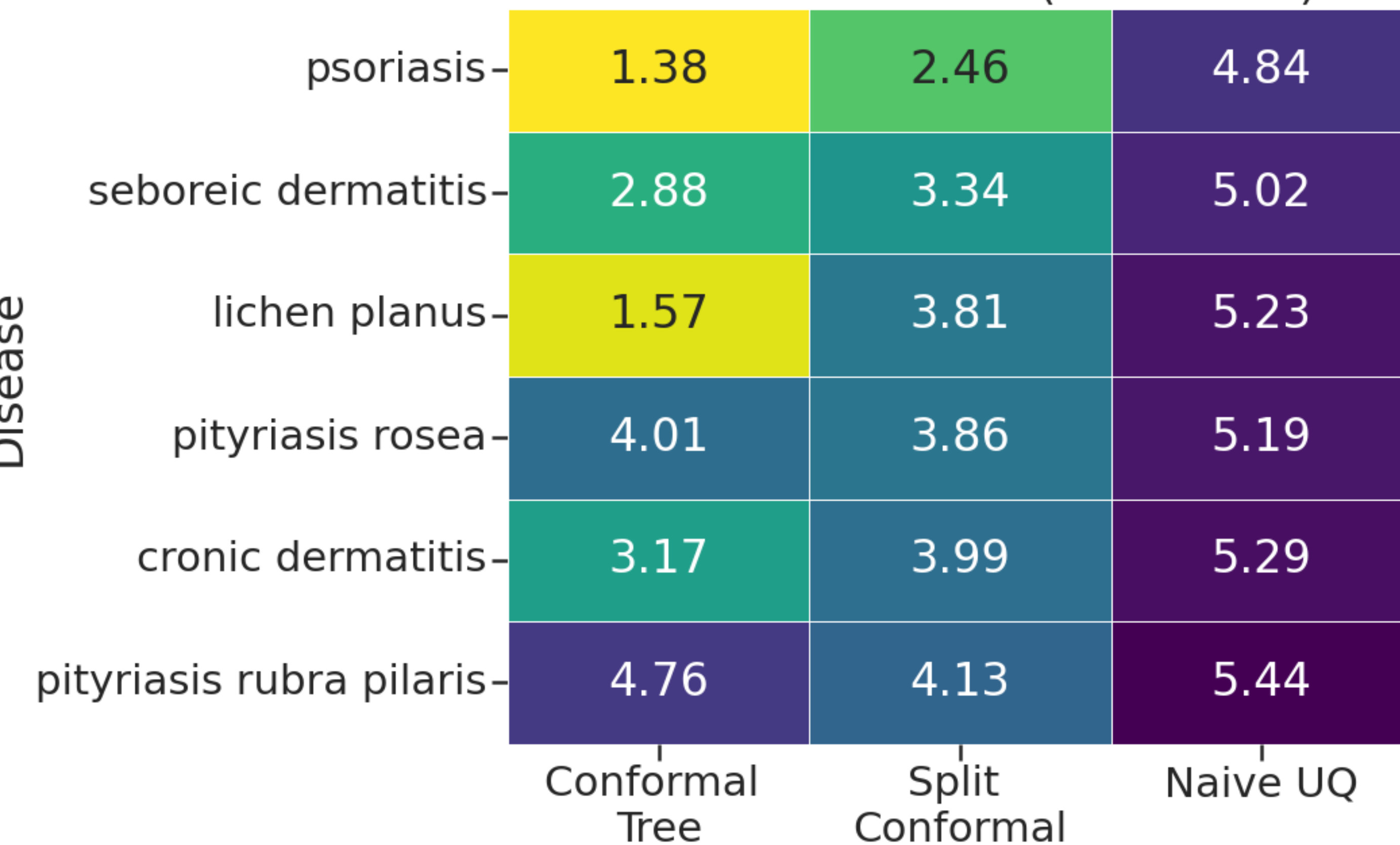
**Covariate: 12 features    Diagnosis: 6 categories**

**Calibration Set: n = 219    Test Set: n = 147**

**Conformal Tree tuning parameter: m=50**

**Multiple Splits (30)**

Prediction Set size ( $1 - \alpha = 0.9$ )



# Our method gives tighter prediction set

**Data Source:** Dermatology (1997) . UCI Machine Learning Repository

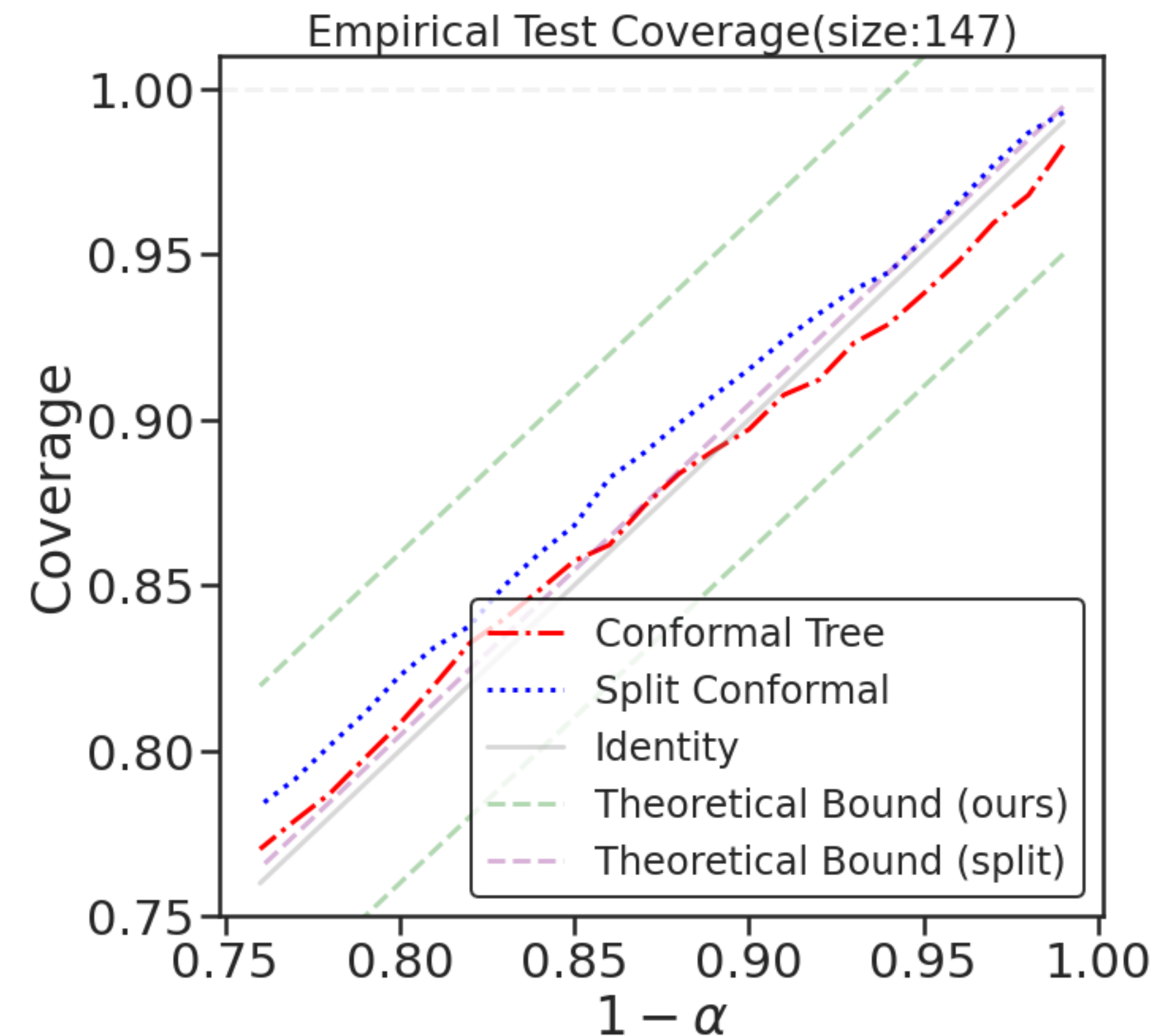
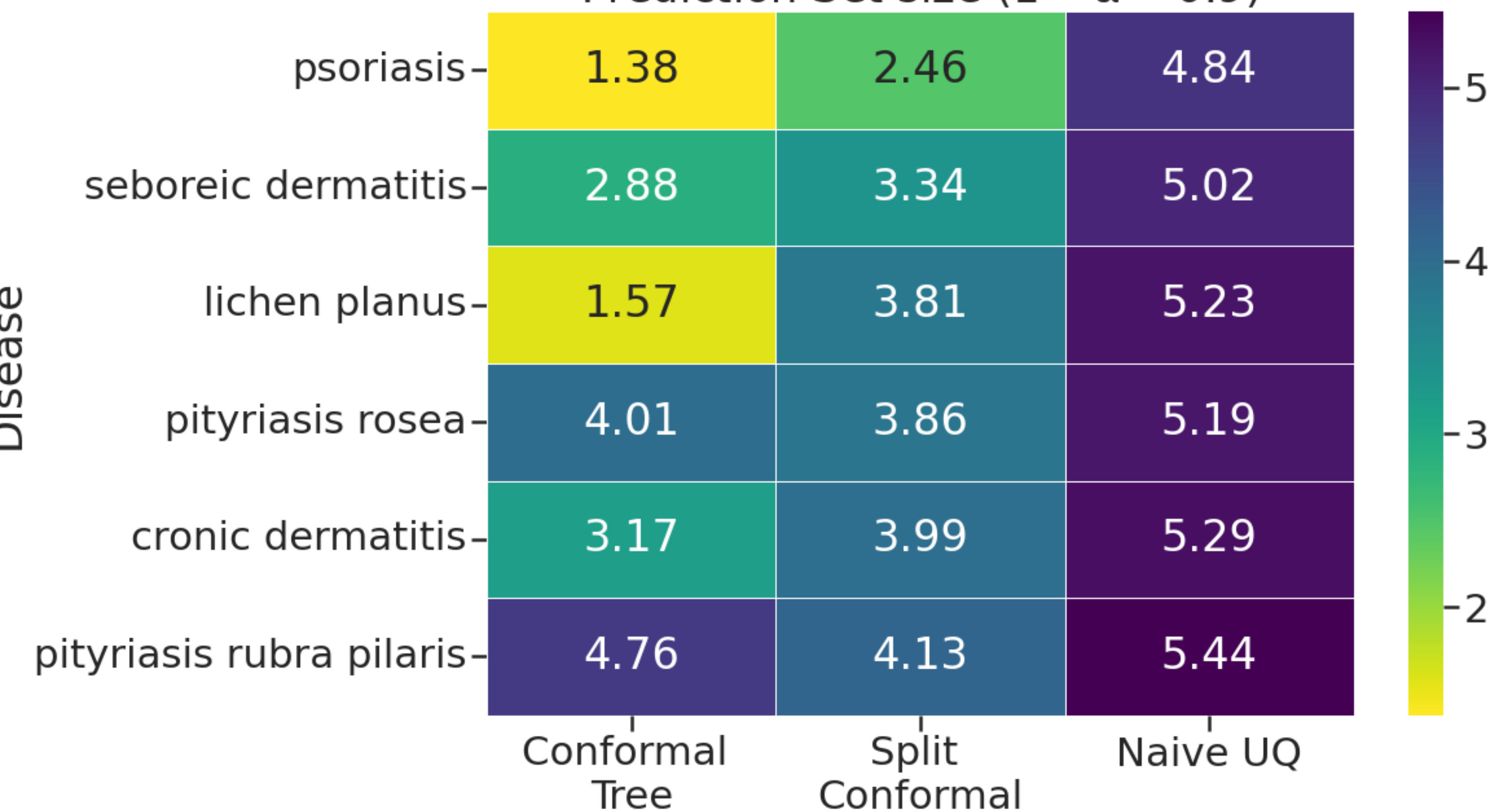
**Covariate: 12 features    Diagnosis: 6 categories**

**Calibration Set: n = 219    Test Set: n = 147**

**Conformal Tree tuning parameter: m=50**

**Multiple Splits (30)**

Prediction Set size ( $1 - \alpha = 0.9$ )



# Our method gives tighter prediction set

**Data Source:** Dermatology (1997) . UCI Machine Learning Repository

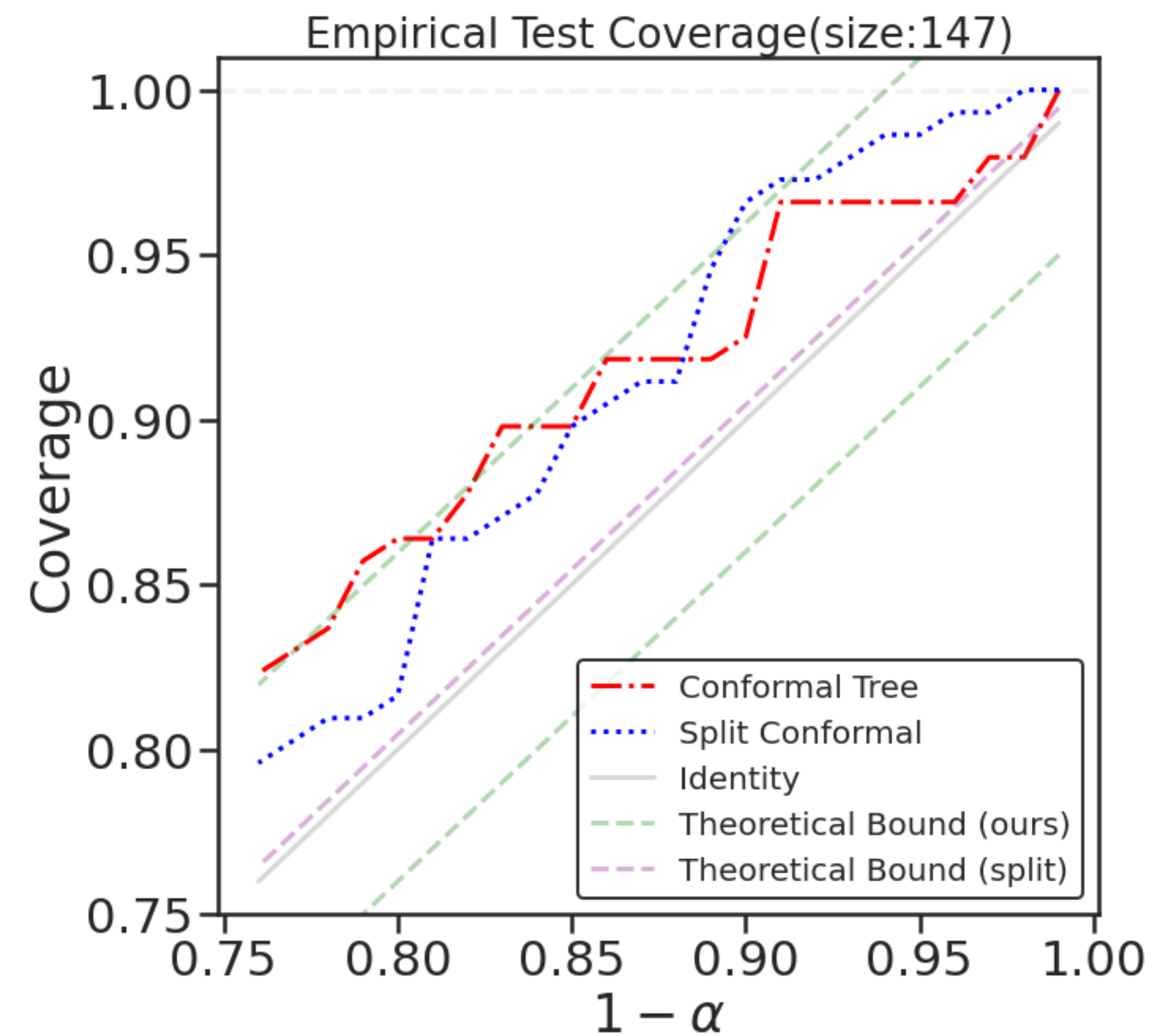
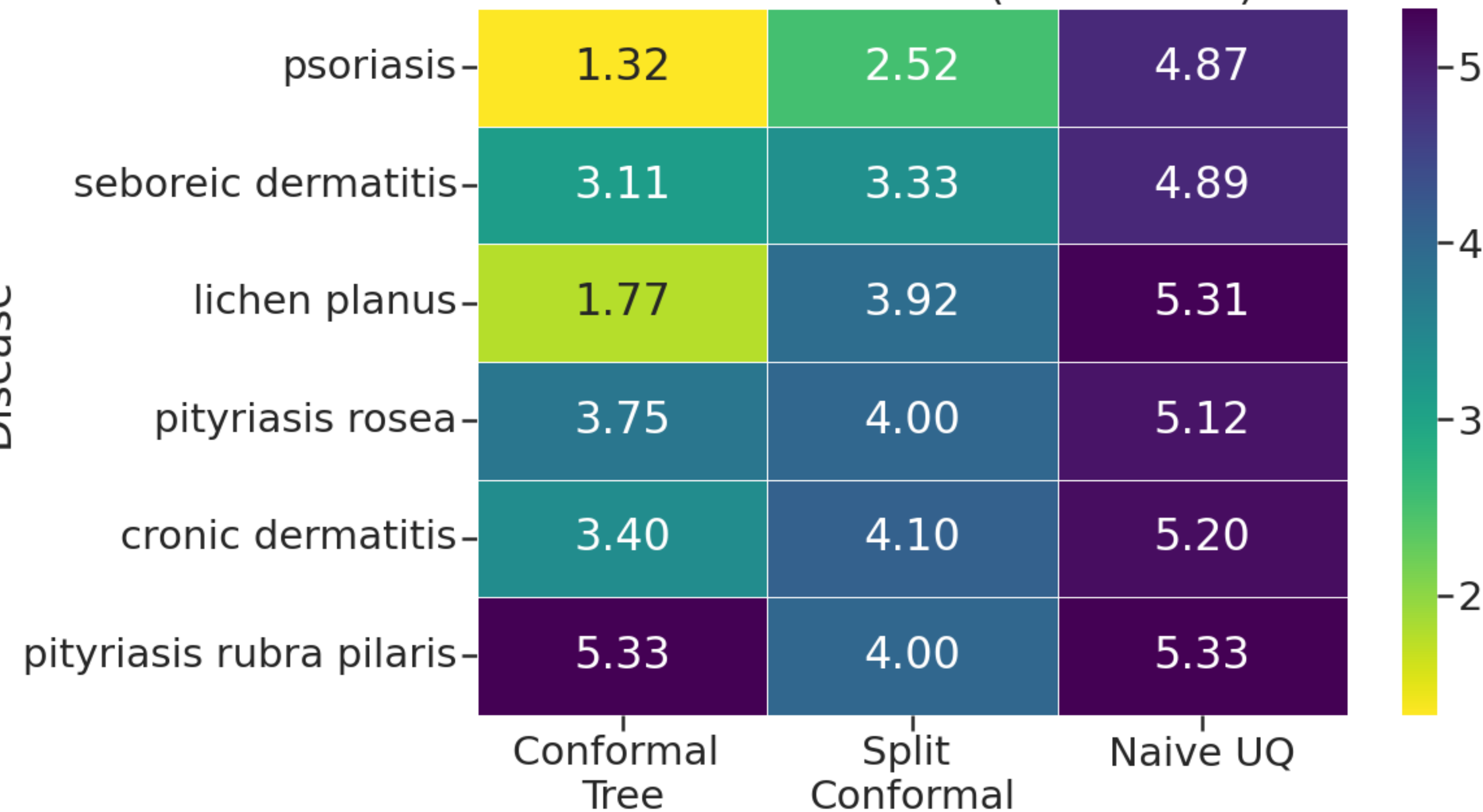
**Covariate: 12 features    Diagnosis: 6 categories**

**Calibration Set: n = 219    Test Set: n = 147**

**Conformal Tree tuning parameter: m=50**

## Single Split

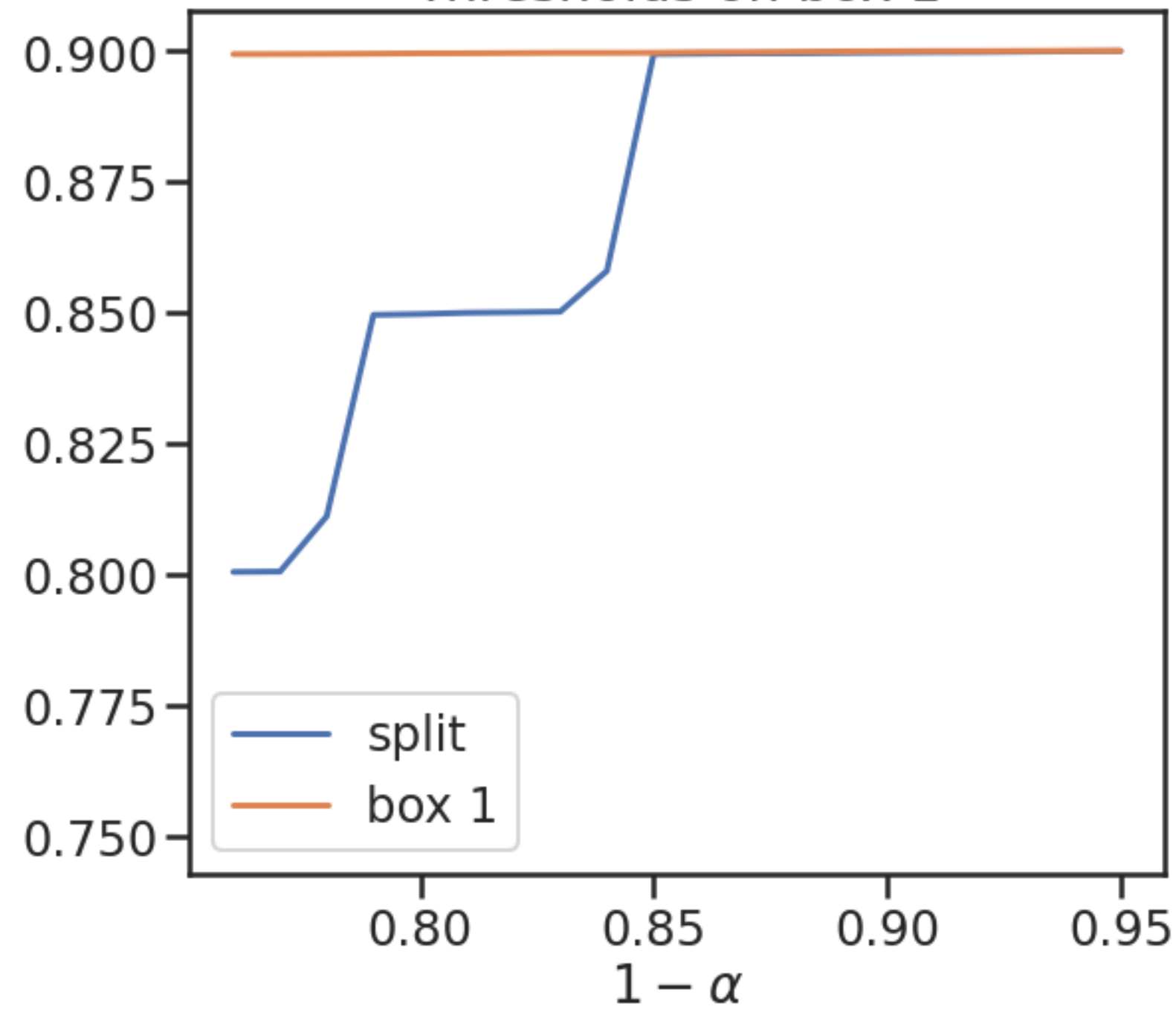
Prediction Set size ( $1 - \alpha = 0.9$ )



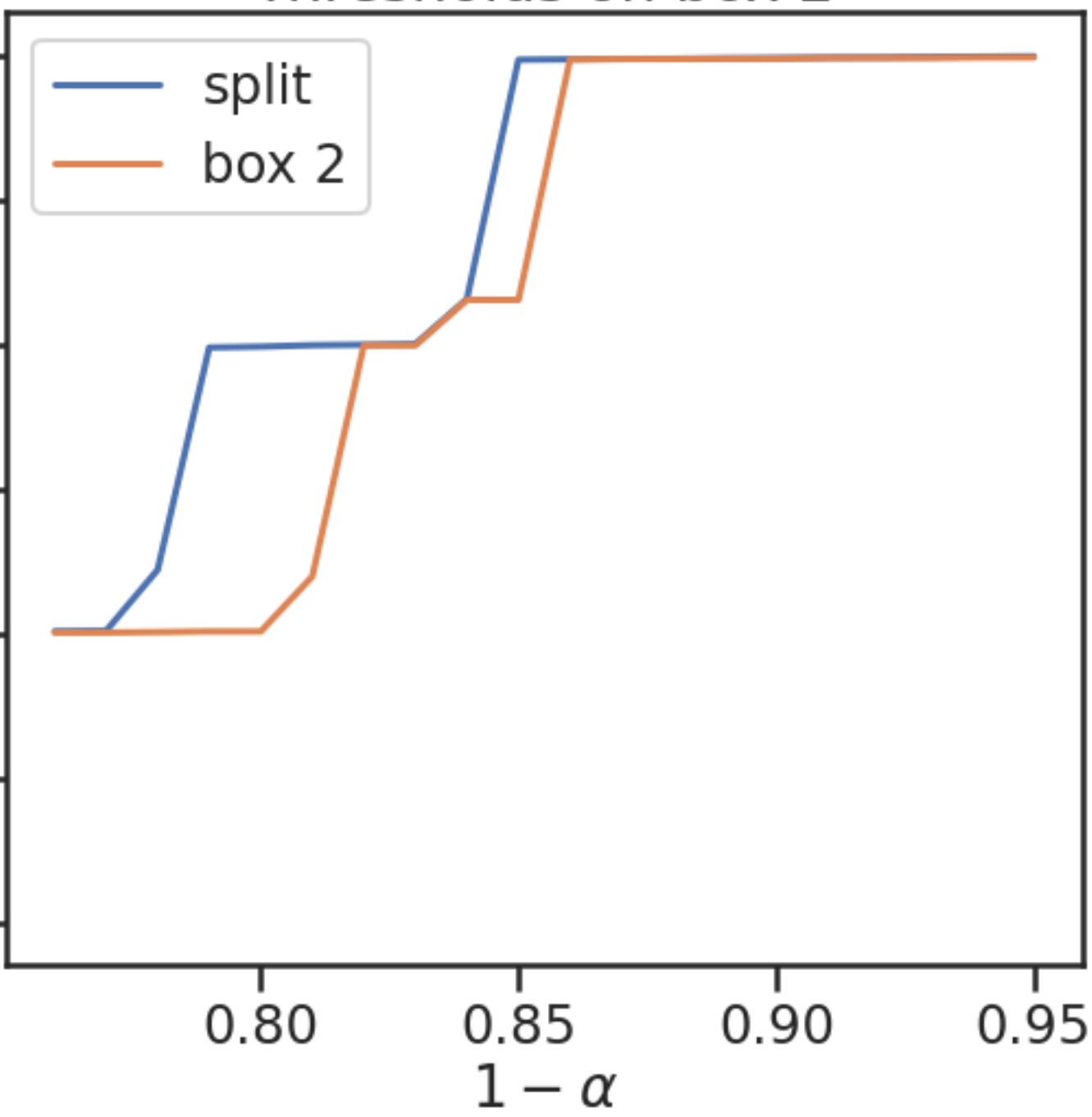
# Block-wise Coverage

Single Split

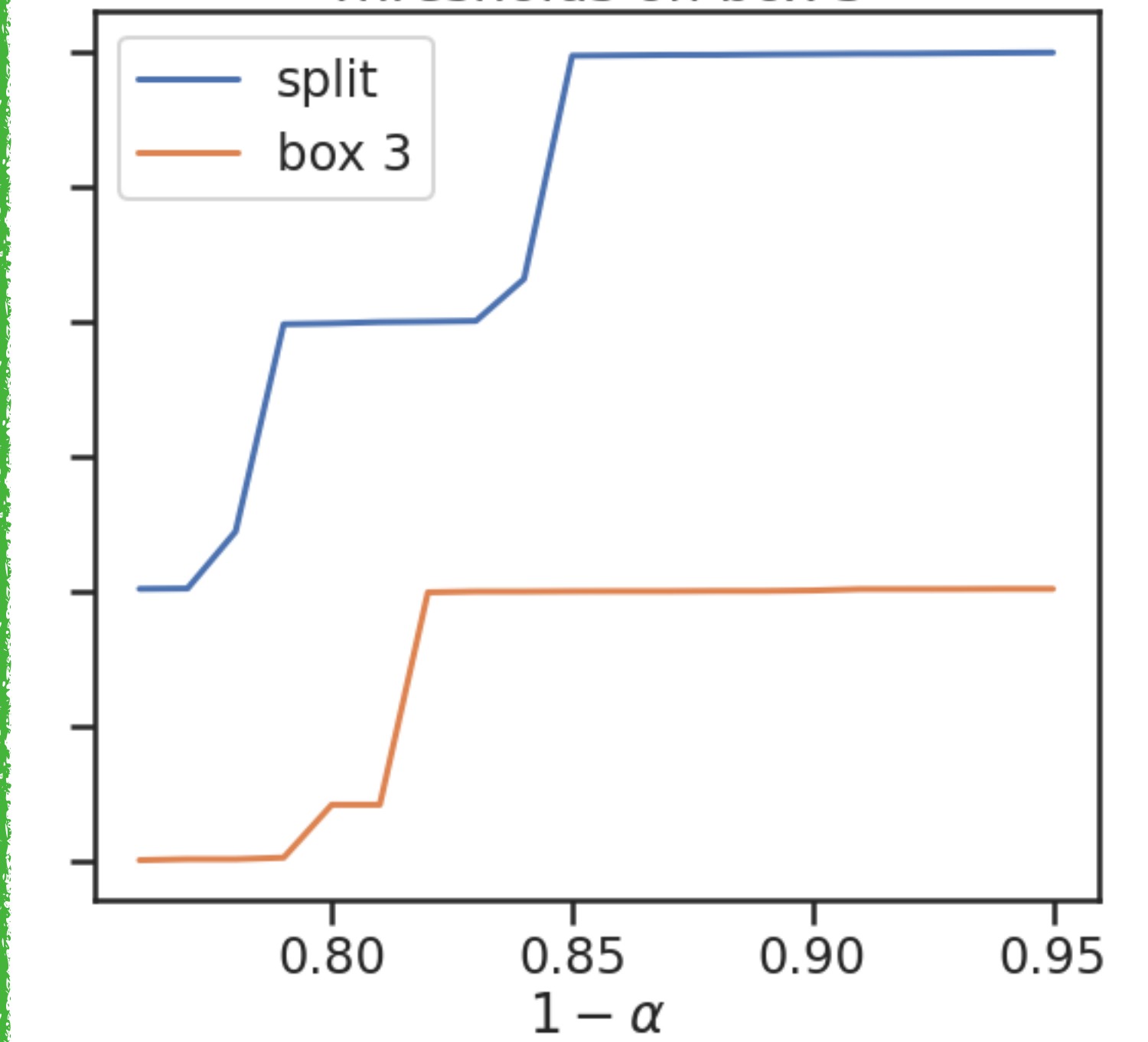
Thresholds on box 1



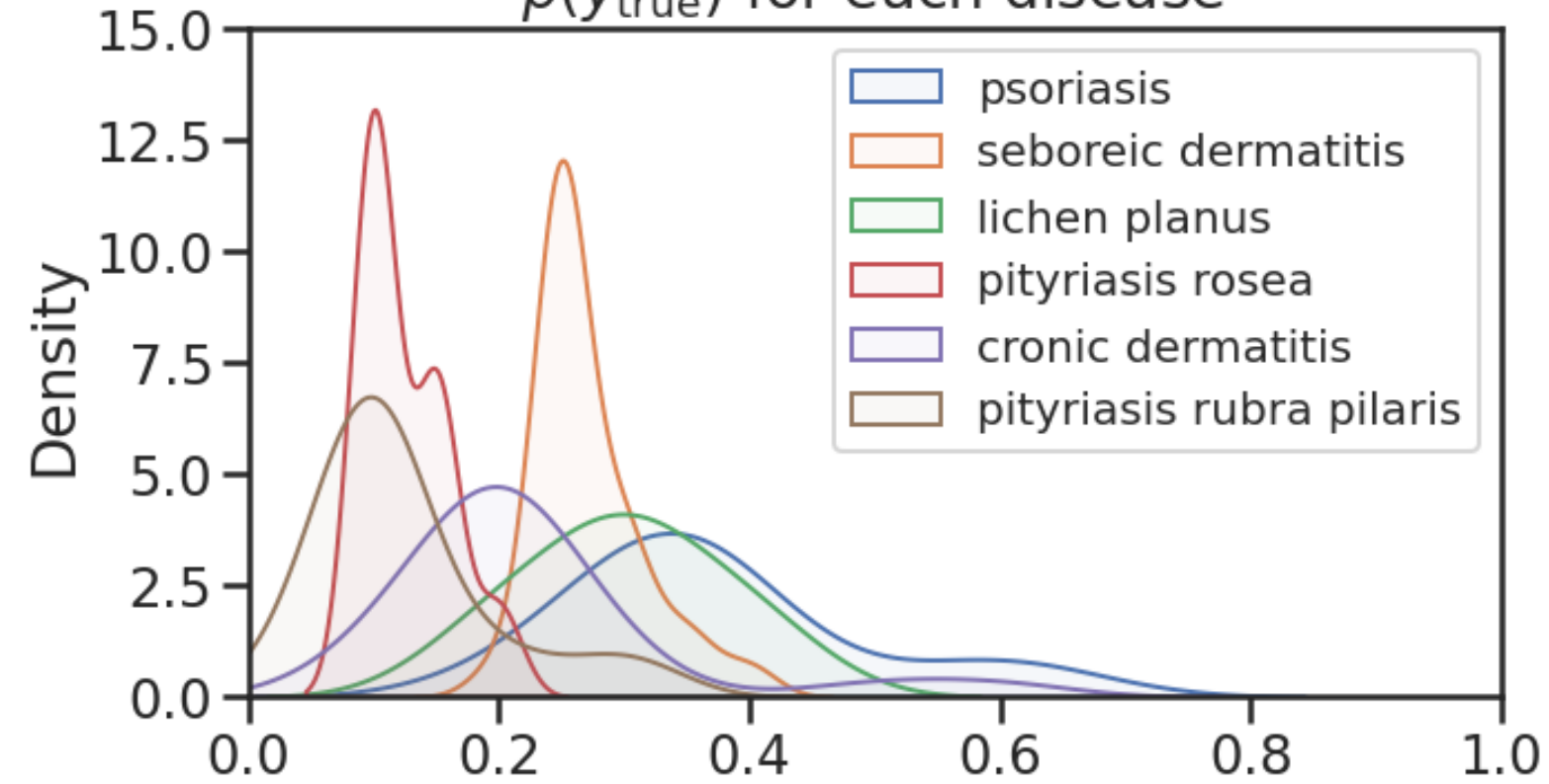
Thresholds on box 2



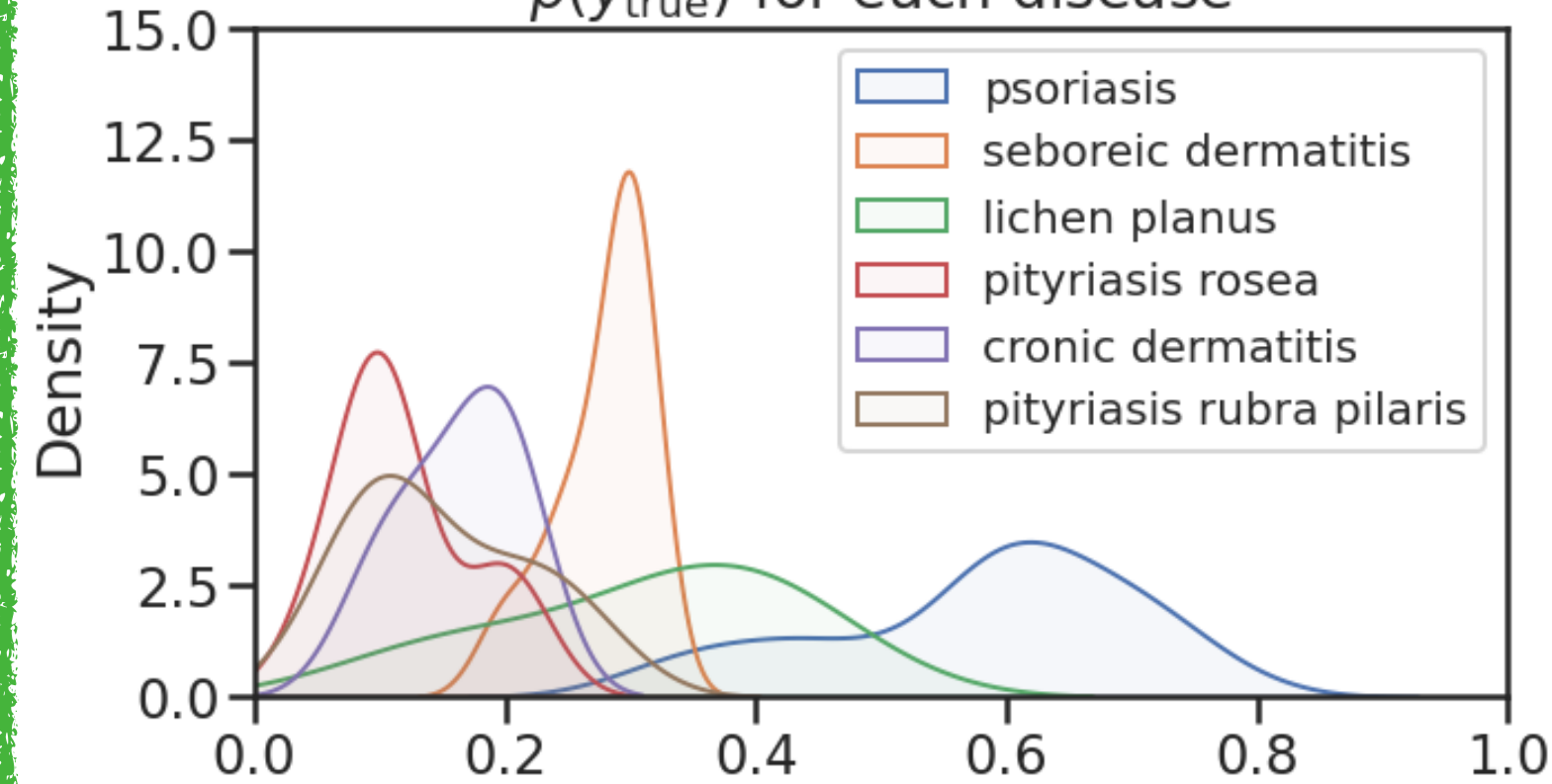
Thresholds on box 3



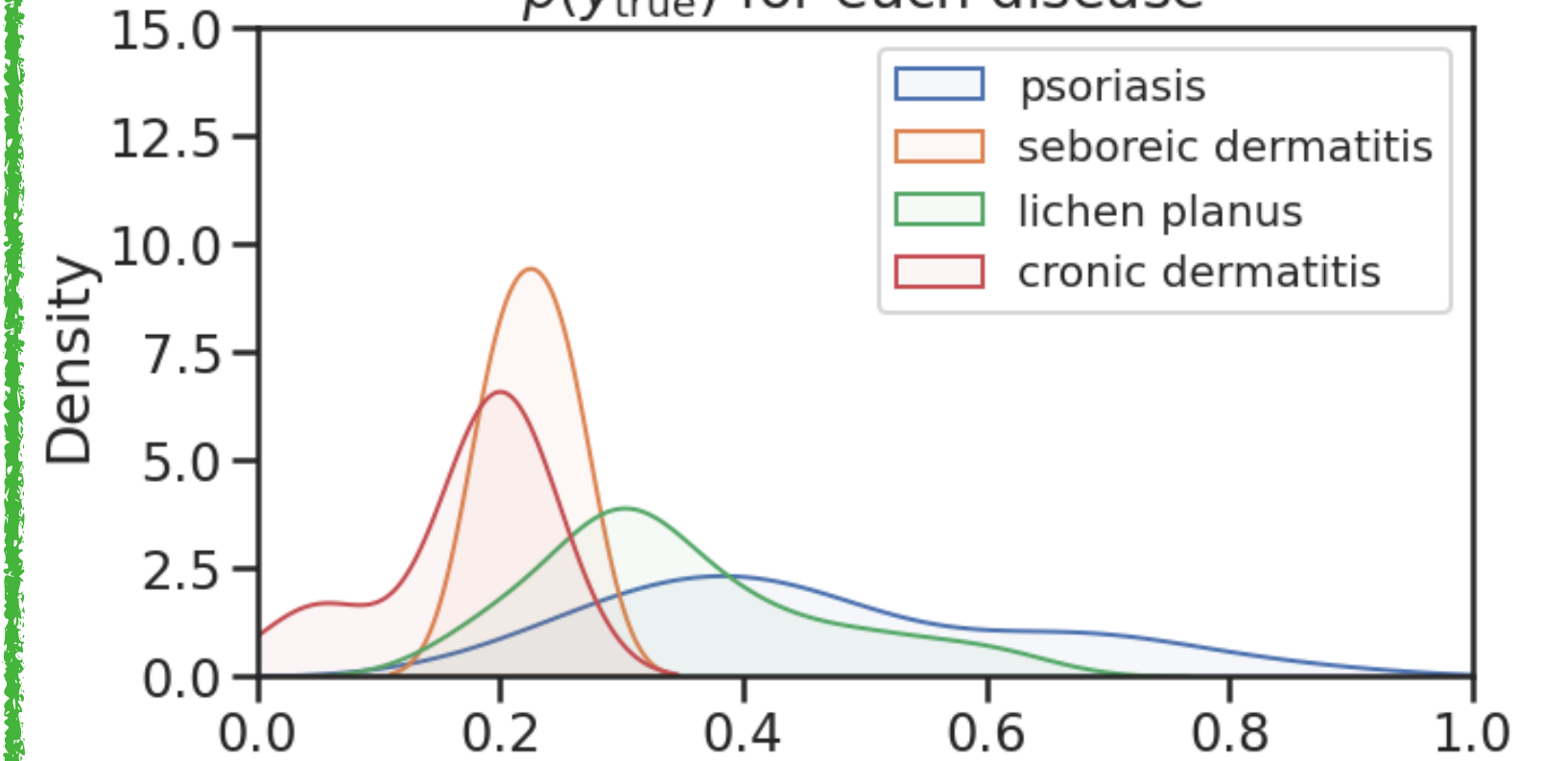
$\hat{p}(y_{\text{true}})$  for each disease



$\hat{p}(y_{\text{true}})$  for each disease



$\hat{p}(y_{\text{true}})$  for each disease



# With Small Calibration Datasets ours is useful

Dataset	Split Conformal		Conformal Tree			CART			Locally Weighted (Forest)			CQR		
	Width	Cov.	Width	Cov.	P.B.	Width	Cov.	P.B.	Width	Cov.	P.B.	Width	Cov.	P.B.
Data 1	4.43	0.91	4.03	0.9	<b>0.66</b>	<b>3.91</b>	0.9	0.62	4.79	0.92	0.56	4.25	0.9	<b>0.66</b>
Data 2	2.83	0.91	3.75	0.92	0.72	3.5	0.91	<b>0.74</b>	4.69	0.91	0.47	3.09	0.9	0.58
bike	1.31	0.9	1.38	0.9	0.52	1.19	0.9	0.59	0.99	0.89	0.74	<b>0.64</b>	0.89	<b>0.91</b>
bio	1.66	0.9	1.64	0.9	0.59	1.6	0.9	0.51	1.28	0.9	<b>0.71</b>	<b>1.16</b>	0.9	0.68
community	1.91	0.9	1.97	0.91	0.46	1.68	0.89	<b>0.68</b>	1.57	0.9	<b>0.68</b>	<b>1.54</b>	0.9	<b>0.68</b>
concrete	0.63	0.89	0.68	0.9	0.57	0.65	0.9	0.59	0.63	0.9	0.59	<b>0.57</b>	0.9	<b>0.65</b>
homes	0.61	0.9	0.64	0.9	0.63	0.62	0.9	0.35	0.58	0.9	0.73	<b>0.56</b>	0.9	<b>0.75</b>
star	0.17	0.9	<b>0.17</b>	0.91	<b>0.59</b>	<b>0.17</b>	0.9	0.44	0.18	0.9	0.45	0.18	0.9	0.4

P.B.: Proportion better (portion of the prediction intervals smaller than that of split conformal prediction)

# With Small Calibration Datasets ours is useful

Dataset	Split Conformal		Conformal Tree			CART			Locally Weighted (Forest)			CQR		
	Width	Cov.	Width	Cov.	P.B.	Width	Cov.	P.B.	Width	Cov.	P.B.	Width	Cov.	P.B.
s1	4.24	0.9	4.11	0.9	<b>0.89</b>	<b>3.93</b>	0.91	0.61	4.68	0.91	0.52	4.13	0.9	0.58
s2	2.61	0.88	<b>2.56</b>	0.88	<b>1.0</b>	4.01	0.9	0.68	4.1	0.89	0.39	3.1	0.89	0.38
s3	5.42	0.89	<b>5.31</b>	0.89	<b>1.0</b>	5.37	0.92	0.48	6.51	0.92	0.51	5.74	0.91	0.5
bike	2.05	0.9	1.98	0.89	<b>0.73</b>	2.13	0.91	0.51	1.96	0.91	0.59	<b>1.79</b>	0.89	0.61
bio	2.27	0.89	2.24	0.89	<b>1.0</b>	2.5	0.91	0.29	2.31	0.89	0.49	<b>1.93</b>	0.89	0.67
community	2.01	0.9	1.86	0.88	0.67	2.26	0.94	0.48	1.97	0.94	0.59	<b>1.75</b>	0.91	<b>0.7</b>
concrete	0.78	0.89	<b>0.75</b>	0.88	<b>0.95</b>	0.83	0.91	0.35	0.76	0.89	0.57	0.89	0.9	0.37
homes	0.94	0.88	<b>0.92</b>	0.88	<b>0.96</b>	1.21	0.9	0.52	0.95	0.9	0.65	0.94	0.9	0.58
star	0.18	0.91	<b>0.17</b>	0.9	<b>0.9</b>	0.2	0.94	0.22	0.2	0.91	0.33	0.2	0.91	0.27

P.B.: Proportion better (portion of the prediction intervals smaller than that of split conformal prediction)

# With Small Calibration Datasets ours is useful

Dataset	Split Conformal		Conformal Tree			CART			Locally Weighted (Forest)			CQR		
	Width	Cov.	Width	Cov.	P.B.	Width	Cov.	P.B.	Width	Cov.	P.B.	Width	Cov.	P.B.
s1	4.24	0.9	4.11	0.9	<b>0.89</b>	<b>3.93</b>	0.91	0.61	4.68	0.91	0.52	4.13	0.9	0.58
s2	2.61	0.88	<b>2.56</b>	0.88	<b>1.0</b>	4.01	0.9	0.68	4.1	0.89	0.39	3.1	0.89	0.38
s3	5.42	0.89	<b>5.31</b>	0.89	<b>1.0</b>	5.37	0.92	0.48	6.51	0.92	0.51	5.74	0.91	0.5
bike	2.05	0.9	1.98	0.89	<b>0.73</b>	2.13	0.91	0.51	1.96	0.91	0.59	<b>1.79</b>	0.89	0.61
bio	2.27	0.89	2.24	0.89	<b>1.0</b>	2.5	0.91	0.29	2.31	0.89	0.49	<b>1.93</b>	0.89	0.67
community	2.01	0.9	1.86	0.88	0.67	2.26	0.94	0.48	1.97	0.94	0.59	<b>1.75</b>	0.91	<b>0.7</b>
concrete	0.78	0.89	<b>0.75</b>	0.88	<b>0.95</b>	0.83	0.91	0.35	0.76	0.89	0.57	0.89	0.9	0.37
homes	0.94	0.88	<b>0.92</b>	0.88	<b>0.96</b>	1.21	0.9	0.52	0.95	0.9	0.65	0.94	0.9	0.58
star	0.18	0.91	<b>0.17</b>	0.9	<b>0.9</b>	0.2	0.94	0.22	0.2	0.91	0.33	0.2	0.91	0.27

Test Dataset

Calibration Dataset



Test Dataset

Calibration Dataset

Adaptivity Train



# Contribution in Conformal Tree

- 😊 First adaptive conformal prediction method, without access to the training dataset and without distributional assumptions
- 😊 First unchangeability for a tree partitioning designed for self-grouping
- 😊 Provide a protective layer for safe use of AI on predictions/classifications to properly communicate uncertainty in practice
- 😊 Very simple for practitioners to implement (single hyperparameter  $m$ )
- 😊 Later, other works followed, e.g. Posterior Conformal Prediction (Zhang and Candès, 2024)  
Our tree is “one time fitting” (not for all possible imputed  $y \in \mathcal{Y}$  values)

**Thank you-!**



# Block-wise Coverage

Single Split

