
MAX PLANCK INSTITUTE
FOR SOFTWARE SYSTEMS

Towards Human-AI Complementarity with Prediction Sets

Giovanni De Toni[1], Nastaran Okati[2], Suhas Thejaswi[2],
Eleni Straitouri[2], Manuel Gomez-Rodriguez[2]

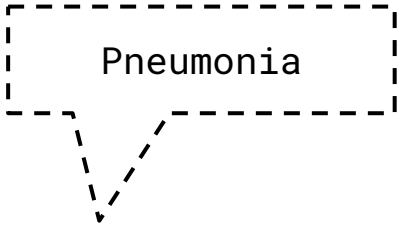
[1]Fondazione Bruno Kessler & University of Trento
[2]Max Planck Institute for Software Systems

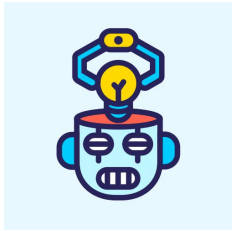
Decision Support System for multiclass classification

$$\mathcal{Y} = \left\{ \begin{array}{c} \text{Pneumothorax} \\ \text{Lung Cancer} \\ \text{Sarcoidosis} \\ \text{Pleural Effusion} \\ \dots \end{array} \right\}$$

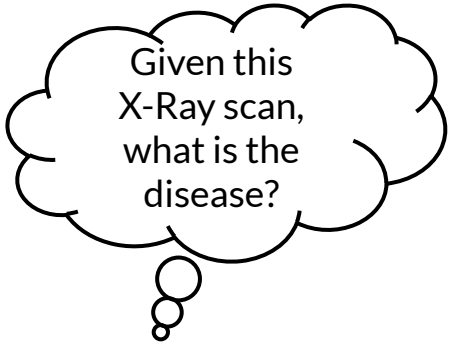
Data Sample

$x \in \mathcal{X}$





Decision Support System

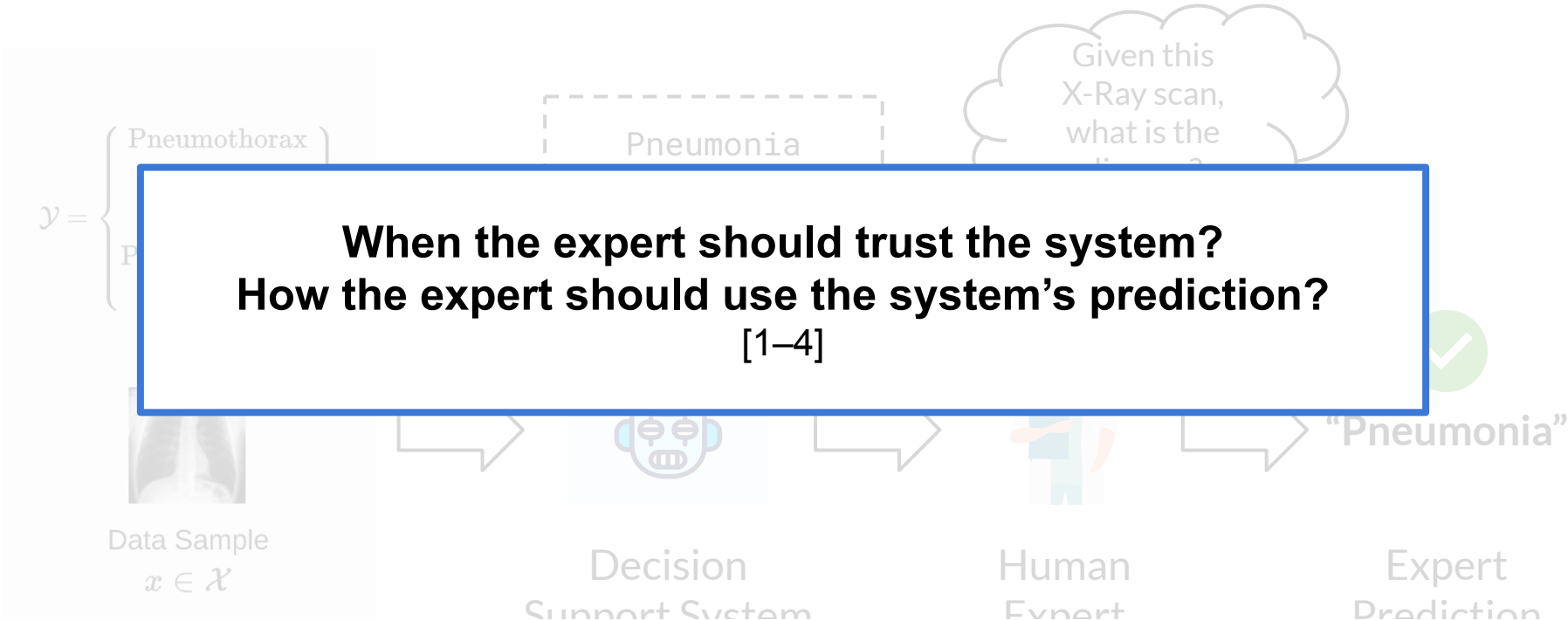


Human Expert

“Pneumonia”

Expert Prediction

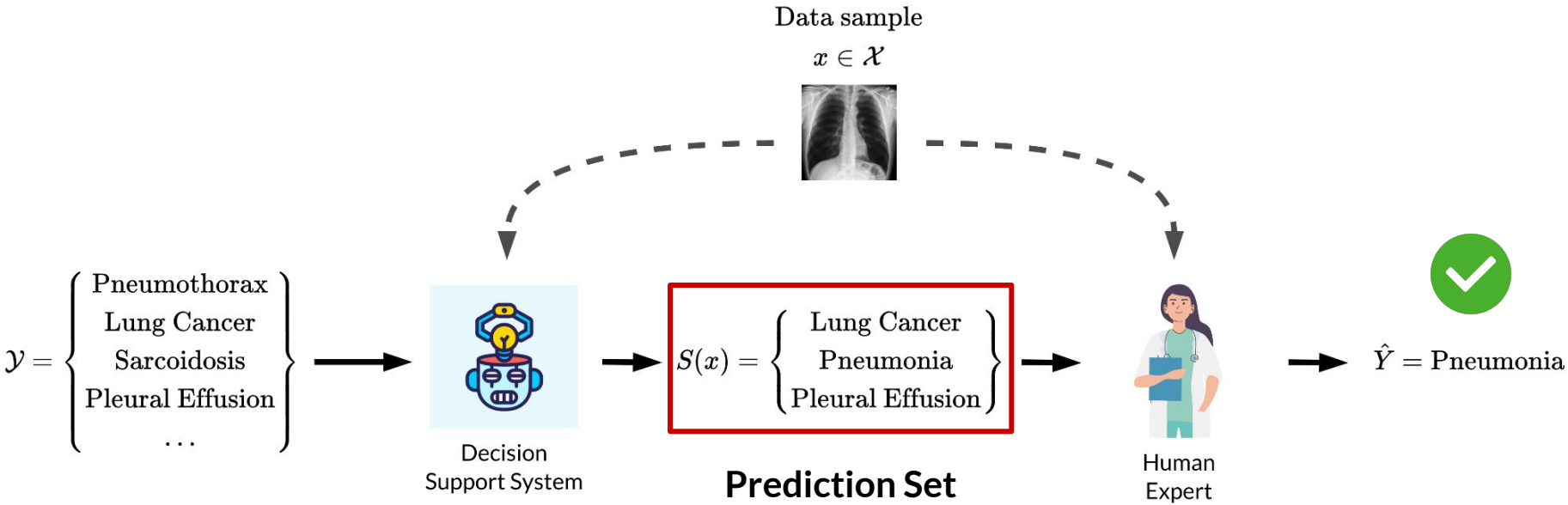
Decision Support System for multiclass classification



When the expert should trust the system?
How the expert should use the system's prediction?
[1–4]

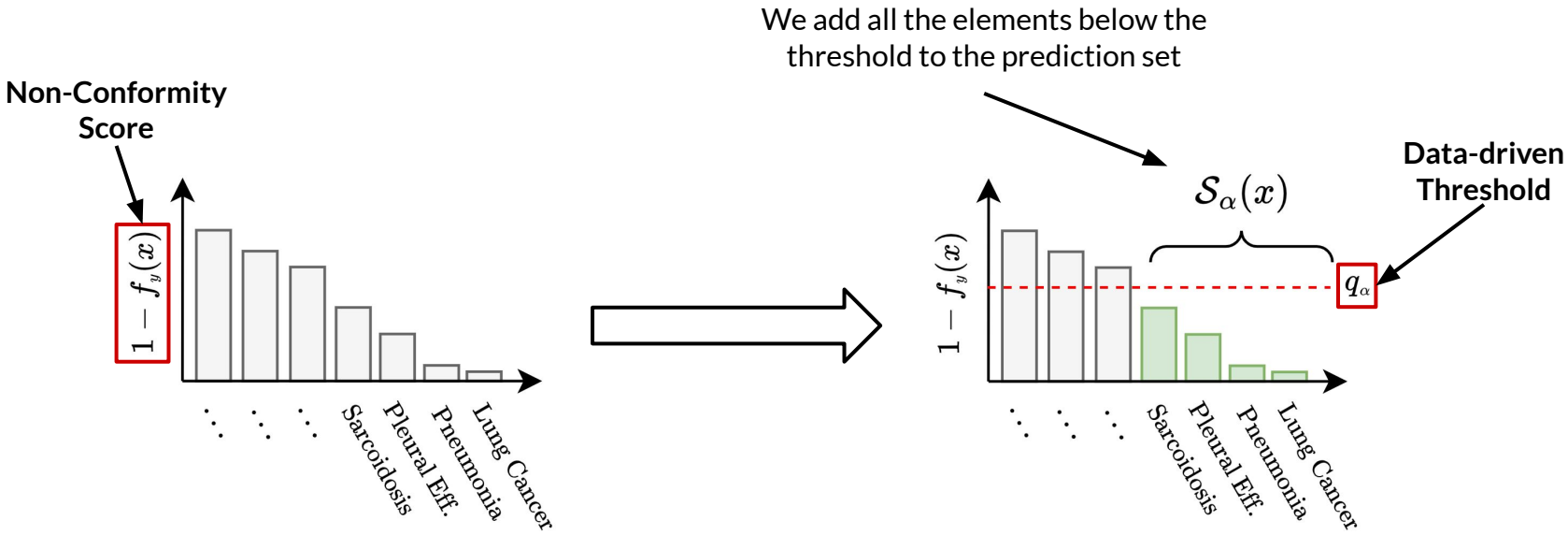
[1] Yin et al., Understanding the effect of accuracy on trust in machine learning models, CHI (2019)
[2] Zhang et al., Effect of confidence and explanation on accuracy and trust calibration in ai-assisted decision making, FAccT (2020)
[3] Suresh et al., Misplaced trust: Measuring the interference of machine learning in human decision-making, WebSci (2020)
[4] Lai et al., Towards a science of human-ai decision making: a survey of empirical studies, Preprint (2021)

Alternative Support Systems based on prediction sets [5,6]



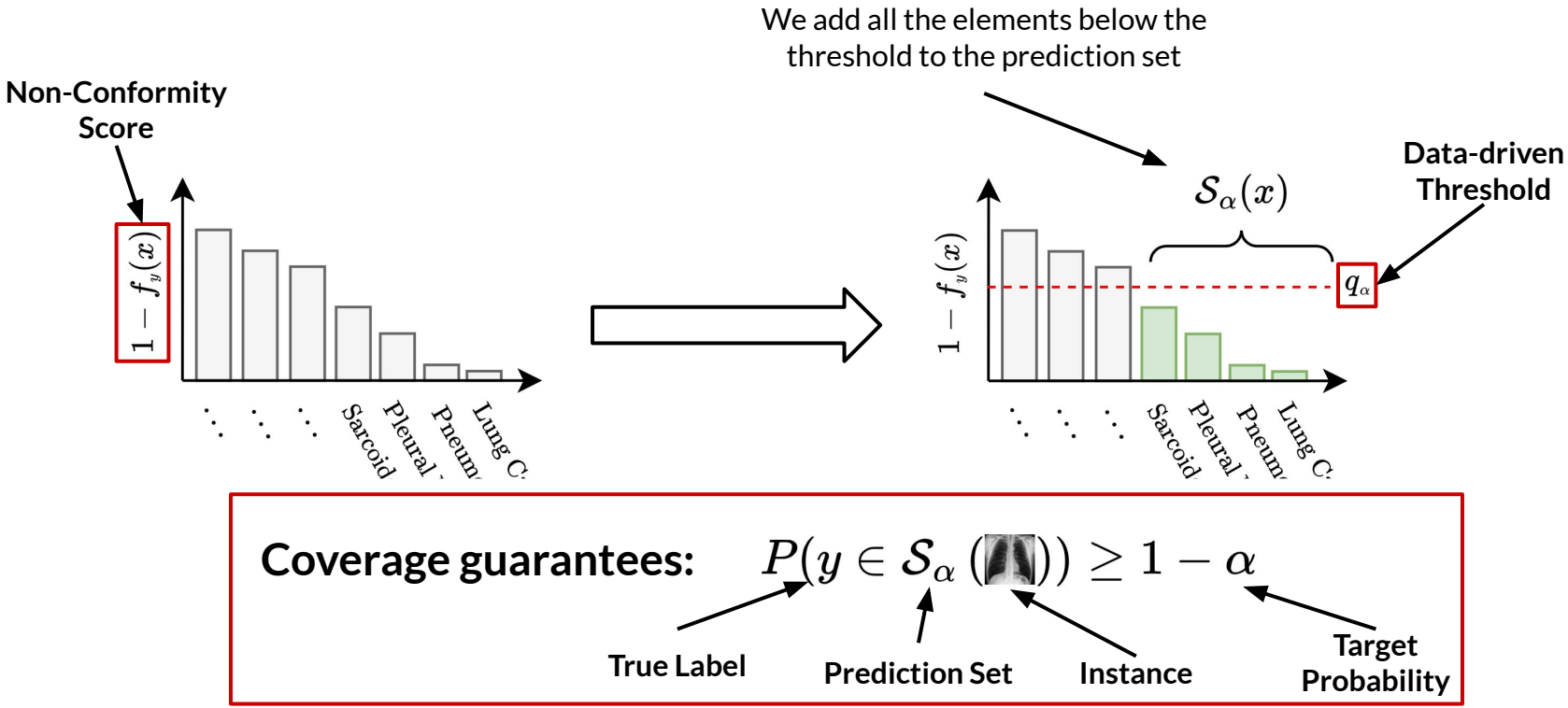
[5] Straitouri et al., Improving expert predictions with conformal prediction, ICML (2023)
[6] Eleni Straitouri and Manuel Gomez-Rodriguez, Designing decision support systems using counterfactual prediction sets, ICML (2024)

Prediction sets with Conformal Prediction [7,8]



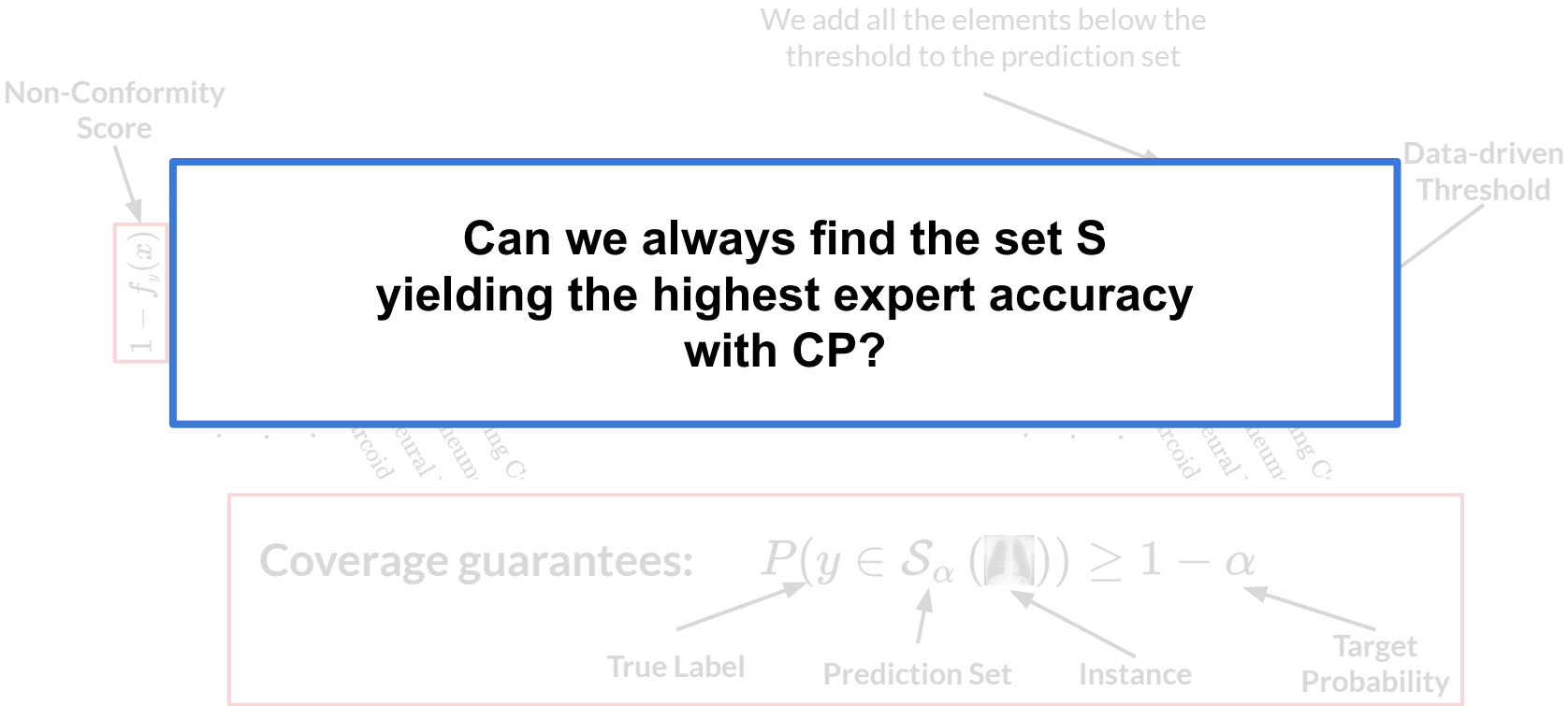
[7] Vladimir Vovk, Alexander Gammerman, and Glenn Shafer, Algorithmic learning in a random world, Vol. 29. New York: Springer (2005)
[8] Anastasios N. Angelopoulos, and Stephen Bates, Conformal prediction: A gentle introduction, Foundations and Trends® in Machine Learning (2023)

Prediction sets with Conformal Prediction [7,8]



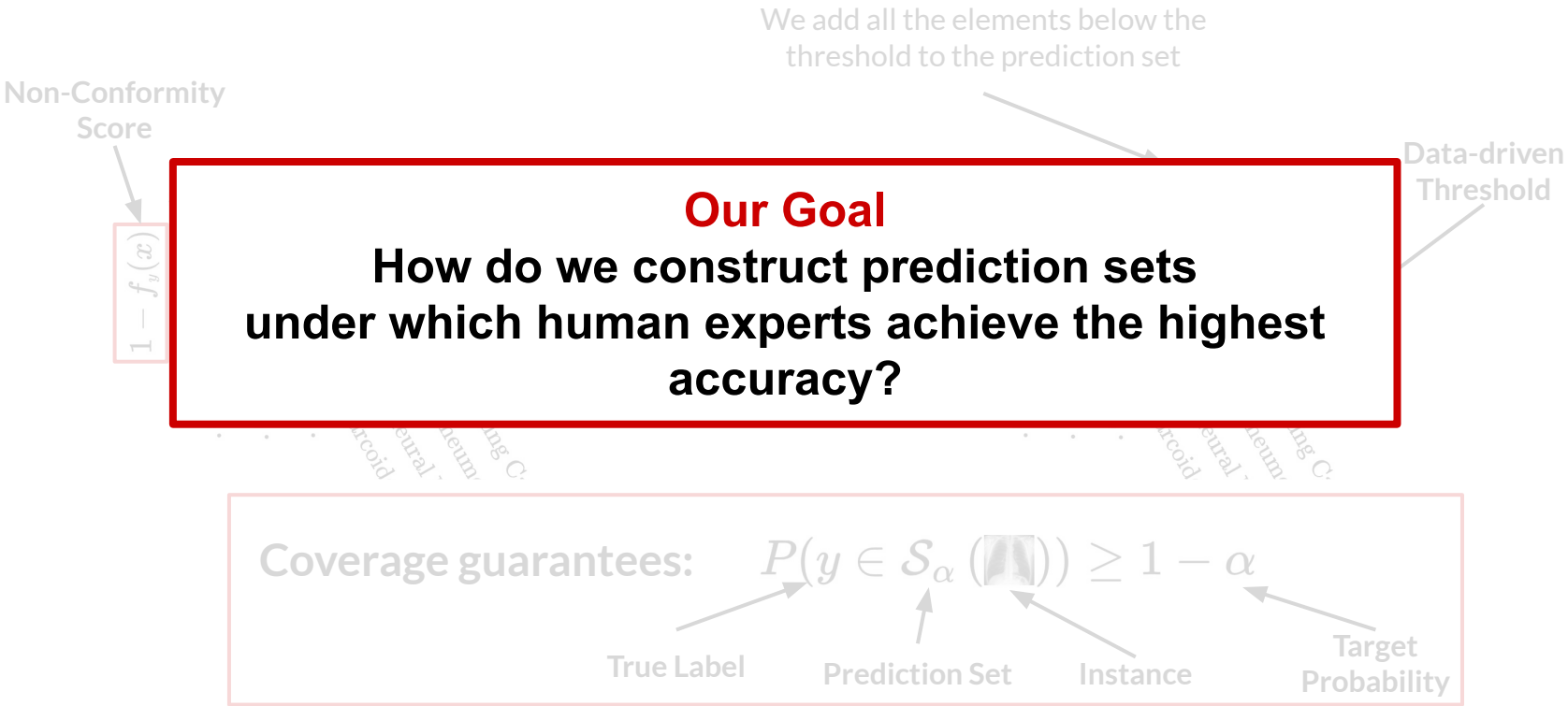
[7] Vladimir Vovk, Alexander Gammerman, and Glenn Shafer, Algorithmic learning in a random world, Vol. 29. New York: Springer (2005)
[8] Anastasios N. Angelopoulos, and Stephen Bates, Conformal prediction: A gentle introduction, Foundations and Trends® in Machine Learning (2023)

Prediction sets with Conformal Prediction [7,8]



[7] Vladimir Vovk, Alexander Gammerman, and Glenn Shafer, Algorithmic learning in a random world, Vol. 29. New York: Springer (2005)
[8] Anastasios N. Angelopoulos, and Stephen Bates, Conformal prediction: A gentle introduction, Foundations and Trends® in Machine Learning (2023)

Prediction sets with Conformal Prediction [7,8]



[7] Vladimir Vovk, Alexander Gammerman, and Glenn Shafer, Algorithmic learning in a random world, Vol. 29. New York: Springer (2005)
[8] Anastasios N. Angelopoulos, and Stephen Bates, Conformal prediction: A gentle introduction, Foundations and Trends® in Machine Learning (2023)

How can we build sets maximizing the expert's accuracy?

We approximate the expert as a multinomial logit model (MNL) [5,9]

Probability the expert picks y from $\mathcal{S}$ when y is the true label

$$g(\mathcal{S} \mid x) = \sum_{y \in \mathcal{S}} f_y(x) \frac{C_{yy}}{\sum_{y' \in \mathcal{S}} C_{y'y}}$$

Estimator of the expert's accuracy

[9] Florian Heiss, Discrete choice methods with simulation, Taylor & Francis (2016)

How can we build sets maximizing the expert's accuracy?

We approximate the expert as a multinomial logit model (MNL) [5,9]

Probability the expert picks y from $\mathcal{S}$ when y is the true label

$$g(\mathcal{S} \mid x) = \sum_{y \in \mathcal{S}} f_y(x) \frac{C_{yy}}{\sum_{y' \in \mathcal{S}} C_{y'y}}$$

Estimator of the expert's accuracy

Goal

Find the prediction set maximizing the expert accuracy:

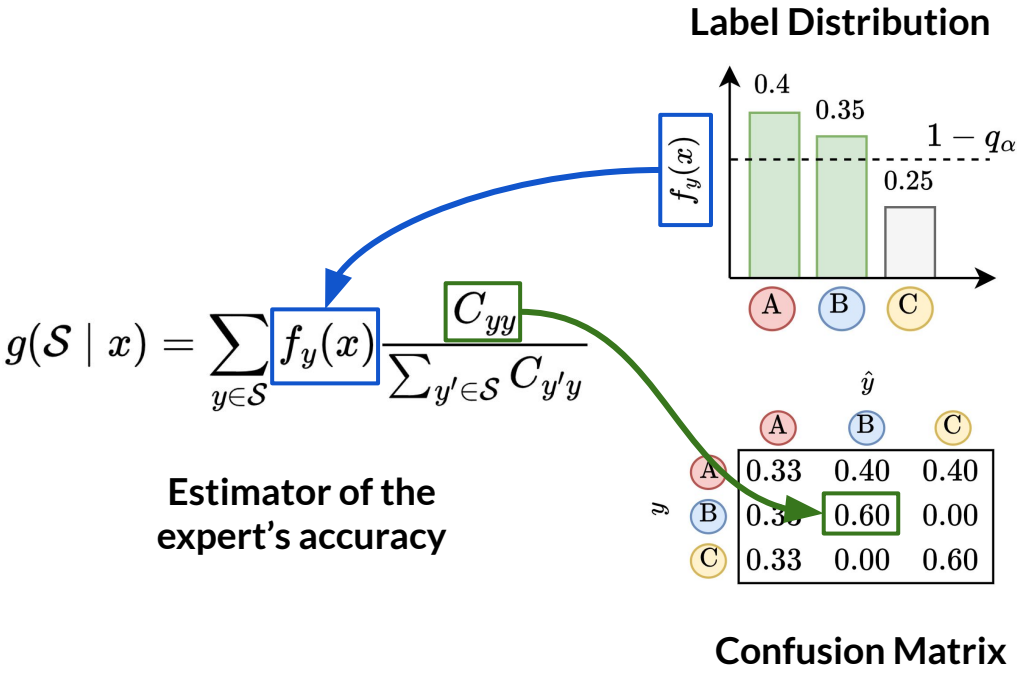
$$\mathcal{S}^* = \text{argmax}_{\mathcal{S} \subseteq \mathcal{Y}}\, g(\mathcal{S} \mid x)$$

where:

$$g(\mathcal{S} \mid x) = \sum_{y \in \mathcal{S}} f_y(x) \frac{C_{yy}}{\sum_{y' \in \mathcal{S}} C_{y'y}}$$

[9] Florian Heiss, Discrete choice methods with simulation, Taylor & Francis (2016)

Conformal Prediction is suboptimal

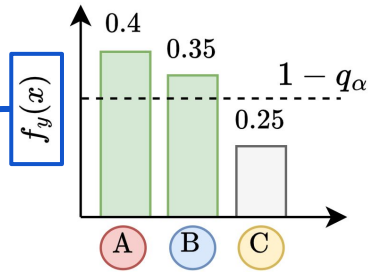


Conformal Prediction is suboptimal

$$g(\mathcal{S} \mid x) = \sum_{y \in \mathcal{S}} f_y(x) \frac{C_{yy}}{\sum_{y' \in \mathcal{S}} C_{y'y}}$$

Estimator of the expert's accuracy

Label Distribution



$\hat{y}$

y	A	B	C
A	0.33	0.40	0.40
B	0.33	0.60	0.00
C	0.33	0.00	0.60

C_{yy}

Confusion Matrix

All possible (conformal) prediction sets

$$g(\mathcal{S}_{cp} \mid x) = \begin{cases} 0.49 & \mathcal{S}_{cp} = \{\text{A}\ \text{B}\ \text{C}\} \\ 0.41 & \mathcal{S}_{cp} = \{\text{A}\ \text{B}\} \\ 0.4 & \mathcal{S}_{cp} = \{\text{A}\} \end{cases}$$

$$g(\{\text{B}\ \text{C}\} \mid x) = 0.60$$

There is a set with a higher objective!

Finding the optimal prediction set is hard

Goal

Find the prediction set maximizing the expert accuracy:

$$\mathcal{S}^* = \text{argmax}_{\mathcal{S} \subseteq \mathcal{Y}}\, g(\mathcal{S} \mid x)$$

where:

$$g(\mathcal{S} \mid x) = \sum_{y \in \mathcal{S}} f_y(x) \frac{C_{yy}}{\sum_{y' \in \mathcal{S}} C_{y'y}}$$

Finding the optimal set is **NP-Hard** and it is **NP-hard** to **approximate** to any non-trivial factor*

* Our reduction is from the *k*-clique and maximum clique problems, respectively.

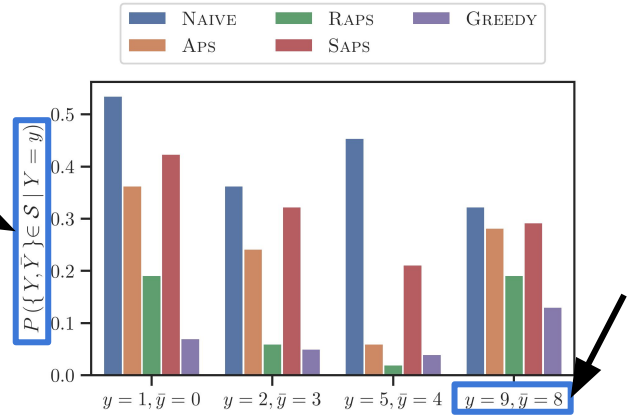
Greedy algorithm guaranteed to offer **equal or greater performance** than CP

Our greedy algorithm

Highest accuracy than CP in simulation studies with both **synthetic** and **real expert predictions**

Less likely to include labels the expert would **mistake** for the true label

Empirical probability that a prediction set includes the two labels



Label most frequently mistaken with the ground truth

Come to our poster on Wednesday 11th 11 a.m. — 2 p.m. PST