SPACR: Single-Pass Adaptive Training of Uncertainty-Aware Conformal Regressors

Soundouss Messoudi **Sylvain Rousseau** **Sébastien Destercke**
Université de technologie de Compiègne, CNRS, Heudiasyc, Compiègne, France

Abstract

Conformal Prediction (CP) provides distribution-free guarantees but often yields inefficient intervals because standard models are not trained with conformal objectives. We propose the Single-Pass Adaptive Conformal Regressor (SPACR), which directly optimizes predictive efficiency and validity using a differentiable calibration proxy within its loss. Unlike prior work (e.g., DOICR), SPACR eliminates the need for batch-splitting or fixed confidence levels during training, enabling a single model to generate valid intervals for any confidence level at inference. Experiments show that SPACR consistently gives tighter intervals and superior efficiency-coverage trade-offs compared to state-of-the-art baselines.

1 Introduction

Uncertainty quantification is critical in safety-critical domains like healthcare and autonomous driving, where users require reliable signals to intervene on potential errors [Vazquez and Facelli, 2022, Michelmore et al., 2020]. Conformal Prediction (CP) addresses this by providing distribution-free, model-agnostic guarantees via prediction sets (for classification) or intervals (for regression) [Vovk et al., 2005, Angelopoulos and Bates, 2021]. However, standard Inductive CP (ICP) is typically applied *post hoc* to "black box" models. Because the underlying model is not optimized for the conformal objective, this often results in inefficient (unnecessarily wide) intervals.

This challenge has motivated the development of methods aimed at directly training models to maximize conformal efficiency. While differentiable objectives exist for classification [Bellotti, 2021, Stutz et al., 2022], regression remains under-explored. The primary existing regression method, DOICR [Lei and Bellotti, 2023], minimizes interval width while enforcing validity, but suffers from sample inefficiency due to mandatory batch splitting and the requirement to retrain for every specific confidence level.

To address these limitations, we propose the *Single-Pass Adaptive Conformal Regressor (SPACR)*, which integrates accuracy, efficiency, and validity into a unified, differentiable objective. Unlike prior work, it ensures calibration without batch splitting and enables a single model to generate valid intervals for multiple confidence levels at inference. Our contributions are:

- We introduce SPACR, a differentiable framework that jointly optimizes predictive efficiency and validity in its loss. This eliminates the disconnect between model training and conformal objectives.
- Unlike DOICR, SPACR eliminates the need for batch-splitting and specific confidence levels during training. A single trained model can thus adaptively generate valid intervals for any confidence level at inference.
- We demonstrate empirically that SPACR achieves tighter intervals and superior coverage-efficiency trade-offs compared to state-of-the-art baselines, while significantly reducing computational costs.

2 Related work

2.1 Conformal Prediction for Regression

Conformal Prediction (CP) [Vovk et al., 2005, Angelopoulos and Bates, 2021] constructs prediction intervals $\hat{C}_{1-\alpha}(x)$ satisfying the coverage property $\mathbb{P}\big(y \in \hat{C}_{1-\alpha}(x)\big) \geq 1-\alpha$, where $\alpha \in (0,1)$ is a user-specified significance level. Inductive Conformal Prediction (ICP) [Papadopoulos, 2008]

Workshop "Towards Trustworthy Predictions: Theory and Applications of Calibration for Modern AI" at AISTATS 2026, Tangier, Morocco.

is the efficient approach that splits data into training $\mathcal{D}_{train} = \{(x_i, y_i)\}_{i=1}^{n}$ and calibration $\mathcal{D}_{cal} = \{(x_j, y_j)\}_{j=1}^{n_{cal}}$ sets, trains a regression model $f : \mathcal{X} \to \mathbb{R}$ on $\mathcal{D}_{train}$, then computes non-conformity scores s_j on $\mathcal{D}_{cal}$.

To achieve adaptive intervals, normalized ICP [Papadopoulos and Haralambous, 2011] scales s_j with an estimated difficulty σ_j:

$$s_j = \frac{|y_j - f(x_j)|}{\sigma_j}. \tag{1}$$

Let q be the $(1-\alpha)$-quantile of these scores. For a new input x_{test}, the prediction interval is:

$$\hat{C}_{1-\alpha}(x_{test}) = \Big[f(x_{test}) - \hat{q}\sigma_{test},\ f(x_{test}) + \hat{q}\sigma_{test}\Big]. \tag{2}$$

This normalized adaptive mechanism generalizes standard ICP that used fixed-width intervals ($\sigma \equiv 1$), improving efficiency. However, the required data splitting in ICP reduces sample efficiency, motivating end-to-end approaches that conformally trains the model.

2.2 Training Conformal Regressors

Recent research aims to align model training with the conformal objective of minimizing interval width while ensuring validity. The state-of-the-art method, Directly Optimized Inductive Conformal Regression (DOICR) [Lei and Bellotti, 2023], adapts differentiable sorting techniques [Stutz et al., 2022] to regression. DOICR optimizes interval width by randomly splitting each training batch into a training set (D_1) and an embedded calibration set (D_2). It computes a differentiable quantile q on D_2 to minimize the loss:

$$\mathcal{L}_{\text{DOICR}} = \frac{2q}{|\mathcal{D}_1|} \sum_{(x_i, y_i) \in \mathcal{D}_1} \exp(u(x_i)),$$

where $\exp(u(x_i))$ represents the uncertainty estimate. However, DOICR has two significant limitations. First, the explicit batch splitting reduces data efficiency. Second, the loss depends on a fixed quantile q corresponding to a specific α, requiring the model to be retrained for every new confidence level.

3 SPACR : Single-Pass Adaptive Conformal Regressor

We introduce SPACR, a framework that integrates the conformal objective into the training loop without data splitting or fixed significance levels. Given a training set $\mathcal{D}_{train}$, our regressor $f_\theta(x)$ outputs a mean $\hat{y}$ and an uncertainty estimate $\hat{\sigma}_i = \exp(u_i)$, where the exponential map ensures positivity and stability.

To align with conformal validity, we minimize a loss function combining three objectives:

$$\mathcal{L}_{\text{SPACR}} = \underbrace{\frac{1}{n}\sum_{i=1}^{n} |\hat{y}_i - y_i|}_{\mathcal{L}_{\text{Accuracy}}} + \underbrace{\frac{1}{n}\sum_{i=1}^{n} \hat{\sigma}_i}_{\mathcal{L}_{\text{Efficiency}}} + \underbrace{\lambda \left[\frac{1}{n}\sum_{i=1}^{n} \phi(y_i, \hat{y}_i, \hat{\sigma}_i)\right]}_{\mathcal{L}_{\text{Validity}}}, \tag{3}$$

where $\mathcal{L}_{\text{Accuracy}}$ (MAE) ensures precision, $\mathcal{L}_{\text{Efficiency}}$ penalizes wide intervals, and λ controls $\mathcal{L}_{\text{Validity}}$ with $\phi(y, \hat{y}, \hat{\sigma}) = \max(|y - \hat{y}| - \hat{\sigma}, 0)$, which penalizes observations falling outside the predicted interval $[\hat{y} \pm \hat{\sigma}]$. This is equivalent to the α-insensitivity loss in SVMs [Vapnik et al., 1996] and forces the model to learn intervals that empirically cover the data.

As shown in Figure 1, this loss encourages the model to dynamically adjust interval widths based on empirical coverage violations, which aligns its training objective with the validity requirement of conformal prediction.

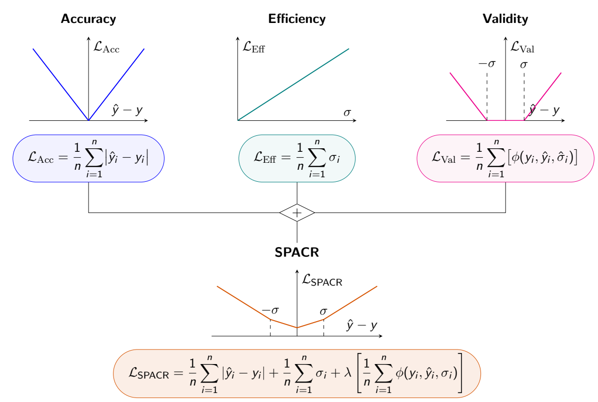


Figure 1: Decomposition of the SPACR loss into Accuracy, Efficiency, and Validity components.

Crucially, SPACR does not require splitting the training batch into sub-sets (see Algorithm 1), maximizing sample efficiency compared to methods like DOICR. Moreover, since $\hat{\sigma}$ is learned independently of a specific α, the model captures a general measure of uncertainty.

Inference. Post-training, SPACR follows the same ICP procedure as traditional conformal methods and other training-based techniques such as DOICR. Given a separate calibration set $\mathcal{D}_{cal}$, we compute non-conformity scores $s_j = \frac{|y_j - \hat{y}_j|}{\hat{\sigma}_j + \beta}$ where $\beta = 0.05\mu_{\hat{\sigma}}$ is a smoothing constant based on the mean calibration uncertainty. For a desired α, we calculate the $(1-\alpha)$-quantile $\hat{q}$ to construct intervals:

$$\hat{C}_{1-\alpha}(x_{test}) = \Big[\hat{y}_{test} - \hat{q}(\hat{\sigma}_{test} + \beta),\ \hat{y}_{test} + \hat{q}(\hat{\sigma}_{test} + \beta)\Big]. \tag{4}$$

Algorithm 1 SPACR Training Procedure

Input: Proper training dataset $\mathcal{D}_{train}$, neural network $f_\theta = (f_\theta^{(1)}, f_\theta^{(2)})$, number of epochs $E_{\max}$, learning rate η

Output: Trained conformalized model f_θ

```
Initialize model parameters θ
for epoch = 1 to E_max do
    Shuffle and partition D_train into batches 𝓑
    for each batch B ∈ 𝓑 do
        Forward pass: For each (x_i, y_i) ∈ B, compute predicted mean ŷ_i = f_θ^(1)(x_i) and uncertainty σ̂_i = exp(f_θ^(2)(x_i))
        Compute loss: Evaluate L_SPACR(B; θ) over the current batch B using Eq. 3
        Backward pass: Compute gradients ∇_θ L_SPACR
        Parameter update: Update θ using an optimizer (e.g., Adam) with learning rate η
    end for
end for
```

Since f_θ is not tied to α during training, a single SPACR model can generate valid intervals for multiple confidence levels simply by re-computing $\hat{q}$, avoiding the computational cost of retraining.

4 Experimental setting

4.1 Methodology

We evaluate SPACR against four baselines using a unified neural network architecture:

- **Standard ICP (SICP)**: Trains a model on MAE loss. Post-hoc calibration produces constant-width intervals (Eq. 2 with $\sigma \equiv 1$).
- **Normalized ICP (NICP)**: Similar to SICP but scales intervals using a trained difficulty estimator (Eq. 2) for adaptivity.
- **Conformalized Quantile Regression (CQR)**: [Romano et al., 2019] trains a model to predict lower and upper quantiles using the pinball loss then ensures validity with conformal calibration.
- **DOICR**: Directly Optimized Inductive Conformal Regression [Lei and Bellotti, 2023], which optimizes calibration during training via batch splitting (see Sec. 2.2).
- **SPACR (Ours)**: Optimizes a joint objective without batch splitting or fixed α (see Sec. 3).

Implementation Details. For tabular data, we use a neural network with three hidden layers (64 units). The output layer varies: SICP/NICP output a single scalar $\hat{y}$; DOICR/SPACR output mean and uncertainty $(\hat{y}, \hat{\sigma})$; CQR outputs two quantiles. NICP uses an additional detached layer for difficulty estimation. For the vision task, we fine-tune a pre-trained ResNet18. All models are trained using Adam ($\eta = 10^{-4}$, batch size 256) for 200 epochs. For SPACR, we set $\lambda = 5$. We report results over 5 random seeds.

Datasets We use diverse tabular benchmarks from [Grinsztajn et al., 2022]. For computer vision, we use the Drift dataset [Li et al., 2024], predicting tree counts from satellite imagery (using the Denmark subset, $n = 13,094$) and the UTK Face dataset [Zhang et al., 2017], consisting of over 20,000 face images, to predict age. All data is split 60/20/20 (train/calibration/test).

Metrics We report **Coverage** (percentage of true values inside the interval, target $1-\alpha$) and **Efficiency** (median interval width, where lower is better).

5 Results

5.1 Performance and Conditional Adaptivity

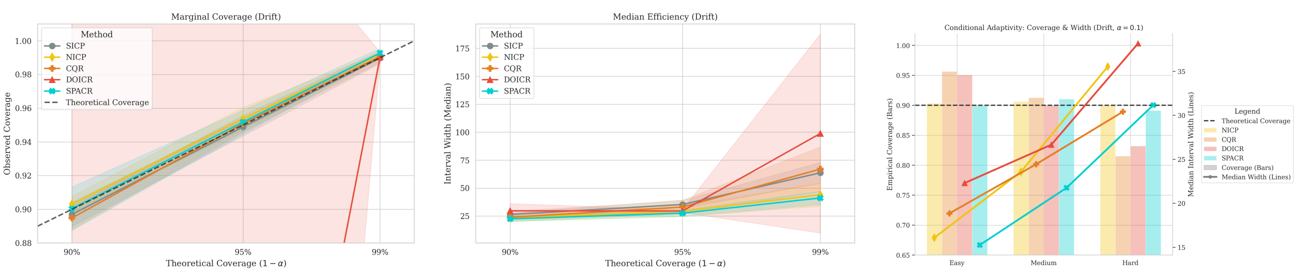
While most methods maintain marginal coverage within $\pm 1.5\%$, SPACR achieves a superior efficiency-coverage trade-off (Table 1, Figure 2). It yields the tightest intervals across most of the 16 datasets at 90% and 95% confidence, and remains highly competitive at 99%. SPACR avoids the severe under-coverage that DOICR exhibits on Brazilian Houses and Drift (Figure 2a). This stability extends conditionally: although exact distribution-free conditional coverage requires strong assumptions [Foygel Barber et al., 2021], SPACR reliably maintains the target 90% coverage across "Easy," "Medium," and "Hard" subsets by scaling interval widths monotonically with difficulty. In contrast, CQR and DOICR severely under-cover "Hard" instances while over-covering "Easy" ones (Figure 2c). Ultimately, SPACR learns a highly adaptive uncertainty representation that is globally efficient and locally valid.

5.2 Sensitivity and Ablation Analysis

Effect of λ. The parameter λ controls the trade-off between efficiency and validity. As $\lambda \to 0$, the model ignores the validity penalty, causing uncertainty estimates to collapse. This results in highly unstable, massive bounds due to calculation errors from vanishing denominators. As $\lambda \to \infty$, the model over-prioritizes

Dataset	Method	$\alpha = 0.10$ (90%) Cov. (%)	$\alpha = 0.10$ (90%) Med. Eff. ↓	$\alpha = 0.05$ (95%) Cov. (%)	$\alpha = 0.05$ (95%) Med. Eff. ↓	$\alpha = 0.01$ (99%) Cov. (%)	$\alpha = 0.01$ (99%) Med. Eff. ↓
Bike Sharing	SICP	90.13 ± 0.50	1.36 ± 0.02	94.79 ± 0.38	1.68 ± 0.02	99.13 ± 0.24	2.57 ± 0.09
	NICP	90.09 ± 0.14	**1.29 ± 0.04**	94.89 ± 0.49	1.55 ± 0.05	99.10 ± 0.27	2.20 ± 0.07
	CQR	89.76 ± 0.22	1.39 ± 0.06	94.69 ± 0.30	1.74 ± 0.05	**99.00 ± 0.35**	3.09 ± 0.08
	DOICR	90.07 ± 0.48	2.48 ± 0.22	95.20 ± 0.34	2.68 ± 0.22	99.03 ± 0.38	3.30 ± 0.23
	SPACR	**90.06 ± 0.52**	1.29 ± 0.09	**95.09 ± 0.41**	**1.54 ± 0.11**	99.02 ± 0.22	**2.07 ± 0.09**
California	SICP	90.41 ± 0.80	0.28 ± 0.01	95.16 ± 0.62	0.36 ± 0.01	**99.01 ± 0.15**	0.56 ± 0.01
	NICP	90.34 ± 0.79	0.28 ± 0.01	95.08 ± 0.60	0.35 ± 0.01	99.01 ± 0.20	0.56 ± 0.02
	CQR	90.14 ± 0.92	0.30 ± 0.01	95.10 ± 0.36	0.38 ± 0.01	99.08 ± 0.17	0.61 ± 0.01
	DOICR	90.18 ± 1.16	0.28 ± 0.02	95.22 ± 0.56	0.34 ± 0.01	99.02 ± 0.06	**0.47 ± 0.01**
	SPACR	**90.06 ± 0.94**	**0.26 ± 0.01**	**95.04 ± 0.57**	**0.33 ± 0.01**	99.02 ± 0.15	0.53 ± 0.03
Diamonds	SICP	**89.99 ± 0.26**	0.38 ± 0.01	95.06 ± 0.15	0.47 ± 0.00	99.05 ± 0.11	0.69 ± 0.01
	NICP	89.89 ± 0.36	0.37 ± 0.01	94.91 ± 0.05	0.45 ± 0.00	99.03 ± 0.10	0.67 ± 0.01
	CQR	89.94 ± 0.23	0.38 ± 0.01	94.91 ± 0.17	0.47 ± 0.01	98.98 ± 0.10	0.83 ± 0.02
	DOICR	89.86 ± 0.27	0.40 ± 0.01	94.91 ± 0.16	0.46 ± 0.01	98.94 ± 0.10	0.61 ± 0.01
	SPACR	90.05 ± 0.36	**0.34 ± 0.00**	**95.02 ± 0.11**	**0.42 ± 0.00**	**99.00 ± 0.08**	**0.60 ± 0.01**
House Sales	SICP	**89.87 ± 0.74**	0.56 ± 0.01	94.76 ± 0.39	0.72 ± 0.02	98.85 ± 0.32	1.13 ± 0.06
	NICP	89.65 ± 0.65	**0.54 ± 0.02**	94.53 ± 0.51	**0.68 ± 0.02**	98.84 ± 0.31	1.04 ± 0.03
	CQR	89.63 ± 0.85	0.54 ± 0.03	94.66 ± 0.44	0.69 ± 0.03	98.80 ± 0.36	1.13 ± 0.05
	DOICR	90.28 ± 0.45	2.30 ± 0.28	**94.99 ± 0.32**	2.56 ± 0.37	98.84 ± 0.08	2.64 ± 0.28
	SPACR	89.83 ± 1.04	0.55 ± 0.02	94.74 ± 0.41	0.68 ± 0.03	**98.87 ± 0.31**	**1.00 ± 0.03**
Medical Charges	SICP	89.85 ± 0.19	0.18 ± 0.00	94.91 ± 0.22	0.27 ± 0.01	98.99 ± 0.12	0.59 ± 0.02
	NICP	89.95 ± 0.15	0.18 ± 0.00	94.93 ± 0.19	0.26 ± 0.01	99.04 ± 0.09	0.55 ± 0.02
	CQR	**89.96 ± 0.08**	0.19 ± 0.01	94.96 ± 0.17	0.29 ± 0.01	98.98 ± 0.11	0.80 ± 0.04
	DOICR	90.12 ± 0.17	0.21 ± 0.01	94.92 ± 0.25	0.26 ± 0.01	99.00 ± 0.05	**0.37 ± 0.01**
	SPACR	89.90 ± 0.09	**0.17 ± 0.00**	**95.00 ± 0.20**	**0.24 ± 0.00**	**99.00 ± 0.07**	0.50 ± 0.01
Superconduct	SICP	89.83 ± 0.56	1.66 ± 0.04	94.88 ± 0.27	2.24 ± 0.06	**98.96 ± 0.28**	3.65 ± 0.18
	NICP	**89.92 ± 0.61**	1.52 ± 0.07	94.86 ± 0.44	2.06 ± 0.09	98.94 ± 0.24	3.50 ± 0.15
	CQR	89.56 ± 0.47	1.59 ± 0.02	94.59 ± 0.23	2.02 ± 0.02	98.93 ± 0.38	3.26 ± 0.11
	DOICR	89.74 ± 0.61	1.55 ± 0.05	94.85 ± 0.34	1.85 ± 0.04	98.91 ± 0.30	**2.60 ± 0.09**
	SPACR	89.79 ± 0.36	**1.29 ± 0.04**	**95.02 ± 0.27**	**1.69 ± 0.05**	98.83 ± 0.16	3.03 ± 0.11
Drift	SICP	89.70 ± 0.71	26.41 ± 2.73	94.89 ± 0.54	35.33 ± 3.64	**98.97 ± 0.32**	63.70 ± 9.28
	NICP	90.31 ± 0.47	23.55 ± 3.32	95.38 ± 0.67	29.35 ± 4.74	99.18 ± 0.11	44.13 ± 10.32
	CQR	89.48 ± 0.69	23.94 ± 3.59	**95.08 ± 0.49**	33.06 ± 6.29	99.06 ± 0.33	66.99 ± 19.81
	DOICR	53.65 ± 48.98	29.70 ± 6.46	57.04 ± 52.08	29.56 ± 1.34	99.04 ± 0.30	98.84 ± 89.02
	SPACR	**90.03 ± 1.29**	**22.51 ± 2.28**	95.17 ± 0.71	**27.59 ± 2.98**	99.27 ± 0.28	**41.19 ± 5.90**
Brazilian Houses	SICP	89.93 ± 0.52	**0.46 ± 0.04**	95.15 ± 0.28	**0.59 ± 0.04**	99.21 ± 0.16	0.94 ± 0.08
	NICP	89.72 ± 0.29	0.45 ± 0.05	94.98 ± 0.32	0.59 ± 0.06	99.23 ± 0.19	**0.93 ± 0.08**
	CQR	**89.97 ± 0.46**	0.51 ± 0.07	94.60 ± 0.73	0.66 ± 0.08	99.23 ± 0.26	1.23 ± 0.15
	DOICR	71.91 ± 40.21	1.53 ± 0.29	75.89 ± 42.42	1.87 ± 0.12	**99.14 ± 0.12**	2.47 ± 0.70
	SPACR	89.86 ± 0.69	1.07 ± 0.12	**95.00 ± 0.88**	1.22 ± 0.15	99.21 ± 0.22	1.59 ± 0.18
CPU Act	SICP	89.75 ± 1.37	**0.15 ± 0.01**	**94.97 ± 1.22**	**0.21 ± 0.02**	99.11 ± 0.31	0.58 ± 0.06
	NICP	89.79 ± 1.36	0.15 ± 0.01	95.25 ± 0.68	0.22 ± 0.02	99.12 ± 0.23	0.66 ± 0.17
	CQR	89.87 ± 0.94	0.16 ± 0.01	95.06 ± 1.18	0.28 ± 0.04	99.28 ± 0.26	0.78 ± 0.11
	DOICR	90.30 ± 0.47	0.97 ± 0.09	95.25 ± 0.76	1.12 ± 0.08	**99.02 ± 0.31**	1.28 ± 0.15
	SPACR	**89.96 ± 0.52**	0.24 ± 0.05	94.78 ± 0.71	0.27 ± 0.05	99.13 ± 0.15	**0.37 ± 0.05**
Fifa	SICP	90.22 ± 0.59	0.30 ± 0.01	95.15 ± 0.52	0.43 ± 0.01	99.08 ± 0.15	0.65 ± 0.02
	NICP	90.11 ± 0.46	0.30 ± 0.01	95.23 ± 0.47	0.44 ± 0.01	**99.04 ± 0.29**	0.63 ± 0.02
	CQR	**90.07 ± 0.70**	0.33 ± 0.00	94.86 ± 0.52	0.36 ± 0.00	99.21 ± 0.27	**0.43 ± 0.01**
	DOICR	90.10 ± 0.47	0.32 ± 0.01	**94.98 ± 0.52**	0.38 ± 0.01	99.08 ± 0.28	0.44 ± 0.02
	SPACR	90.27 ± 0.51	**0.28 ± 0.00**	94.90 ± 0.57	**0.36 ± 0.01**	98.93 ± 0.16	0.61 ± 0.04
Isolet	SICP	90.29 ± 1.69	0.88 ± 0.03	94.73 ± 1.45	1.19 ± 0.08	99.40 ± 0.12	2.33 ± 0.22
	NICP	89.71 ± 1.49	0.87 ± 0.03	**95.06 ± 1.60**	1.20 ± 0.11	99.33 ± 0.11	2.23 ± 0.17
	CQR	89.68 ± 1.51	0.87 ± 0.03	94.69 ± 1.47	1.15 ± 0.03	**99.00 ± 0.18**	2.01 ± 0.04
	DOICR	**90.23 ± 0.89**	1.13 ± 0.04	94.82 ± 0.41	1.29 ± 0.04	99.03 ± 0.28	1.70 ± 0.09
	SPACR	89.05 ± 1.26	**0.78 ± 0.06**	94.54 ± 0.70	**1.01 ± 0.07**	99.13 ± 0.28	**1.61 ± 0.11**
Pol	SICP	90.07 ± 0.80	0.35 ± 0.02	95.12 ± 0.77	0.94 ± 0.08	98.97 ± 0.15	3.44 ± 0.34
	NICP	**90.06 ± 0.83**	0.34 ± 0.02	95.11 ± 0.71	0.91 ± 0.07	99.03 ± 0.13	3.33 ± 0.27
	CQR	90.28 ± 0.68	0.49 ± 0.08	95.27 ± 0.23	1.32 ± 0.19	**98.99 ± 0.16**	2.48 ± 0.14
	DOICR	89.39 ± 0.87	0.79 ± 0.15	**94.92 ± 0.19**	1.20 ± 0.12	98.95 ± 0.32	1.72 ± 0.14
	SPACR	90.37 ± 0.43	**0.09 ± 0.03**	95.67 ± 0.37	**0.14 ± 0.04**	99.14 ± 0.17	**0.33 ± 0.08**
Wine Quality	SICP	90.46 ± 1.03	**0.34 ± 0.01**	**94.95 ± 0.64**	0.43 ± 0.02	**99.00 ± 0.37**	0.65 ± 0.06
	NICP	91.31 ± 1.22	0.36 ± 0.02	95.32 ± 0.89	0.44 ± 0.03	98.92 ± 0.26	0.67 ± 0.05
	CQR	90.26 ± 1.06	0.35 ± 0.01	94.80 ± 0.70	0.46 ± 0.02	99.11 ± 0.46	0.77 ± 0.05
	DOICR	**90.02 ± 0.75**	0.43 ± 0.04	95.46 ± 0.77	0.50 ± 0.04	99.18 ± 0.20	0.68 ± 0.04
	SPACR	90.48 ± 0.85	0.35 ± 0.01	95.31 ± 0.80	**0.43 ± 0.01**	99.02 ± 0.33	**0.65 ± 0.05**
UTK Face	SICP	89.97 ± 0.67	32.15 ± 4.12	94.98 ± 0.45	40.67 ± 5.42	98.99 ± 0.23	60.17 ± 6.94
	NICP	90.16 ± 0.35	30.36 ± 9.72	94.92 ± 0.34	38.97 ± 13.57	99.08 ± 0.03	58.48 ± 16.74
	CQR	**90.01 ± 0.33**	26.29 ± 3.96	95.01 ± 0.43	41.82 ± 8.57	**99.00 ± 0.14**	55.22 ± 6.05
	DOICR	89.46 ± 0.29	31.67 ± 2.57	**95.00 ± 0.22**	38.38 ± 6.59	98.94 ± 0.25	47.42 ± 5.11
	SPACR	90.13 ± 0.46	**21.84 ± 0.84**	95.02 ± 0.30	**27.49 ± 0.98**	98.95 ± 0.14	**42.51 ± 2.06**

Table 1: Comparison of methods across datasets for Marginal Coverage and Median Efficiency (Mean ± Std). Best results are highlighted in bold, invalid coverage is marked in red.



(a) Coverage vs. calibration line. (b) Median efficiency. (c) Conditional coverage & efficiency.

Figure 2: Performance figures for the different approaches for the Drift dataset.

coverage, severely inflating intervals at the cost of efficiency. Moderate values ($\lambda \in [1, 50]$) strikes the optimal balance, yielding tight intervals that dynamically adapt to the difficulty of the data, as demonstrated by the efficiency curves for the Isolet dataset (Figure 3).

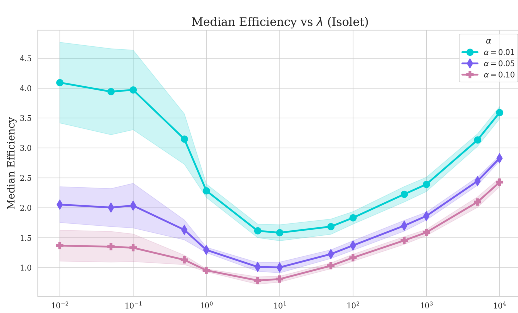


Figure 3: Efficiency Curves for varying λ values for the Isolet Dataset.

Ablation Study. We validated each loss component by evaluating variants of SPACR with specific terms removed. Omitting the efficiency penalty ($\mathcal{L}_{\text{Efficiency}}$) eliminates the optimization pressure for compactness, completely destroying conditional adaptability. This causes the interquartile range (IQR) of the predicted interval widths to collapse. For instance, on the California and Medical Charges datasets, the IQR drops to exactly 0.00 (at 90% confidence), meaning the model outputs a rigid, constant-width interval for all instances regardless of difficulty. Conversely, removing the validity penalty ($\mathcal{L}_{\text{Validity}}$) destabilizes training by ignoring empirical coverage violations, forcing the post-hoc calibration step to compensate with impractically wide intervals. On image datasets at 90% confidence, this degradation is severe: the median interval width inflates from 22.51 to 96.38 on the Drift dataset, and from 21.84 to 127.94 on the UTK Face dataset. Thus, $\mathcal{L}_{\text{Efficiency}}$ is crucial for instance-specific adaptability, while $\mathcal{L}_{\text{Validity}}$ is essential to prevent uninformative, exploding bounds.

5.3 Computational Cost

A critical advantage of SPACR is its $\mathcal{O}(1)$ single-pass design. Unlike DOICR and CQR, which require retraining for each of k confidence levels (scaling as $\mathcal{O}(k)$), SPACR learns a generalized uncertainty representation in one run ($\mathcal{O}(1)$), enabling dynamic inference for any α without additional training costs.

This yields significant empirical time savings: on the Drift dataset, SPACR trained in 32 minutes, whereas both DOICR and CQR required over 1.5 hours to support just three confidence levels $\alpha \in \{0.01, 0.05, 0.10\}$. Consequently, SPACR is highly suited for systems requiring dynamic, real-time calibration.

6 Conclusion

We introduce SPACR, a differentiable framework that trains uncertainty-aware regressors without batch-splitting or fixed confidence levels. By optimizing a joint objective, SPACR outperforms state-of-the-art methods like DOICR, yielding tighter intervals and enabling efficient multi-level inference with a single model. Future work will investigate alternative robust loss functions [Barron, 2019, Ghosh et al., 2017] and asymmetric variants [Huang et al., 2013] to further refine the interplay between the loss components.

References

[Angelopoulos and Bates, 2021] Angelopoulos, A. N. and Bates, S. (2021). A gentle introduction to conformal prediction and distribution-free uncertainty quantification. *arXiv preprint arXiv:2107.07511*.

[Barron, 2019] Barron, J. T. (2019). A general and adaptive robust loss function. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4331–4339.

[Bellotti, 2021] Bellotti, A. (2021). Optimized conformal classification using gradient descent approximation. *arXiv preprint arXiv:2105.11255*.

[Foygel Barber et al., 2021] Foygel Barber, R., Candes, E. J., Ramdas, A., and Tibshirani, R. J. (2021). The limits of distribution-free conditional predictive inference. *Information and Inference: A Journal of the IMA*, 10(2):455–482.

[Ghosh et al., 2017] Ghosh, A., Kumar, H., and Sastry, P. S. (2017). Robust loss functions under label noise for deep neural networks. In *Proceedings of the AAAI conference on artificial intelligence*, volume 31.

[Grinsztajn et al., 2022] Grinsztajn, L., Oyallon, E., and Varoquaux, G. (2022). Why do tree-based models still outperform deep learning on typical tabular data? *Advances in neural information processing systems*, 35:507–520.

[Huang et al., 2013] Huang, X., Shi, L., and Suykens, J. A. (2013). Support vector machine classifier with pinball loss. *IEEE transactions on pattern analysis and machine intelligence*, 36(5):984–997.

[Lei and Bellotti, 2023] Lei, H. and Bellotti, A. (2023). Reliable prediction intervals with directly optimized inductive conformal regression for deep learning. *Neural Networks*, 168:194–205.

[Li et al., 2024] Li, S., Gominski, D., Brandt, M., Tong, X., and Ciais, P. (2024). Get your embedding space in order: Domain-adaptive regression for forest monitoring. In *European Conference on Computer Vision*, pages 94–111. Springer.

[Michelmore et al., 2020] Michelmore, R., Wicker, M., Laurenti, L., Cardelli, L., Gal, Y., and Kwiatkowska, M. (2020). Uncertainty quantification with statistical guarantees in end-to-end autonomous driving control. In *2020 IEEE international conference on robotics and automation (ICRA)*, pages 7344–7350. IEEE.

[Papadopoulos, 2008] Papadopoulos, H. (2008). *Inductive conformal prediction: Theory and application to neural networks*. INTECH Open Access Publisher Rijeka.

[Papadopoulos and Haralambous, 2011] Papadopoulos, H. and Haralambous, H. (2011). Reliable prediction intervals with regression neural networks. *Neural Networks*, 24(8):842–851.

[Romano et al., 2019] Romano, Y., Patterson, E., and Candes, E. (2019). Conformalized quantile regression. *Advances in neural information processing systems*, 32.

[Stutz et al., 2022] Stutz, D., Dvijotham, K., Cemgil, A. T., and Doucet, A. (2022). Learning optimal conformal classifiers. In *International Conference on Learning Representations*.

[Vapnik et al., 1996] Vapnik, V., Golowich, S., and Smola, A. (1996). Support vector method for function approximation, regression estimation and signal processing. *Advances in neural information processing systems*, 9.

[Vazquez and Facelli, 2022] Vazquez, J. and Facelli, J. C. (2022). Conformal prediction in clinical medical sciences. *Journal of Healthcare Informatics Research*, 6(3):241–252.

[Vovk et al., 2005] Vovk, V., Gammerman, A., and Shafer, G. (2005). *Algorithmic learning in a random world*. Springer Science & Business Media.

[Zhang et al., 2017] Zhang, Z., Song, Y., and Qi, H. (2017). Age progression/regression by conditional adversarial autoencoder. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 5810–5818.