CROSS-IMAGE PIXEL CLASSIFICATION OF BIOMEDICAL MULTI-EXCITATION HYPERSPECTRAL IMAGES

Naira Matosyan[1], *Fernando Villarruel*[2],
Narine Sarvazyan[1,2,3], *Varduhi Yeghiazaryan*[1], *and Aram Butavyan*[1]

[1]Akian College of Science and Engineering, American University of Armenia, Yerevan, Armenia
[2]L. A. Orbeli Institute of Physiology NAS RA, Yerevan, Armenia
[3]George Washington University, Washington, DC, United States

ABSTRACT

Cross-image pixel classification of multi-excitation hyperspectral imaging (ME-HSI) data is essential for deploying deep learning models to new tissue samples without requiring additional manual annotations. We present a progressive self-training approach for classifying ablated regions in ME-HSI of cardiac tissue, addressing the significant domain shift between source and target images. Our method combines confidence-based pseudo-label selection, using margin-based filtering, per-class pruning, and progressive budget expansion with a minimum class ratio constraint derived from source statistics to prevent class collapse. Evaluated on six target images from an ME-HSI dataset of radiofrequency ablated bovine heart atria, our approach improves average overall accuracy from 0.75 to 0.85 and Cohen's κ from 0.26 to 0.35 compared to the baseline model without adaptation. For images with poor baseline performance, improvements exceed 30% in overall accuracy. In one sample, self-training surpasses same-image-supervised performance, demonstrating that effective domain adaptation can outperform limited target-domain supervision.

Index Terms— Hyperspectral imaging, domain adaptation, deep learning, self-training, pseudo-labeling, atrial tissue classification

1. INTRODUCTION

Atrial fibrillation (AF) is the most prevalent cardiac arrhythmia among adults, associated with significant cardiovascular complications [1]. Radiofrequency ablation (RFA) using a catheter is an effective minimally invasive treatment that uses heat generated by high-frequency electrical currents to isolate the pulmonary vein, thus eliminating the abnormal electrical activity. However, accurately identifying the extent of tissue damage during the procedure remains a major challenge. Postoperative AF recurrence is often attributed to viability gaps—intact myocardial regions between ablated zones that continue to conduct electrical impulses [2]. This limitation drives the need for real-time, non-invasive imaging modalities capable of visualizing ablation effects with high specificity and spatial resolution.

Multi-excitation hyperspectral imaging (ME-HSI) captures rich spectral information across multiple excitation wavelengths, producing 4D data structures that enhance tissue differentiation compared to traditional single-excitation hyperspectral imaging (HSI) [3]. Recent work has demonstrated the potential of deep learning methods for classifying ablated regions in ME-HSI data of atrial tissue [4, 5]. However, these approaches assume that training and test data come from the same distribution—an assumption that often fails in practice when models trained on one tissue sample are applied to new, unseen samples.

Cross-image ME-HSI classification presents significant challenges due to domain shift between source and target images. When a model trained on labeled source images is deployed on unlabeled target images, performance often degrades substantially due to differences in tissue properties, imaging conditions, and spectral characteristics. Domain adaptation methods aim to bridge this gap by aligning feature distributions across domains, typically through pseudo-labeling strategies where confident predictions on target data are used to iteratively refine the model.

Several recent approaches explore confident learning for pseudo-label quality assessment in HSI domain adaptation. [6] proposes the CLDA method, which combines adversarial domain adaptation with confident learning to select high-confidence target samples and iteratively refine pseudo-labels. Their joint distribution matrix estimates the relationship between pseudo-labels and true labels to guide per-class pruning. To prevent error accumulation from incorrect pseudo-labels, [7] introduces a dual classification head strategy in DHSNet, where one head generates pseudo-labels while another is trained on them, ensuring that pseudo-label errors do not directly propagate to the main classifier. The importance of target domain class distribution is addressed

Thanks to the Afeyan Family Foundation, RA HESC (24PostDoc/2-2G002), and the EU (ERA-CHAIR NAS-SAR 101087403) for funding.

in PCDM-UDA [8], which uses pseudo-class distribution constraints to correct pseudo-labels through zero-one programming optimization, while also introducing frequency domain phase features as domain-stable representations. More recently, large vision models have been explored for pseudo-label generation; [9] proposes the IPG framework, which utilizes the Segment Anything Model with a spectral voting strategy for iteratively refining pseudo-labels.

In this work, we address the cross-image classification problem for ME-HSI of ablated atrial tissue through progressive self-training with confidence-based pseudo-label selection. Our approach builds upon the LeViT architecture [10], which demonstrated strong performance for this tissue classification task [4], and extends it with a domain adaptation strategy that progressively increases the pseudo-label budget across multiple self-training rounds. We incorporate class-balanced constraints derived from source domain statistics to prevent class collapse, a common failure mode in pseudo-label-based methods where the model becomes increasingly biased toward the majority class. We evaluate our approach on a dataset of eight bovine left atrial tissue samples imaged with ME-HSI at four excitation wavelengths, demonstrating improved cross-image generalization compared to the baseline model.

2. DATASET

The potential of HSI to reveal tissue damage during RFA procedures in atrial heart tissue was first explored in [11]. Building on this work, [12] presented a comparative analysis of deep learning classification methods for identifying ablated regions in atrial tissue HSI, revealing that many architectures known to perform well on common HSI datasets yielded subpar results on this data, likely due to substantially smaller training sets and higher target complexity. More recently, [3] introduced the ME-HSI approach for enhanced tissue differentiation resulting in improved target identification.

In this work, we use the ME-HSI dataset of RFA lesions on bovine left atrial tissue introduced in [4], comprising eight samples from three different animals. The data were acquired using an HSI setup consisting of a 300 W xenon lamp coupled to four bandpass filters centered at 360, 370, 380, and 390 nm (10 nm FBHW) as illumination source, and a Nuance FX imaging system as a detector. For each excitation, 31 emission wavelengths were recorded (420–720 nm with 10 nm step), creating a 4D hyperspectral data structure with dimensions H $\times$ W $\times$ 31 $\times$ 4. Ground truth masks were generated by lesions' diameter determination using TTC (triphenyltetrazolium chloride) staining of histological sections. The inner 70% of each lesion diameter was labeled as 'ablated' and the outer 30% as 'unsure', to account for tissue slicing effects. Background pixels were identified by filtering HSI data based on signal intensity and excluded from analysis. The complete description of ground truth generation can be found in [4].

For cross-image evaluation, we adopt a leave-two-out protocol: two images Animal-2, Sample-3 (A2-S3) and Animal-3, Sample-2 (A3-S2) serve as labeled source data for training, while the remaining six images are used as unlabeled target domains for adaptation and evaluation. This setup simulates realistic deployment scenarios where a model trained on a limited set of labeled samples must generalize to new, unseen tissue samples. Prior work [4, 5] demonstrated that ME-HSI configurations using data from all four excitations show improved classification performance over individual HSI cubes, with the concatenated 4D setup achieving best results when sufficient training data is available.

3. METHODOLOGY

Building on findings from our previous work on ME-HSI classification [4, 5], we address the challenge of cross-image classification where models trained on labeled source images must generalize to new, unlabeled target images. Our approach combines a spatial–spectral deep learning architecture with progressive self-training for unsupervised domain adaptation.

We use the LeViT architecture [10], a vision transformer that merges convolutional and attention-based designs. LeViT demonstrated promising results in our prior experiments [4] by effectively utilizing both spatial context through convolutional patch embeddings and long-range dependencies through self-attention mechanisms. For the 4D ME-HSI data, we retain the four excitation cubes as separate 31-band inputs rather than concatenating them along the spectral dimension. We use a late-fusion LeViT model with an independent LeViT branch for each excitation; each branch produces class logits from spatial patches centered at each pixel. The excitation-wise logits are then normalized and fused by a class-wise weighted average over the four excitations, allowing the contribution of each excitation to vary by class. This enables pixel-wise classification while incorporating neighborhood context and excitation-specific cues.

For domain adaptation, we implement a pseudo-labeling strategy inspired by confident learning principles [6]. Given model predictions on unlabeled target images, we select high-confidence samples as pseudo-labels using the following criteria.

Margin-based confidence: For each pixel, we compute the margin as the absolute difference between the two foreground class logits: $m = |l_{\text{ablated}} - l_{\text{non-ablated}}|$. Pixels with margin below a threshold $\tau = 0.5$ are excluded, as low-margin predictions indicate classifier uncertainty.

Per-class pruning: Following CLDA [6], we apply per-class pruning where the lowest-confidence 20% of samples within each predicted class are discarded. This prevents confident but potentially incorrect predictions from dominating the pseudo-label set.

Table 1. Classification performance across progressive self-training rounds for six ME-HSI images. Same-image uses the 5-click circle protocol (split 0) and is reported for reference only. OA: Overall Accuracy, AA: Average Accuracy, κ: Cohen's Kappa, Sens: Sensitivity (ablated recall). Best results per image among Baseline and self-training rounds are shown in **bold**.

	Same-image				Baseline				Round 1				Round 2				Round 3				Round 4				Round 5			
Image	OA	AA	κ	Sens	OA	AA	κ	Sens	OA	AA	κ	Sens	OA	AA	κ	Sens	OA	AA	κ	Sens	OA	AA	κ	Sens	OA	AA	κ	Sens
A1-S1	0.91	0.93	0.45	0.96	**0.94**	0.49	−0.01	0.00	0.62	0.32	−0.09	0.00	0.54	0.74	0.08	**0.98**	0.92	0.82	**0.41**	0.72	0.87	**0.90**	0.35	0.94	0.14	0.54	0.01	**0.98**
A2-S1	0.58	0.68	0.16	0.81	**0.87**	**0.70**	**0.40**	**0.49**	0.86	0.69	0.38	0.46	0.85	0.66	0.33	0.41	**0.87**	0.67	0.36	0.40	**0.87**	0.61	0.28	0.28	**0.87**	0.58	0.22	0.21
A2-S2	0.85	0.82	0.32	0.78	0.51	**0.74**	0.10	**0.99**	0.52	0.72	0.09	0.95	0.59	0.69	0.09	0.81	0.72	0.73	0.15	0.74	**0.85**	0.62	0.15	0.35	**0.85**	0.63	**0.17**	0.38
A2-S4	0.88	0.89	0.45	0.91	0.90	0.66	0.29	0.39	0.85	0.66	0.22	**0.44**	0.90	**0.68**	0.31	0.42	0.92	**0.68**	0.37	0.40	0.91	**0.68**	0.34	0.41	**0.93**	**0.68**	**0.39**	0.40
A3-S1	0.91	0.92	0.58	0.93	0.95	0.92	0.71	0.89	0.93	0.93	0.63	0.93	0.81	0.88	0.37	**0.97**	0.93	0.93	0.64	0.92	**0.98**	**0.95**	0.84	0.91	**0.98**	**0.95**	**0.86**	0.93
A3-S3	0.63	0.77	0.19	0.94	0.31	0.62	0.05	**1.00**	0.27	0.60	0.04	**1.00**	0.29	0.61	0.05	**1.00**	0.33	0.63	0.06	**1.00**	**0.64**	**0.68**	**0.14**	0.74	0.61	**0.68**	0.13	0.77
Avg	0.79	0.84	0.36	0.89	0.75	0.69	0.26	0.63	0.68	0.65	0.21	0.63	0.66	0.71	0.21	**0.77**	0.78	**0.74**	0.33	0.70	**0.85**	**0.74**	**0.35**	0.61	0.73	0.68	0.30	0.61

Budget constraints: We control the proportion of pseudo-labels selected from each image using a budget parameter β, which specifies the fraction of target pixels to include. This relates to the class distribution considerations in PCDM-UDA [8].

Minimum class ratio: To prevent class collapse—where the model becomes increasingly biased toward the majority class—we enforce a minimum class ratio constraint derived from the labeled source (gold) images. Specifically, we then require that at least 13% of selected pseudo-labels be ablated; if natural confidence-based selection falls short, we force-select additional ablated pixels with the highest class-specific confidence until this floor is met.

Our self-training protocol works over multiple rounds with progressively increasing pseudo-label budgets, inspired by iterative refinement concepts from IPG [9] and gradual adaptation in CLDA [6]. Critically, we decouple the *teacher* used for pseudo-label generation from the *student* initialization used for fine-tuning: each round generates improved pseudo-labels from the previous round's adapted model, but always fine-tunes from the original source-trained baseline. This mitigates catastrophic forgetting of source knowledge while still benefiting from improved target pseudo-labels. Fine-tuning is performed on target pseudo-labels only, using class-balanced focal loss to handle class imbalance. The procedure is as follows.

1. **Initialization:** Train the base model on labeled source images to obtain the baseline checkpoint θ_0.
2. **Pseudo-label generation:** Apply the current teacher model $\theta_{r-1}^{\text{teach}}$ (baseline for $r = 1$; otherwise the model from round $r - 1$) to each target image and select pseudo-labels using the confidence-based criteria with budget β_r for round r.
3. **Model update:** Initialize from the fixed baseline θ_0 and fine-tune on the selected target pseudo-labels only, employing class-balanced focal loss.
4. **Iteration:** Set the teacher for the next round to the adapted model from step 3, then repeat steps 2–3 while increasing the budget ($\beta_r = r\%$ for rounds $r = 1, \ldots, 5$).

4. RESULTS

We evaluate our progressive self-training approach on six target images from the ME-HSI dataset, using images A2-S3 and A3-S2 as labeled source data. Table 1 presents the classification performance metrics for same-image inference, the baseline model, and five self-training rounds.

The baseline model, trained solely on source images without any target-domain adaptation, shows substantial variation in cross-image generalization. Image A3-S1 achieves excellent baseline performance ($\text{OA} = 0.95$, $\kappa = 0.71$), indicating strong domain similarity with the source data. In contrast, image A3-S3 shows poor baseline performance ($\text{OA} = 0.31$, $\kappa = 0.05$) with sensitivity of 1.00, indicating the model over-predicts the ablated class. Image A1-S1 demonstrates the opposite failure mode—high OA (0.94) but near-zero κ and sensitivity, indicating the baseline model predicts almost exclusively non-ablated tissue and fails to detect ablated regions entirely.

On average, self-training yields substantial improvements: OA increases from 0.75 (baseline) to 0.85 (Round 4), representing a 10 pp gain. Cohen's κ improves from 0.26 to 0.35, a relative increase of 35% that indicates meaningfully better agreement beyond chance.

The largest improvements occur for images with poor baseline performance. Image A2-S2 improves from $\text{OA} = 0.51$ to $\text{OA} = 0.85$ (+34 pp) by Round 4, as the model corrects its initial over-prediction of ablated tissue (baseline $\text{Sens} = 0.99$) toward better class balance. Similarly, image A3-S3 improves from $\text{OA} = 0.31$ to $\text{OA} = 0.64$ (+33 pp), with the model learning to reduce false positive ablated predictions while maintaining reasonable sensitivity.

Image A3-S1, which already performs well at baseline, shows continued improvement through self-training, with κ increasing from 0.71 to 0.86 by Round 5. Image A2-S4 demonstrates steady gains, with κ improving from 0.29 to 0.39 (+34% relative improvement).

The challenging image A1-S1, where the baseline model completely fails to detect ablated tissue ($\text{Sens} = 0$), shows gradual improvement through self-training. By Round 2, sensitivity reaches 0.98 and κ becomes positive (0.08), indicating

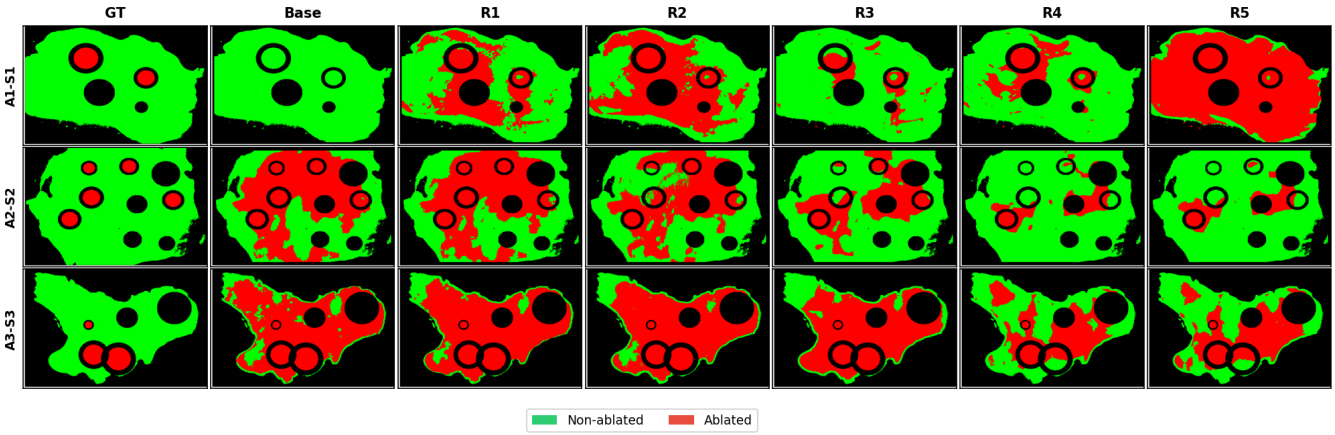


Fig. 1. Classification predictions across progressive self-training rounds for three images. First column shows ground truth (GT), followed by baseline predictions and Rounds 1–5. Green indicates non-ablated tissue, red indicates ablated tissue, and black indicates unsure and background regions excluded from analysis.

the model begins learning to identify ablated regions despite the significant domain shift.

Figure 1 presents classification maps across three images and all self-training rounds. The ground truth (GT column) shows ablated regions in red and non-ablated tissue in green.

Correcting over-prediction: Images A2-S2 and A3-S3 show progressive correction of excessive ablated predictions. The baseline predictions appear predominantly red (over-predicting ablated), but by Round 4–5, the predictions show better spatial correspondence with the ground truth ablated regions.

Learning to detect ablated tissue: Image A1-S1 demonstrates the model gradually learning to detect ablated regions that were entirely missed at baseline. While predictions remain imperfect, later rounds show emerging red regions that increasingly correspond to ground truth ablated areas.

Maintaining strong performance: Images A3-S1 and A2-S4 (not shown for brevity) maintain and improve their already-good baseline predictions, with ablated region boundaries becoming more precise through successive rounds.

Figure 2 visualizes the pseudo-labels selected at each round, overlaid on ground truth. The progressive expansion of pseudo-labeled regions from 1% to 5% budget is visible across rounds. Notably, the selection strategy tends to choose high-confidence predictions near class boundaries, providing informative training signal for subsequent rounds.

For images where baseline predictions are biased (e.g., A3-S3 over-predicting ablated), early pseudo-labels reflect this bias. However, as predictions improve through self-training, later rounds select more balanced and spatially accurate pseudo-labels, creating a positive feedback loop that drives continued improvement.

The Same-image column in Table 1 provides reference performance using the 5-click circle annotation protocol on each target image directly. While same-image training achieves higher average κ (0.36 vs. 0.35), our cross-image self-training approach reaches competitive performance without requiring any manual annotation on the target images. For image A3-S1, self-training actually exceeds same-image performance ($\kappa = 0.86$ vs. 0.58), demonstrating that effective domain adaptation can outperform limited same-image supervision when sufficient source data is available.

5. CONCLUSION

We presented a progressive self-training approach for cross-image pixel classification of ME-HSI datasets of RFA lesions on bovine cardiac tissue. Our method addressed the challenge of domain shift between source and target images through confidence-based pseudo-label selection with class-balanced constraints, enabling adaptation to new tissue samples without requiring manual annotations.

The experimental results demonstrated substantial improvements over the baseline model trained solely on source images. Average overall accuracy improved from 0.75 to 0.85, and Cohen's κ increased from 0.26 to 0.35, indicating meaningfully better classification agreement. The largest gains occurred for challenging images where the baseline model exhibited severe class imbalance—either over-predicting or completely missing ablated regions. For image A3-S1, self-training achieved $\kappa = 0.86$, exceeding the same-image supervised reference ($\kappa = 0.58$), suggesting that domain adaptation with sufficient source data can outperform limited target-domain supervision.

Several directions remain for future work. Extending evaluation to additional tissue samples would strengthen generalization claims. Combining self-training with other domain adaptation techniques, such as adversarial feature

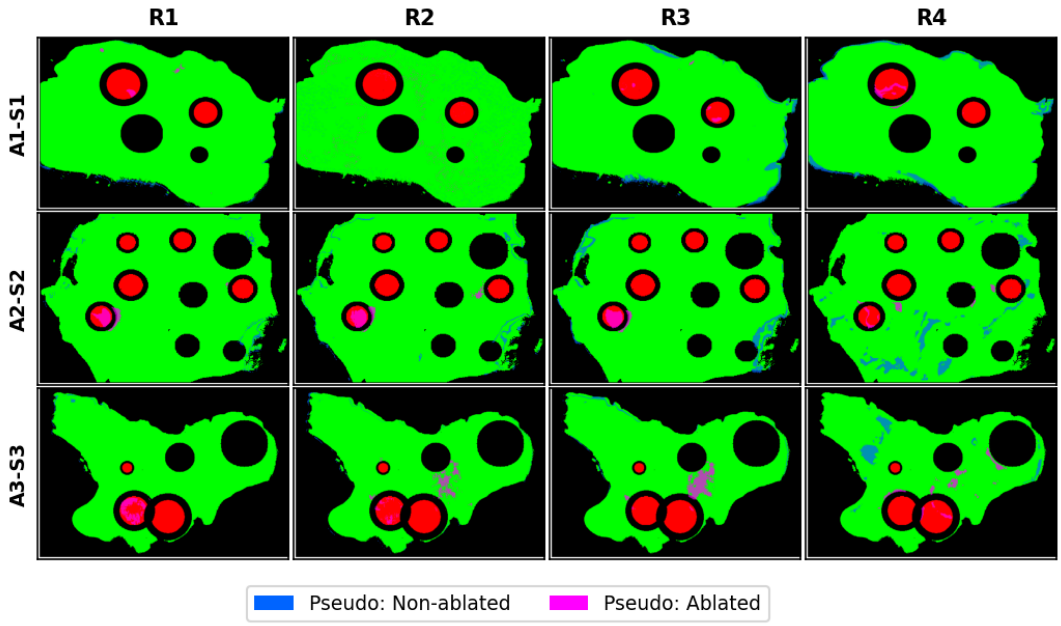


Fig. 2. Confidence-based pseudo-label selections across self-training Rounds 1–4 for three images. Magenta indicates pixels selected as non-ablated pseudo-labels, blue indicates pixels selected as ablated pseudo-labels. Pseudo-labels are generated for training the subsequent round, hence round 5 (final) has no pseudo-label selection.

alignment, may yield further improvements. Finally, optimizing inference speed and integrating the approach into real-time surgical guidance systems represents an important step toward clinical deployment of ME-HSI for RFA monitoring.

6. REFERENCES

[1] Ayodele Odutayo et al., "Atrial fibrillation and risks of cardiovascular disease, renal disease, and death: systematic review and meta-analysis," *BMJ*, vol. 354, pp. i4482, 2016.

[2] Hakan Oral et al., "Clinical significance of early recurrences of atrial fibrillation after pulmonary vein isolation," *JACC*, vol. 40, no. 1, pp. 100–104, 2002.

[3] Narek Chilingaryan et al., "4D hyperspectral imaging for intraoperative tissue classification," in *MI 2025: CBI*. SPIE, 2025, vol. 13410, pp. 255–260.

[4] Arpi Hunanyan et al., "Deep learning classification of ablated atrial tissue in multi-excitation hyperspectral images," in *WHISPERS*. IEEE, 2025.

[5] Ter-Petrosyan et al., "Comparative analysis of multi-excitation hyperspectral image configurations for ablated atrial tissue classification," in *WHISPERS*. IEEE, 2025.

[6] Zhuoqun Fang et al., "Confident learning-based domain adaptation for hyperspectral image classification," *IEEE TGRS*, vol. 60, pp. 1–16, 2022.

[7] Rong Liu et al., "Dual classification head self-training network for cross-scene hyperspectral image classification," *arXiv preprint*, 2025, arXiv:2502.17879.

[8] Jingpeng Gao et al., "Pseudo-class distribution guided multi-view unsupervised domain adaptation for hyperspectral image classification," *IJEOG*, vol. 136, pp. 104356, 2025.

[9] Zheng Zhao et al., "An iterative pseudo label generation framework for semi-supervised hyperspectral image classification using the segment anything model," *FPS*, vol. 15, pp. 1515403, 2024.

[10] Benjamin Graham et al., "LeViT: A vision transformer in ConvNet's clothing for faster inference," in *IEEE/CVF ICCV*, 2021, pp. 12259–12269.

[11] Luther M. Swift et al., "Hyperspectral imaging for label-free in vivo identification of myocardial scars and sites of radiofrequency ablation lesions," *HR*, vol. 15, no. 4, pp. 564–575, 2018.

[12] Yeva Gabrielyan et al., "Comparative analysis of deep learning methods for classification of ablated regions in hyperspectral images of atrial tissue," *IEEE Access*, vol. 13, pp. 35029–35047, 2025.