NEURAL INFORMATION PROCESSING SYSTEMS

NeurIPS 2026 · Paris

Stochastic Grouping Conformal Prediction

for Effective Subgroup Reliability

Meihui Zhong[1]* Wenxin Tai[1] Ting Zhong[1] Fan Zhou[1,2]†

[1] University of Electronic Science and Technology of China

[2] Intelligent Digital Media Technology, Key Laboratory of Sichuan Province

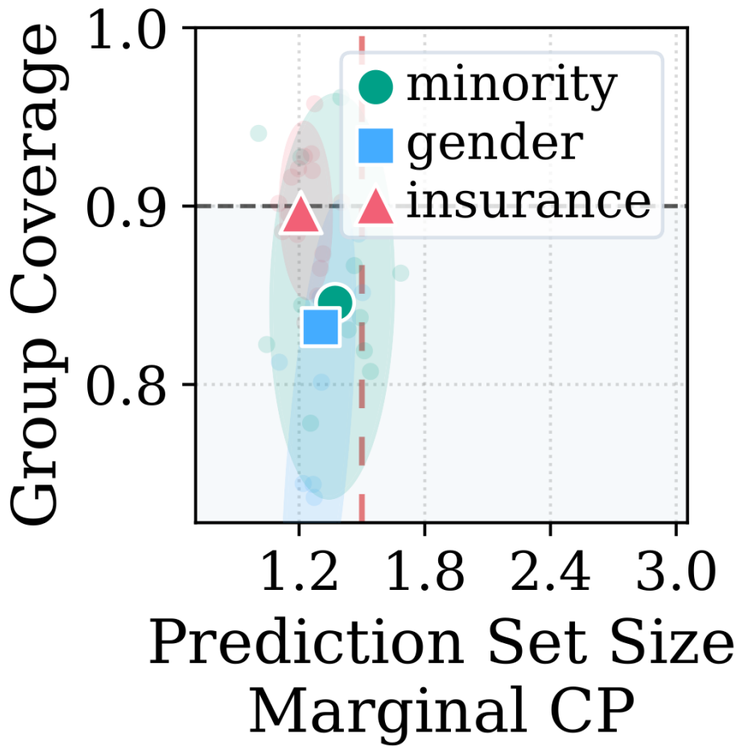
Marginal coverage can hide subgroup failures

90%

Nominal population coverage

A global threshold can average undercoverage in some groups with overcoverage in others

Adapt to local error patterns while sharing calibration evidence

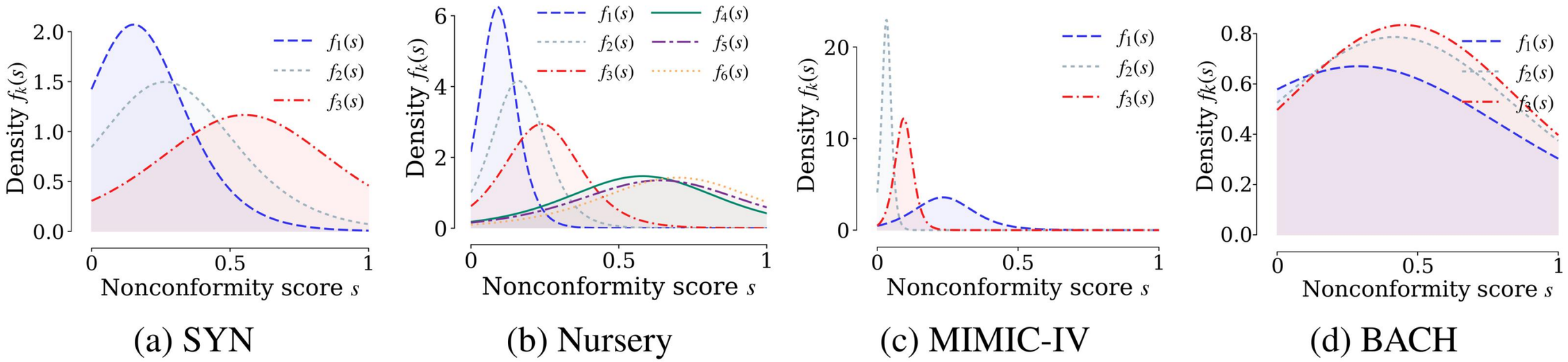


MIMIC-IV · Paper Fig. 1 (Marginal CP panel)

A local score law for each sample

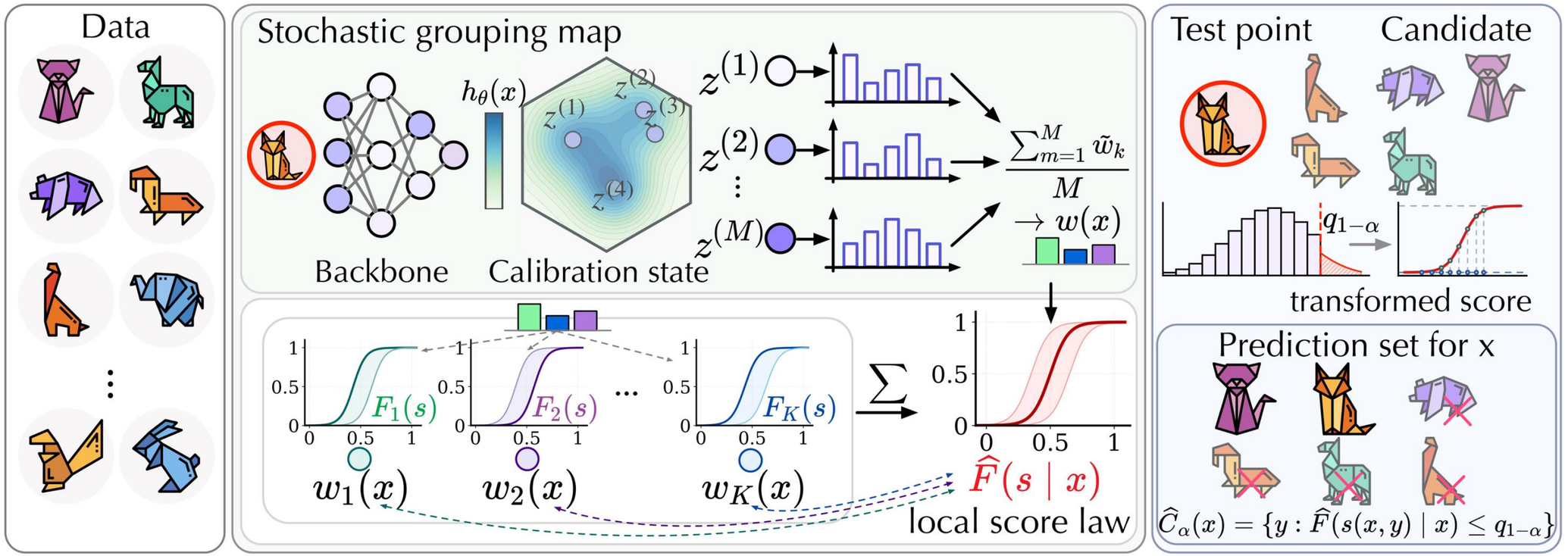
$$\widehat{F}(s \mid x) = \sum_{k=1}^{K} w_k(x) F_k(s)$$

Sample-specific memberships combine shared score distributions



Shared components capture different score regimes · Paper Fig. 3

From shared evidence to prediction sets



Stochastic memberships

Average over latent states

Shared score distributions

Fit component CDFs on D_1

Independent calibration

Set the threshold on Dcal

Independent calibration retains marginal validity

FINITE-SAMPLE MARGINAL COVERAGE

$$\Pr\{Y \in \hat{C}_\alpha(X)\} \geq 1 - \alpha$$

Fix the learned score transform, then calibrate on an independent split

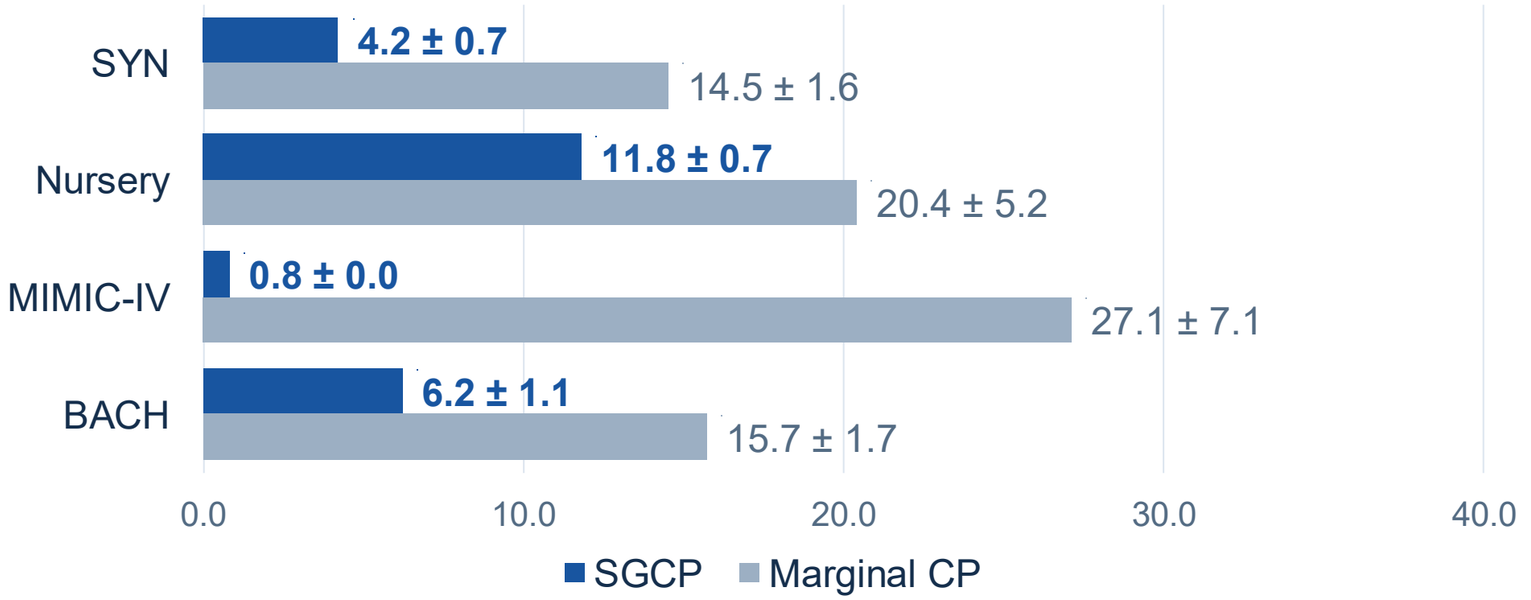
Requires exchangeable calibration and test examples
The score transform is fitted independently of both

Paper Proposition 2

Smaller subgroup gaps and prediction sets

Compared with Marginal CP · 90% target coverage · 20 runs · largest sample setting

Subgroup coverage gap ↓



Mean absolute subgroup coverage error, in percentage points
Bars show means; labels report mean ± SD

Average set size ↓

Dataset	Marginal CP	SGCP
SYN	2.309	**2.225**
Nursery	2.170	**1.089**
MIMIC-IV	1.226	**1.107**
BACH	2.894	**2.224**

Mean labels per prediction set.

Paper Table 1