

Conformal Prediction for Image Segmentation Using Morphological Prediction Sets

Luca Mossina[†] and Corentin Friedrich

IRT Saint Exupéry, Toulouse, France

{luca.mossina, corentin.friedrich}@irt-saintexupery.com

Abstract. Image segmentation is a challenging task influenced by multiple sources of uncertainty, such as the data labeling process or the sampling of training data. In this paper we focus on binary segmentation and address these challenges using conformal prediction, a family of model- and data-agnostic methods for uncertainty quantification that provide finite-sample theoretical guarantees and applicable to any pretrained predictor. Our approach involves computing nonconformity scores, a type of prediction residual, on held-out calibration data not used during training. We use dilation, one of the fundamental operations in mathematical morphology, to construct a margin added to the borders of predicted segmentation masks. At inference, the predicted set formed by the mask and its margin contains the ground-truth mask with high probability, at a confidence level specified by the user. The size of the margin serves as an indicator of predictive uncertainty for a given model and dataset. We work in a regime of minimal information as we do not require any feedback from the predictor: only the predicted masks are needed for computing the prediction sets. Hence, our method is applicable to any segmentation model, including those based on deep learning; we evaluate our approach on several medical imaging applications. Our code is available at <https://github.com/deel-ai-papers/consema>.

Keywords: Image Segmentation · Conformal Prediction · Uncertainty Quantification.

1 Introduction

Uncertainty Quantification (UQ) is essential for ensuring the reliability of Machine Learning (ML) models in critical fields like medical imaging [21]. In image segmentation, uncertainties can stem from various sources, including data labeling and sampling. If such predictions are part of a complex system, such as an automated aid in medical diagnostics, one needs to rigorously quantify the prediction errors. We use Conformal Prediction (CP) [38,18], a framework that provides model- and data-agnostic methods for UQ with finite-sample theoretical guarantees, applicable to any pretrained predictor. It constructs prediction

[†] Corresponding author.

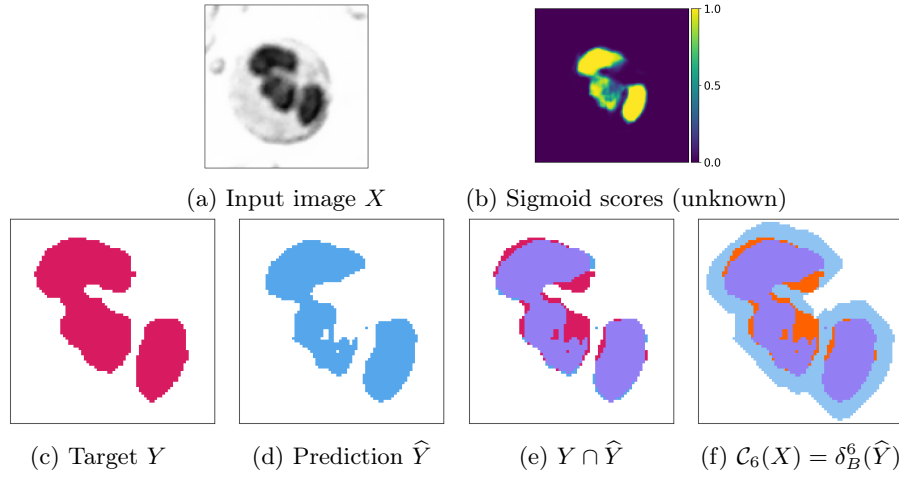


Fig. 1: Example: White Blood Cell (WBC) dataset [39], prediction (nucleus) with UniverSeg [11]. (a) Input image X . (b) Sigmoid scores $\hat{f}(X)$, assumed to be unavailable. (c) ground-truth mask Y . (d) predicted mask \hat{Y} . (e) intersection of Y and \hat{Y} in purple (true positives). (f) prediction set $\mathcal{C}_\lambda(\hat{Y})$: adding a margin via $\lambda = 6$ dilations of \hat{Y} by structuring element $B = \begin{bmatrix} \blacksquare & \blacksquare & \blacksquare \\ \blacksquare & \blacksquare & \blacksquare \\ \blacksquare & \blacksquare & \blacksquare \end{bmatrix}$, the missing pixels (e, in red) are covered, as per nonconformity score in Eq. (3). **Colors** \blacksquare : true positives; light blue : dilation margin; orange : false negatives recovered.

sets that contain the truth at a confidence level defined by the user, using held-out i.i.d.¹ calibration data from the same distribution as production data. We focus on binary segmentation, where each pixel is classified as either belonging to the object (e.g., a tumor) or the background.

Our contribution. We propose a novel approach to CP for image segmentation, working with a minimal set of hypotheses: only the binary prediction masks (\hat{Y} , Fig. 1d) are needed. Unlike existing methods, we do not require access to the predictor $\hat{f}(\cdot)$ nor its feedback (e.g., logits); thus, our method is applicable to black-box predictors, e.g., embedded in third-party software and machines or derived from complex foundation models [24].

We build CP sets as margins to be added on the contours of masks (Fig. 1f) using morphological dilation. The size of these margins depends on the nonconformity scores (Eq. 3) measured on held-out calibration data. This method can be used to validate a model (knowing the typical error we incur into, on production data), but also to provide a set of pixels (the “conformal margin”) that are likely to contain the part of ground truth we might have missed. Although we focus on medical imaging, it is applicable to any use case and any segmentation model.

¹ CP also applies to *exchangeable* data, which is a less strict requirement.

2 Background

Conformal Prediction (CP) [38,3,12] constructs prediction sets $\mathcal{C}(X)$ that contain the ground truth Y with probability $\mathbb{P}\{Y \in \mathcal{C}(X)\} \geq 1 - \alpha$, where $\alpha \in (0, 1)$ is a user-specified error level (also “risk”). We use inductive (or “split”) CP [28], which computes nonconformity scores (prediction residuals) on held-out, labeled calibration data that are independent of the training data and follow the same distribution as the test data. The size of $\mathcal{C}(X)$ is often interpreted as a measure of uncertainty,² as it depends on a quantile of the nonconformity scores.

Conformal Prediction in image segmentation. Using a threshold $\lambda \in [0, 1]$ on the sigmoid scores, [6,4,9] construct prediction sets $\mathcal{C}_\lambda(X) = \{\text{all pixels whose sigmoid score is } \geq 1 - \lambda\}$ with distribution-free risk-controlling procedures in binary segmentation. In [27], they extend the method of [4] to account for multiple classes at once, where each class channel can be seen as a binary mask. Furthermore, [13] builds inner and outer prediction sets that capture the ground truth with high probability, and they propose a nonconformity score based on the distance to the boundary of the masks. The methods of [23,7] use a spatially-aware weighting of the scores, under the hypothesis of pixel-wise exchangeability. Our work is also related to CP for object detection [17,22,2,1,36], where a “conformal” margin is added around the bounding boxes.

2.1 Nested prediction sets

Following the formulation of CP based on nested prediction sets of [18], let X and Y be the input features and target; let $\{\mathcal{C}_\lambda(X)\}_{\lambda \in \Lambda}$ be a sequence of prediction sets, where Λ is an ordered set (e.g., $\Lambda \subset \mathbb{R}^+$ or $\Lambda \subset \mathbb{N}$). This is said to be a *sequence of nested sets* when, for any $\lambda \leq \lambda'$, we have $\mathcal{C}_\lambda(X) \subseteq \mathcal{C}_{\lambda'}(X)$. The nonconformity score induced by $\mathcal{C}_\lambda(X)$ is then $r(X, Y) = \inf\{\lambda \in \Lambda : Y \in \mathcal{C}_\lambda(X)\}$. In words, the score is the smallest parameter λ such that the prediction set built from features X contains the true Y (see [18] for more examples). Given calibration data $(X_i, Y_i)_{i=1}^n$, one can compute the empirical quantile $\hat{\lambda}$ in (1) which is then used at inference to build $\mathcal{C}_{\hat{\lambda}}(X_{\text{test}})$.

$$\hat{\lambda} = \lceil (n+1)(1-\alpha) \rceil\text{-th largest score in } (r(X_i, Y_i))_{i=1}^n. \quad (1)$$

3 Methods



Let $(X_i, Y_i)_{i=1}^n$ be a sequence of calibration points where X_i is the input image, and Y_i its ground-truth mask labeled by an expert. We notate with \hat{Y} the predicted mask and we assume that the segmentation predictor \hat{f} is unknown or inaccessible, based on deep learning or any other algorithm. Note that Y and \hat{Y} are sets of points (i.e. pixels), hence the usual set notation applies.

For binary segmentation, we aim to avoid false negatives in the prediction: that is, every ground-truth pixel is contained in the predicted mask, denoted as

² CP does not distinguish [27] between *aleatoric* and *epistemic* uncertainty [20].

$Y \subseteq \hat{Y}$. To do so, the core of our proposal is to add a margin $\mu(\hat{Y})$ around the prediction \hat{Y} so that we cover all true positives: the prediction set is $\mathcal{C}_\lambda(X) = \hat{Y} \cup \mu(\hat{Y})$ and the condition becomes $Y \subseteq \mathcal{C}_\lambda(X)$.³ With respect to Fig. 1.c, our method statistically covers the red pixels (false negatives), whereas the blue pixels are false positives and deemed innocuous.

3.1 Nested sets via morphological dilations

We use *morphological dilation*, the fundamental operator of mathematical morphology [26,31,33,30,16,10]. Binary dilation $\delta_B(\cdot)$ on a discrete set (e.g., a binary digital image) is performed using a structuring element (SE) B , which defines pixel connectivity. Common choices include a 3×3 cross (, 4-connectivity) and a 3×3 square (, 8-connectivity) [16]. One iteration of dilation passes B over the image and assigns a value of 1 to all zero-valued pixels that have at least one neighboring 1-pixel under B .

Our proposal for the prediction set is to choose a structuring element B and apply a dilation $\lambda \in A \subseteq \mathbb{N}$ times:

$$\mathcal{C}_\lambda(X) := \underbrace{(\delta_B \circ \delta_B \circ \dots \circ \delta_B)}_{\lambda \text{ iterations}}(\hat{Y}) = \delta_B^\lambda(\hat{Y}), \quad (2)$$

with $\delta^0(\hat{Y}) := \hat{Y}$ and $\delta_B^\lambda(\hat{Y}) := \delta_B(\delta_B^{\lambda-1}(\hat{Y}))$. Also, the set $\mu^\lambda(\hat{Y}) = \mathcal{C}_\lambda(X) \setminus \hat{Y}$ is what we call the *margin* of $\mathcal{C}_\lambda(X)$. Morphological dilation is extensive, so the dilated set always increases in size and contains the original set, until the whole image is covered. It follows that for any (nonempty) prediction mask \hat{Y} , the sequence $(\delta_B^\lambda(\hat{Y}))_{\lambda \in A}$ forms a sequence of nested sets. Furthermore, it would be straightforward to extend this method to negative margins, using erosions on the background if the predictions were over-covering the ground truth.

Note that any operation that preserves the nested conditions of Sec. 2.1 is applicable, such as chaining several structuring elements to induce specific shape on the margin or having SE's of variable size. For example, it is possible to obtain the distance-based score of [13] doing a single dilation with a structuring element $B(\lambda)$ that grows with λ , so that $B(\lambda) \subset B(\lambda')$ for any $\lambda < \lambda'$. Then, the prediction set is $\mathcal{C}_\lambda(X_i) = \delta_{B(\lambda)}^1(\hat{Y}_i)$, where, for $\lambda = 0$, we have $B(0) := \emptyset$ and $\delta_{\emptyset}(\hat{Y}) := \hat{Y}$. This can be a discrete approximation of a disc or any other shape. As above, it also holds that $(\delta_{B(\lambda)}^1(\hat{Y}))_{\lambda \in A}$ is a sequence of nested sets.

One must rely exclusively on prior knowledge or training data when selecting the morphological operation and B . Using the calibration data for this purpose would violate the i.i.d. or exchangeability assumptions required by CP.

Nonconformity score. For a calibration pair (X_i, Y_i) , we define the score as the smallest value $\lambda \in \mathbb{N}$ (i.e., number of dilations) such that at least $\tau \times 100\%$

³ We write $\mathcal{C}_\lambda(X)$ to be consistent with the literature, although we could write $\mathcal{C}_\lambda(\hat{Y})$, since we do not need access to X nor the underlying predictor \hat{f} .

of the ground-truth pixels in Y_i are contained within the prediction set $\mathcal{C}_\lambda(X_i)$, where the hyperparameter $\tau \in [0, 1]$ is the *coverage ratio* [27]. More formally,

$$r(X_i, Y_i) = \inf \left\{ \lambda \in \mathbb{N} : \frac{|Y_i \cap \mathcal{C}_\lambda(X_i)|}{|Y_i|} \geq \tau \right\}, \quad (3)$$

where $|\cdot|$ is the number of elements (pixels) in a set. In some rare events, demanding to cover the entire truth can be overly conservative, that's why we introduce the hyperparameter τ that enables a trade-off. For instance, a $\tau = 0.999$ implies that the user can accept up to 0.1% of false negatives in $\mathcal{C}_\lambda(X)$.

Since by construction we have that for any $\lambda \leq \lambda'$, $\mathcal{C}_\lambda(X_i) \subseteq \mathcal{C}_{\lambda'}(X_i)$, this can be applied into the formulation of conformal nested prediction sets [18,4] (see Sec. 2.1) and the following holds:

Theorem 1. *Let $\hat{\lambda}$ be computed as in Eq. (1). Under the hypotheses of inductive conformal prediction [28,18,4], for the nonconformity score in (3) induced by prediction sets (2), it holds true that, for a new point $(X_{\text{test}}, Y_{\text{test}})$,*

$$\mathbb{P} \left[\frac{|Y_{\text{test}} \cap \mathcal{C}_{\hat{\lambda}}(X_{\text{test}})|}{|Y_{\text{test}}|} \geq \tau \right] \geq 1 - \alpha. \quad (4)$$

Proof. For any $\tau \in [0, 1]$, it suffices to set a binary loss to $\ell(\mathcal{C}_\lambda(X), Y) = \mathbb{1}\{\frac{|Y \cap \mathcal{C}_\lambda(X)|}{|Y|} \not\geq \tau\}$ (monotone in λ) and apply Conformalized Risk Control (CRC) as per Theorem 1 in [4], where they show that CRC with binary losses and CP are statistically equivalent. \square

The CP guarantee in Eq.(4) is said to hold *marginally*, i.e., on average over all possible inputs X_{test} and on average over repeated draws of the calibration and test samples; see [3] for statistical details. Eq. (1) implies that, for an *a priori* fixed α , the sample size must be $n \geq \frac{1}{\alpha} - 1$. Similarly, for a fixed calibration set of size n , the user can choose (prior to calibration) an error value $\alpha \geq \frac{1}{n+1}$.

Conformalization algorithm The “conformalization” of a (unknown) pre-trained segmentation predictor \hat{f} boils down to:

1. set $\alpha \in (0, 1)$ and collect labeled calibration data $(X_i, Y_i)_{i=1}^n$, with $n \geq \frac{1}{\alpha} - 1$;
2. fix a coverage ratio $\tau \in [0, 1]$ and a B for prediction set $\mathcal{C}_\lambda(\cdot)$ in Eq. (2);
3. compute the nonconformity scores $(r(X_i, Y_i))_{i=1}^n$ as per Eq. (3);
4. compute the empirical quantile $\hat{\lambda}$ as in Eq. (1);
5. for a test prediction \hat{Y}_{test} , use $\hat{\lambda}$ in $\mathcal{C}_\lambda(\cdot)$ and compute the dilated mask.

4 Experiments

We ran our experiments with two segmentation models and three dataset groups. First, we used the pretrained UniverSeg model [11].⁴ We tested it on two public

⁴ <https://github.com/JJG0/UniverSeg>, accessed 2024-04-08.

datasets as also evaluated in their paper: WBC (White Blood Cells) [39]⁵ and OASIS [25,19]⁶, a neuroimaging dataset. As in [4], we also ran experiments using the PraNet [14] model⁷ and the collection of datasets it was trained on, covering polyp segmentation in colonoscopy images (referred to as Polyps): ETIS [32], CVC-ClinicDB [8], CVC-ColonDB [34], EndoScene [37], and Kvasir [29].

For conformalization, we randomly shuffled and partitioned (with a ratio of 50/50) the original test set into calibration and proper test sets, yielding 50 calibration samples for WBC and OASIS, and 250 for the Polyps dataset. We then applied the algorithm in Sec. 3.1 and compute the following metrics:

$$\text{Empirical coverage: } \text{Cov}(\hat{\lambda}; \mathcal{C}, \tau) = \frac{1}{n_{\text{test}}} \sum_{i=1}^{n_{\text{test}}} \mathbb{1} \left\{ \frac{|Y_i \cap \mathcal{C}_{\hat{\lambda}}(X_i)|}{|Y_i|} \geq \tau \right\}, \quad (5)$$

$$\text{Stretch [2]: } \phi(\hat{\lambda}; \mathcal{C}) = \frac{1}{n_{\text{test}}} \sum_{i=1}^{n_{\text{test}}} \frac{|\mathcal{C}_{\hat{\lambda}}(X_i)|}{|\hat{Y}_i|}. \quad (6)$$

Statistical coverage being a random quantity, the empirical coverage is the evaluation of a realization of Eq. (4). The stretch ϕ tells, on average, how much larger the prediction sets are with respect to \hat{Y} (lower is better). We also report the average empirical quantile $\hat{\lambda}$, which indicates how many dilations were necessary to attain the specified coverage. For all the metrics, we report the average and standard deviation over 36 runs (i.e., shuffling and partitioning the data).

4.1 Results

The results in Table 1 show that, as expected, the empirical coverage is greater than the nominal value $1 - \alpha$ on average over multiple runs, that is, our CP procedure constructs statistically valid prediction sets ($\text{Cov} \geq 1 - \alpha$). As for the size of the prediction set, we see how the stretch and $\hat{\lambda}$ increase for higher α and τ , to compensate for the stricter requirements imposed by the user. In Table 1 we can also see how CP can be used to evaluate a model: for a given risk α and coverage ratio τ , the margin is small (e.g., 1.056 for WBC and UniverSeg at $\alpha = 0.1$ and $\tau = 0.9$) when the underlying predictor is already satisfying (empirically) the statistical requirements.

Finally, despite working in the restrictive setting without feedback from the predictor (no sigmoid scores), we show that there are cases where more information does not improve the prediction sets: for PraNet on the Polyps dataset, we also compute conformal sets using a threshold on the sigmoid (see Sec. 2, so that as it is lowered, more pixels are included. The average stretch across several configurations (Tab. 2) is considerably larger than with our method. In this model

⁵ Distributed as open source at: https://github.com/zxaoyou/segmentation_WBC. We acknowledge the *Jiangxi Tecom Science Corporation, China*, and the *CellaVision blog* (<http://blog.cellavision.com/>) for providing the data.

⁶ Source: sites.wustl.edu/oasisbrains, obtained via github.com/JJGO/UniverSeg.

⁷ We reused the precomputed predictions and dataset as partitioned (training, test) by the authors of [4]. See github.com/aangelopoulos/conformal-prediction.

Model	Dataset	$1 - \alpha$	τ	Cov	ϕ	avg $\hat{\lambda}$
PraNet	Polyps	0.9	0.9	0.909 (0.023) [†]	1.253 (0.131) [†]	6.083 (2.980) [†]
		0.9	0.99	0.898 (0.029)	1.780 (0.270)	17.500 (5.593)
		0.9	0.999	0.904 (0.023)	2.031 (0.309)	22.500 (5.969)
		0.95	0.9	0.953 (0.020)	3.093 (0.857)	41.389 (14.349)
		0.95	0.99	0.955 (0.019)	4.144 (0.971)	58.611 (14.946)
		0.95	0.999	0.962 (0.019)	4.840 (0.959)	69.167 (14.238)
UniverSeg	WBC	0.9	0.9	0.952 (0.044)	1.108 (0.065)	1.056 (0.630)
		0.9	0.99	0.932 (0.055)	1.646 (0.261)	6.278 (2.514)
		0.9	0.999	0.928 (0.061)	1.865 (0.284)	8.389 (2.686)
		0.95	0.9	0.984 (0.033)	1.449 (0.365)	4.361 (3.523)
		0.95	0.99	0.985 (0.024)	2.448 (0.664)	13.667 (5.826)
		0.95	0.999	0.985 (0.025)	2.776 (0.735)	16.528 (6.236)
UniverSeg	OASIS	0.9	0.9	0.969 (0.040)	1.774 (0.061)	4.500 (0.507)
		0.9	0.99	0.941 (0.049)	2.198 (0.034)	12.333 (0.756)
		0.9	0.999	0.940 (0.038)	2.247 (0.030)	15.417 (0.996)
		0.95	0.9	0.997 (0.007)	1.845 (0.040)	5.194 (0.401)
		0.95	0.99	0.990 (0.014)	2.232 (0.033)	14.222 (1.124)
		0.95	0.999	0.993 (0.014)	2.259 (0.028)	17.917 (0.937)

Table 1: Experiments with structuring element $B = \begin{bmatrix} \oplus & \oplus \\ \oplus & \oplus \end{bmatrix}$ averaged across 36 runs.
[†]: standard deviation.

and dataset configuration, our method performs better than approaches that use more information (e.g., sigmoid scores): not only it produces statistically valid sets, but it also avoids introducing artifacts in the conformal margin.

An example of this phenomenon is visible in Fig. 2d, where the conformalized margins are uninformative; possibly due to the type of training loss [5], the values of the sigmoid fail to serve as a proxy of uncertainty. On the other hand, in cases where distant zones are missed by the segmentation mask \hat{Y} , our method cannot recover those areas except by using a large margin. In this case, using the sigmoid (if available), may be a viable option; in a sense, our approach can be seen as complementary to other CP methods.

Model	Dataset	$1 - \alpha$	τ	$\phi_{\text{morphology}}$	$\phi_{\text{thresholding}}$
PraNet	Polyps	0.9	0.9	1.253	1.218
		0.9	0.99	1.780	8.950
		0.9	0.999	2.031	15.001
		0.95	0.9	3.093	3.010
		0.95	0.99	4.144	15.189
		0.95	0.999	4.840	16.062

Table 2: Comparison with the first rows in Tab. 1. Here, conformalization is done by morphological dilation and thresholding of the sigmoid [4] (see Sec. 2).

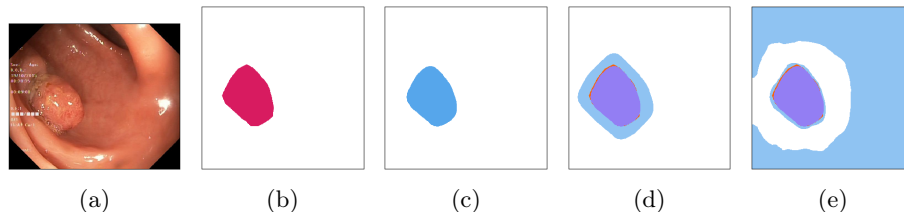


Fig. 2: Example: Polyps dataset (see Sec. 4), prediction with PraNet [14]. For $\alpha = 0.10$ and $\tau = 0.99$: (a) input image, (b) ground-truth mask, (c) predicted mask, (d) prediction set via dilation (Eq. 2), (e) prediction set via thresholding on sigmoid (as in [4]). Pixels in light blue (■) are the margin: in (d) it contains only pixels contiguous to the prediction (c) while in (e), it does not necessarily do so because of the underlying sigmoid scores (not shown). As shown in Tab. 2, for this model configuration the latter has much larger stretch (Eq. 6, lower is better). White pixels represent the background.

5 Conclusion

In this paper, we proposed an approach that combines a fundamental operation in mathematical morphology, dilation, with Conformal Prediction to construct statistically valid prediction sets for image segmentation. We achieved this in a restrictive framework with no internal knowledge of the model (e.g., sigmoid scores), where only the prediction masks are required. We managed to preserve the original shape of the prediction and to remain robust to aberrant scores from such models.

Although we applied our algorithm to several benchmarks in medical imaging, our method can be used with any segmentation method that returns binary masks (e.g., standard thresholding/clustering approaches or more advanced ML models). This includes ML models that lack transparency (black boxes) and whose details are hidden from the end users, as well as algorithms that were not originally conceived for uncertainty quantification.

Perspectives. A promising next step is to extend morphological sets to multiclass and instance segmentation, which are commonly used in the field of medical imaging. In their basic form, CP sets do not adapt to the input instance, and the theoretical guarantee holds on average: some images may be “harder” and require a larger margin, and vice versa. This is an active field of research in CP [15,9], and the literature on Mathematical Morphology could provide new tools to build adaptive morphological prediction sets using training data. Finally, we consider combining morphological sets with other approaches (e.g., thresholding) to leverage their respective strengths [35].

Acknowledgments. This work was carried out within the DEEL project,⁸ which is part of IRT Saint Exupéry and the ANITI AI cluster. The authors acknowledge the financial support from DEEL’s Industrial and Academic Members and the France 2030 program – Grant agreements n°ANR-10-AIRT-01 and n°ANR-23-IACL-0002.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Andéol, L., Fel, T., de Grancey, F., Mossina, L.: Confident object detection via conformal prediction and conformal risk control: an application to railway signaling. In: Proceedings of the Twelfth Symposium on Conformal and Probabilistic Prediction with Applications. vol. 204, pp. 36–55. PMLR (2023)
2. Andéol, L., Fel, T., de Grancey, F., Mossina, L.: Conformal prediction for trustworthy detection of railway signals. *AI and Ethics* **4**(1), 157–161 (Jan 2024). <https://doi.org/10.1007/s43681-023-00400-7>
3. Angelopoulos, A.N., Barber, R.F., Bates, S.: Theoretical foundations of conformal prediction. arXiv preprint arXiv:2411.11824 (2024)
4. Angelopoulos, A.N., Bates, S., Fisch, A., Lei, L., Schuster, T.: Conformal risk control. In: The Twelfth International Conference on Learning Representations, ICLR 2024, Vienna, Austria, May 7–11, 2024 (2024), <https://openreview.net/forum?id=33XGfHLtZg>
5. Azad, R., Heidary, M., Yilmaz, K., Hüttemann, M., Karimijafarbigloo, S., Wu, Y., Schmeink, A., Merhof, D.: Loss functions in the era of semantic segmentation: A survey and outlook. arXiv preprint arXiv:2312.05391 (2023)
6. Bates, S., Angelopoulos, A., Lei, L., Malik, J., Jordan, M.: Distribution-free, risk-controlling prediction sets. *J. ACM* **68**(6) (2021). <https://doi.org/10.1145/3478535>
7. Bereska, J.I., Karimi, H., Samavi, R.: SACP: Spatially-aware conformal prediction in uncertainty quantification of medical image segmentation. In: Medical Imaging with Deep Learning (2025), <https://openreview.net/forum?id=uQaPr1wU1W>
8. Bernal, J., Sánchez, F.J., Fernández-Esparrach, G., Gil, D., Rodríguez, C., Vilarinho, F.: Wm-dova maps for accurate polyp highlighting in colonoscopy: Validation vs. saliency maps from physicians. *Comp. Med. Imaging Graph.* **43**, 99–111 (2015)
9. Blot, V., Angelopoulos, A.N., Jordan, M., Brunel, N.J.B.: Automatically adaptive conformal risk control. In: The 28th International Conference on Artificial Intelligence and Statistics (2025), <https://openreview.net/forum?id=z046aYZqrG>
10. Blusseau, S., Puybureau, É.: Morphologie mathématique et traitement d’images. *Informatique Mathématique Une photographie en 2023* pp. Chapitre–3 (2023)
11. Butoi, V.I., Ortiz, J.J.G., Ma, T., Sabuncu, M.R., Gutttag, J., Dalca, A.V.: Universeg: Universal medical image segmentation. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). pp. 21438–21451 (2023)
12. Da Veiga, S.: Tutorial on conformal prediction and related methods - ETICS 2024 Research School (Sep 2024), <https://hal.science/hal-04690218>
13. Davenport, S.: Conformal confidence sets for biomedical image segmentation. arXiv preprint arXiv:2410.03406 (2024)

⁸ <https://www.deel.ai/>

14. Fan, D.P., Ji, G.P., Zhou, T., Chen, G., Fu, H., Shen, J., Shao, L.: Pranel: Parallel reverse attention network for polyp segmentation. In: Medical Image Computing and Computer Assisted Intervention – MICCAI 2020. pp. 263–273. Springer (2020)
15. Gibbs, I., Cherian, J.J., Candès, E.J.: Conformal prediction with conditional guarantees. *Journal of the Royal Statistical Society Series B: Statistical Methodology* p. qkaf008 (03 2025). <https://doi.org/10.1093/jrsssb/qkaf008>
16. Gonzalez, R.C., Woods, R.E.: Digital Image Processing. Pearson, 4 edn. (3 2017)
17. de Grancey, F., Adam, J.L., Alecu, L., Gerchinovitz, S., Mamalet, F., Vigouroux, D.: Object detection with probabilistic guarantees: A conformal prediction approach. In: SAFECOMP 2022 Workshops. pp. 316–329. Springer (2022)
18. Gupta, C., Kuchibhotla, A.K., Ramdas, A.: Nested conformal prediction and quantile out-of-bag ensemble methods. *Pattern Recognition* **127**, 108496 (2022)
19. Hoopes, A., Hoffmann, M., Greve, D.N., Fischl, B., Gutttag, J., Dalca, A.V.: Learning the effect of registration hyperparameters with hypermorph. *J Mach Learn Biomed Imaging* **1**, 1–30 (2022)
20. Hüllermeier, E., Waegeman, W.: Aleatoric and epistemic uncertainty in machine learning: an introduction to concepts and methods. *Mach. Learn.* **110**(3), 457–506 (2021)
21. Lambert, B., Forbes, F., Doyle, S., Dehaene, H., Dojat, M.: Trustworthy clinical ai solutions: a unified review of uncertainty quantification in deep learning models for medical image analysis. *Artificial Intelligence in Medicine* p. 102830 (2024)
22. Li, S., Park, S., Ji, X., Lee, I., Bastani, O.: Towards pac multi-object detection and tracking. *arXiv preprint arXiv:2204.07482* (2022)
23. Liu, K., Sun, T., Zeng, H., Zhang, Y., Pun, C.M., Vong, C.M.: Spatial-aware conformal prediction for trustworthy hyperspectral image classification. *IEEE Transactions on Circuits and Systems for Video Technology* pp. 1–1 (2025). <https://doi.org/10.1109/TCSVT.2025.3558753>
24. Ma, J., He, Y., Li, F., Han, L., You, C., Wang, B.: Segment anything in medical images. *Nature Communications* **15**(1), 654 (2024)
25. Marcus, D.S., Wang, T.H., Parker, J., Csernansky, J.G., Morris, J.C., Buckner, R.L.: Open access series of imaging studies (oasis): cross-sectional mri data in young, middle aged, nondemented, and demented older adults. *Journal of cognitive neuroscience* **19**(9), 1498–1507 (2007)
26. Matheron, F.: Random sets and integral geometry. John Wiley & Sons (2 1975)
27. Mossina, L., Dalmau, J., Andéol, L.: Conformal semantic image segmentation: Post-hoc quantification of predictive uncertainty. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops. pp. 3574–3584 (June 2024)
28. Papadopoulos, H., Proedrou, K., Vovk, V., Gammerman, A.: Inductive confidence machines for regression. In: Machine Learning: ECML 2002 (2002)
29. Pogorelov, K., Randel, K.R., Griwodz, C., Eskeland, S.L., de Lange, T., Johansen, D., Spampinato, C., Dang-Nguyen, D.T., Lux, M., Schmidt, P.T., Riegler, M., Halvorsen, P.: Kvasir: A multi-class image dataset for computer aided gastrointestinal disease detection. In: Proceedings of ACM MMSys. p. 164–169 (2017)
30. Schmitt, M., Mattioli, J.: Morphologie mathématique. Presses des MINES (2013)
31. Serra, J.: Image analysis and mathematical morphology: V.1. Image Analysis & Mathematical Morphology Series, Academic Press, San Diego, CA (Jan 1984)
32. Silva, J., Histace, A., Romain, O., Dray, X., Granado, B.: Toward embedded detection of polyps in wce images for early diagnosis of colorectal cancer. *International Journal of Computer Assisted Radiology and Surgery* **9**(2), 283–293 (Sep 2013). <https://doi.org/10.1007/s11548-013-0926-3>

33. Soille, P.: *Morphological Image Analysis*. Springer (2004)
34. Tajbakhsh, N., Gurudu, S.R., Liang, J.: Automated polyp detection in colonoscopy videos using shape and context information. *IEEE T-MI* **35**(2), 630–644 (2015)
35. Teneggi, J., Tivnan, M., Stayman, W., Sulam, J.: How to trust your diffusion model: A convex optimization approach to conformal risk control. In: *Proceedings of the 40th International Conference on Machine Learning. Proceedings of Machine Learning Research*, vol. 202, pp. 33940–33960. PMLR (23–29 Jul 2023)
36. Timans, A., Straehle, C.N., Sakmann, K., Nalisnick, E.: Adaptive bounding box uncertainties via two-step conformal prediction. In: *Computer Vision – ECCV 2024*. pp. 363–398. Springer Nature Switzerland, Cham (2025)
37. Vázquez, D., Bernal, J., Sánchez, F.J., Fernández-Esparrach, G., López, A.M., Romero, A., Drozdal, M., Courville, A.: A benchmark for endoluminal scene segmentation of colonoscopy images. *J Healthc Eng* **2017**(1), 4037190 (2017)
38. Vovk, V., Gammerman, A., Shafer, G.: *Algorithmic learning in a random world*, vol. 29. Springer (2005)
39. Zheng, X., Wang, Y., Wang, G., Liu, J.: Fast and robust segmentation of white blood cell images by self-supervised learning. *Micron* **107**, 55–71 (2018)