# Uncertainty-Aware Meta-Learning for Robust Stent Segmentation via Selective Ensemble

Hyojin Ko[1], Jihoon Kweon[1], Jung-Min Ahn[1], Yong-Hak Kim[1]

Asan Medical Center, University of Ulsan College of Medicine
Seoul, Republic of Korea

## Abstract

*Assessing stent expansion on fluoroscopic images during Percutaneous Coronary Intervention (PCI) requires coronary stent segmentation, but achieving consistent performance with a single model is challenging due to low stent visibility and morphological variability.*

*We propose a selective ensemble of five EfficientNet-B7/Feature Pyramid Network (FPN) base models, each trained with BCEDice (binary cross-entropy + Dice) plus one auxiliary loss (clDice, Hausdorff distance, perimeter regularity, TopK, or Wasserstein topological homology) and search-optimized weights. A Convolutional Block Attention Module (CBAM) meta-learner integrates the fluoroscopic image, base model probability maps, and pixel-wise uncertainty through two-phase label-free training (random-fit and nearest-neighbor-fit). Internal five-fold cross-validation (942 frames, 826 patients) achieved a Dice Similarity Coefficient (DSC) of 90.43%, exceeding the best-performing individual base model and hard voting by 1.08 and 0.21 percentage points, respectively. External validation (361 frames, 212 patients) reached a DSC of 83.55%, up to 3.68 percentage points above the lowest-performing base model.*

*This uncertainty-aware framework enables robust stent segmentation without additional annotation, supporting automated stent analysis during PCI.*

## 1. Introduction

Percutaneous Coronary Intervention (PCI) is widely used to relieve myocardial ischemia from obstructive coronary artery disease, but stent underexpansion remains common, particularly in calcified lesions where rigid calcium limits lesion preparation and stent expansion [1]. Intravascular ultrasound or optical coherence tomography can guide stent implantation in complex PCI, but routine use remains limited by cost and the need for dedicated equipment and trained personnel [1].

To facilitate angiography-based assessment of stent expansion, dedicated stent-enhancement techniques such as StentBoost [2] have been developed, though these require dedicated X-ray acquisitions and increase radiation exposure [3]. More recently, a deep learning-based ensemble of segmentation models delineated coronary stents on fluoroscopic images and applied AI-based Quantitative Coronary Angiography (AI-QCA) to detect stent underexpansion using already-acquired images [4].

Quantitative assessment of stent expansion therefore depends on robust ensemble-based segmentation. In coronary artery segmentation, selective ensemble methods that adaptively weight models by morphological features or predicted quality have outperformed hard voting [5], suggesting applicability to other coronary X-ray targets. However, stent segmentation has distinct challenges because stents are less visible on fluoroscopy than contrast-filled coronary arteries and consist of thin struts in a grid-like pattern. Complex configurations, including overlapping and bifurcation stenting, further increase variability, making it difficult to establish generalizable weighting criteria.

To address these challenges, we developed a selective ensemble framework that integrates model predictions via pixel-wise uncertainty, requiring no additional meta-learner annotations. We hypothesized that the proposed framework would outperform conventional ensemble methods in fluoroscopic stent segmentation. Performance and generalizability were evaluated using patient-level cross-validation and an independent external cohort.

## 2. Methods

### 2.1. Data Acquisition

The internal dataset comprised post-procedural fluoroscopic sequences from patients who underwent PCI at Asan Medical Center between February and December 2016. A total of 942 frames from 826 patients were extracted from fluoroscopic sequences targeting the three major coronary arteries, the Right Coronary Artery (RCA), Left Anterior Descending artery (LAD), and Left Circumflex artery (LCX). Interventional cardiologists with over 10 years of experience selected an end-diastolic frame and manually annotated the ground-truth stent masks. For external validation, 361 fluoroscopic frames were obtained from 212 patients at Chungnam National University Hospital between 2016 and 2018. The internal dataset was divided into five patient-level folds with no overlap. The entire external cohort served as a fixed hold-out set, evaluated independently by each of the five internally trained models.

## 2.2. Preprocessing

Fluoroscopic frames were min-max normalized to [0,1] per image, rescaled to 8-bit for augmentation, and returned to [0,1] before network input. Augmentation was applied using Albumentations [6]. Geometric augmentation, applied identically to images and masks with nearest-neighbor mask interpolation: shift-scale-rotate (shift/scale ±10%, rotation ±20°, p=1), elastic transform (p=0.2), and grid distortion (p=0.2). Intensity augmentation comprised random brightness/contrast adjustment (brightness -0.2 to +0.3, contrast ±0.2, p=0.5), Gaussian noise (variance 0 to 0.1, p=0.5), and CLAHE (clip limit=4, p=0.2).

## 2.3. Base Model Training

Each base model combined an EfficientNet-B7 encoder [7] with a Feature Pyramid Network (FPN) decoder [8], initialized with ImageNet-pretrained weights. Five combined losses were defined by adding one auxiliary term to BCEDice: centerline Dice loss (clDice) [9], Hausdorff distance loss [10], perimeter regularity loss [11], TopK loss [12], and Wasserstein topological homology loss [13].

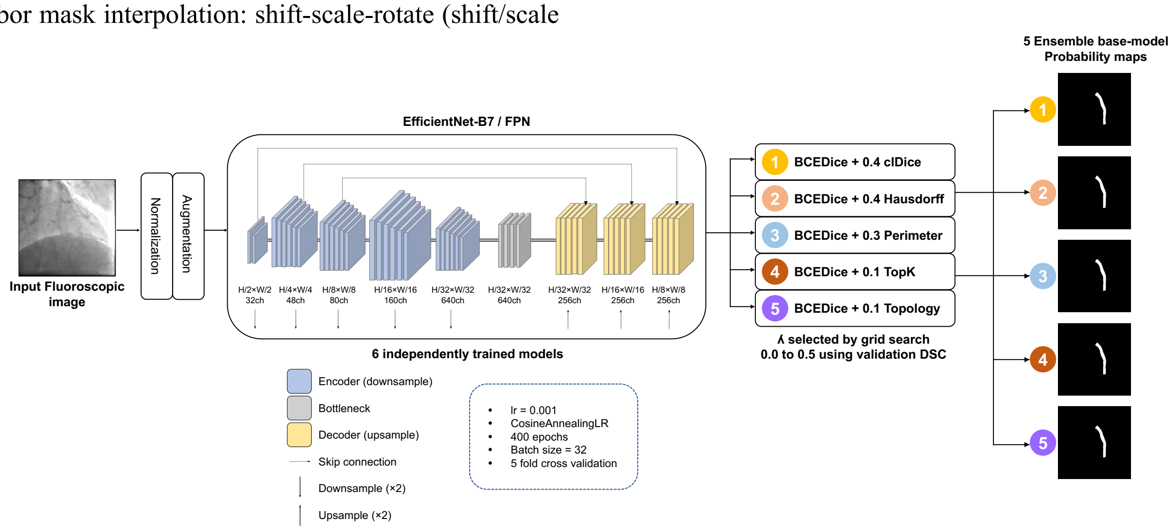



Figure 1. Training process of five base models with five BCEDice-based loss functions.

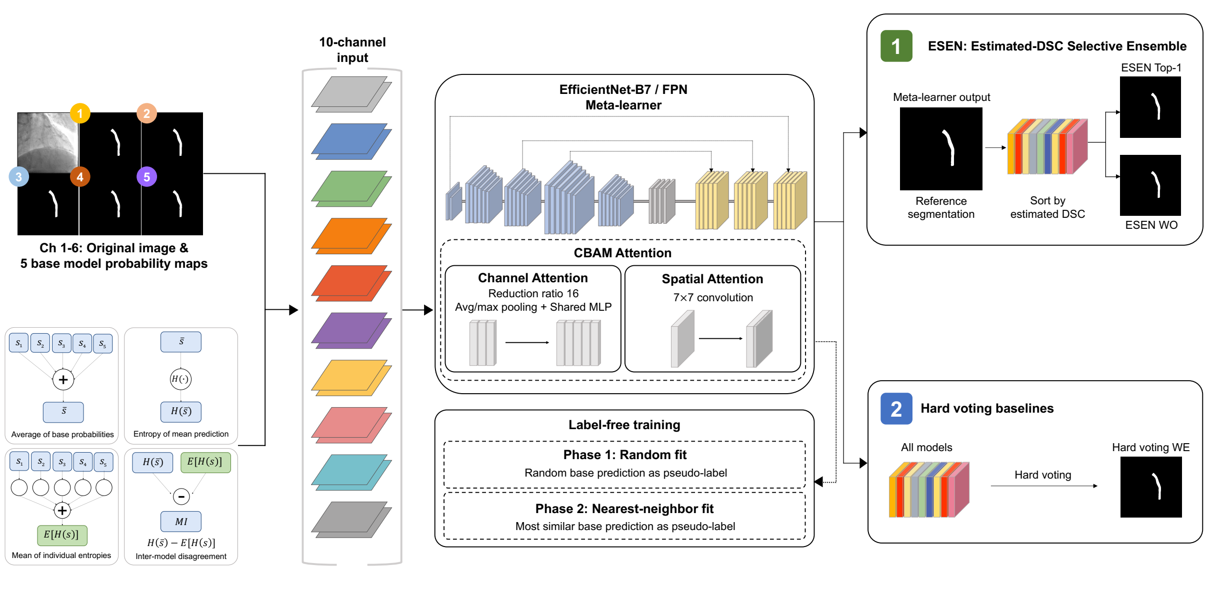



Figure 2. Label-free CBAM meta-learner training and selective ensemble inference

Each combined loss was defined as:

$$\mathcal{L}_{compound} = \mathcal{L}_{BCEDice} + \lambda \cdot \mathcal{L}_{aux} \quad (1)$$

where $\lambda$ is the auxiliary term weight, selected by grid search from 0.0 to 0.5 before cross-validation based on validation DSC. The resulting $\lambda$ values were 0.4 for clDice, 0.4 for Hausdorff distance, 0.3 for perimeter regularity, 0.1 for TopK, and 0.1 for Wasserstein topological homology. The BCEDice-only model was used as a performance reference.

All models were trained for 400 epochs with a batch size of 32 using the Adam optimizer, an initial learning rate of $10^{-3}$, and a CosineAnnealingLR scheduler. For cross-validation, data were divided at the patient level into training, validation, and test sets (3:1:1), with assignments permuted across the five folds.

## 2.4. CBAM-Based Uncertainty Aware Meta-Learner and Selective Ensemble

The meta-learner shared the base models' FPN/EfficientNet-B7 backbone, with a CBAM module [14] applied to its 10-channel input stack: the fluoroscopic image, five base model probability maps $s_l$ ($l = 1,\ldots, 5$), and four auxiliary maps comprising the mean probability map and three pixel-wise uncertainty maps.

The four auxiliary channels were defined as the mean probability $\bar{s} = \frac{1}{L}\sum_{l=1}^{L} s_l$, the predictive entropy $H(\bar{s}) = -\bar{s}\,log(\bar{s}) - (1-\bar{s})log(1-\bar{s})$, the expected entropy $E[H(s)] = \frac{1}{L}\sum_l H(s_l)$, and the mutual information $MI = H(\bar{s}) - E[H(s)]$, each scaled to [0, 1]. Channel attention (reduction ratio 16, average- and max-pooling with a shared multilayer perceptron) preceded spatial attention (7×7 convolution).

The meta-learner was trained in two label-free phases using the same fold splits as the frozen base models. In the first phase (50 epochs), the target was the mean of the five probability maps. In the second phase (30 epochs), the target for each image was the probability map of the base model most similar to the current meta-learner output, with similarity assessed by soft Dice and a small random perturbation. BCEDice loss was minimized using the Adam optimizer (learning rate $5 \times 10^{-4}$, polynomial decay) and a batch size of 12. Validation masks were used only for final checkpoint selection.

Base models were then ranked for each image following the Estimated-DSC Selective Ensemble (ESEN) of Park et al. [5], using the meta-learner output as a pseudo-reference segmentation to estimate the DSC of each base-model prediction and determine its rank. Top-1 retained only the highest-ranked prediction, WO averaged all five predictions using per-image softmax weights from these scores, and hard voting (WE) used equal weights.

Post-processing was applied uniformly to all final predictions, comprising binary hole-filling, Gaussian smoothing with a 15×15 kernel, and removal of connected components with less than 800 pixels.

## 2.5. Evaluation Methods

Five baseline models and five ensemble/meta-learner strategies were evaluated using five-fold cross-validation on the internal dataset, and their performance was assessed on the external cohort.

# 3. Results

## 3.1. Internal Validation

Across the five internal folds (942 frames, 826 patients), the proposed method (WO aggregation) achieved a mean DSC of 90.43% (Table 1), compared with 89.35% for the best-performing individual base model (+1.08 percentage points) and 90.22% for hard voting (+0.21 percentage points). The percentile-DSC curves (Fig. 3(a)) show the proposed method's curve lying above every baseline across nearly the entire distribution, with the margin widening in the lower-percentile range. Fig. 4 shows two representative internal cases with segmentation overlays for five representative methods, in which the proposed method attained the highest DSC in both cases.

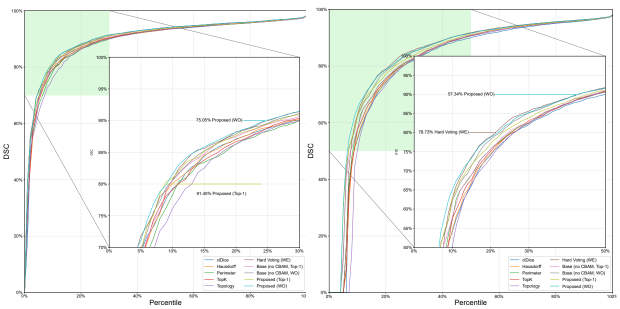
## 3.2. External Validation

On an external cohort comprising 361 frames from 212 patients, the proposed method achieved a DSC of 83.55% (Table 1), 3.68 percentage points higher than the lowest-performing individual base model. The performance gap between the proposed method and the baselines was more pronounced on the external cohort than on the internal dataset (Fig. 3(b)), and the hard voting curve was below the proposed method in the low-to-mid percentile range.

## 3.3. Ablation and Reproducibility

Ablation studies assessed the individual contributions of CBAM attention and the aggregation method on both datasets (Table 1). CBAM attention improved DSC regardless of aggregation strategy, with DSC increasing from 89.75%/89.96% to 90.00%/90.43% under Top-1/WO aggregation internally (+0.25/+0.47 percentage points), and from 81.99%/82.98% to 82.74%/83.55% externally (+0.75/+0.57 percentage points).

WO aggregation also improved DSC over Top-1 aggregation regardless of whether CBAM attention was used, and the combination of CBAM and WO aggregation achieved the best overall performance. The proposed method demonstrated narrow variability across the five permutations, ranging from 89.97% to 91.14% internally and from 83.33% to 83.81% externally, showing limited variation across fold assignments.




(a). Internal dataset (b). External dataset

Figure 3. Cumulative distribution of per-case DSC of internal (a) and external (b) datasets.

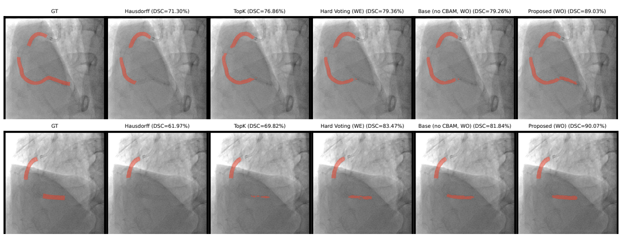



Figure 4. Representative segmentation examples from the internal cohort for five representative methods.

Table 1. Segmentation Performance (DSC, %).

| Method | Internal | External |
|---|---|---|
| clDice | 89.21 | 80.48 |
| Hausdorff | 89.32 | 81.01 |
| Perimeter | 89.24 | 81.44 |
| TopK | 89.35 | 81.05 |
| Topology | 88.79 | 79.87 |
| Hard Voting (WE) | 90.22 | 82.90 |
| Base (no CBAM, Top-1) | 89.75 | 81.99 |
| Base (no CBAM, WO) | 89.96 | 82.98 |
| Proposed (Top-1) | 90.00 | 82.74 |
| Proposed (WO) | 90.43 | 83.55 |

## 4. Discussion and Conclusion

The uncertainty-aware selective ensemble proposed in this study, integrated via a label-free meta-learner, outperformed every comparator configuration evaluated (Table 1), with the largest gains observed under external validation, indicating that uncertainty-aware attention is particularly effective under distribution shift. Limitations include external validation at a single additional institution and the computational cost of running all five base models plus the meta-learner at inference. These results support the feasibility of this label-free, robust framework for automated stent segmentation, warranting future prospective multi-center validation.

## References


[1] C. Vrints et al., “2024 ESC Guidelines for the management of chronic coronary syndromes,” Eur.Heart J., vol. 45, pp. 3415-3537, Oct.2024.

[2] F. Cura et al., “StentBoost visualization for the evaluation of coronary stent expansion during percutaneous coronary interventions,” Cardiol. Ther., vol. 2, no. 2, pp. 171-180, Dec. 2013.

[3] Z. Jin et al., “Impact of StentBoost subtract imaging on patient radiation exposure during percutaneous coronary intervention,” Int. J. Cardiovasc. Imaging, vol. 29, pp. 1207-1213, Aug. 2013.

[4] H. Lee et al., “TCT-191 Fully Automated Evaluation of Coronary Stent Underexpansion Using Coronary Angiography-Based Deep Learning,” J. Am. Coll. Cardiol., vol. 84, no. 18_Suppl, pp. B12, Oct. 2024.

[5] J. Park et al., “Selective ensemble methods for deep learning segmentation of major vessels in invasive coronary angiography,” Med. Phys., vol. 50, no. 12, pp. 7822–7839, Dec. 2023.

[6] A. Buslaev et al., “Albumentations: fast and flexible image augmentations,” *Information*, vol. 11, no. 2, p. 125, Feb. 2020.

[7] M. Tan and Q. Le, “Efficientnet: Rethinking model scaling for convolutional neural networks,” in Proc. Int. Conf. Mach. Learn. (ICML), 2019, pp. 6105–6114.

[8] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, “Feature pyramid networks for object detection,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 2017, pp. 2117–2125.

[9] S. Shit et al., “clDice-a novel topology-preserving loss function for tubular structure segmentation,” in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), 2021, pp. 16560–16569.

[10] D. Karimi and S. E. Salcudean, “Reducing the Hausdorff distance in medical image segmentation with convolutional neural networks,” IEEE Trans. Med. Imaging, vol. 39, no. 2, pp. 499–513, Feb. 2020.

[11] R. E. L. Jurdi, C. Petitjean, P. Honeine, V. Cheplygina, and F. Abdallah, “A surprisingly effective perimeter-based loss for medical image segmentation,” in Proc. Med. Imaging Deep Learn. (MIDL), 2021, pp. 158–167.

[12] Z. Wu, C. Shen, and A. V. D. Hengel, “Bridging category-level and instance-level semantic image segmentation,” arXiv preprint arXiv:1605.06885, 2016.

[13] S. Ozcelik, S. Unver, I. A. Gurses, R. Turkay, and C. Gunduz-Demir, “Topology-aware loss for segmentation in computed tomography images,” Biomed. Signal Process. Control, vol. 117, p. 109512, 2026.

[14] S. Woo, J. Park, J.-Young Lee, and I. S. Kweon, “Cbam: Convolutional block attention module,” in Proc. Eur. Conf. Comput. Vis. (ECCV), 2018, pp. 3–19.



Address for correspondence:

Jihoon Kweon.
Department of Biomedical Engineering, Asan Medical Center, University of Ulsan College of Medicine, 88 Olympic-ro, 43-Gil, Songpa-gu, Seoul, 05505, Republic of Korea
kjihoon2@naver.com