COMPARATIVE ANALYSIS OF MULTI-EXCITATION HYPERSPECTRAL IMAGE CONFIGURATIONS FOR ABLATED ATRIAL TISSUE CLASSIFICATION

Nazeli Ter-Petrosyan[1*], *Arpi Hunanyan*[1,2*], *Fernando Villarruel*[3], *Narine Sarvazyan*[1,3,4], *Aram Butavyan*[1], *and Varduhi Yeghiazaryan*[1]

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

ABSTRACT

Multi-excitation hyperspectral imaging (ME-HSI) is a novel technique that records detailed spectral signatures for each spatial pixel under multiple excitation wavelengths. The use of ME-HSI has been recently explored for biomedical applications such as tissue differentiation. The purpose of this work is to investigate the contribution of individual excitation wavelength in an ME-HSI setup to deep learning classification performance. We use an ME-HSI dataset of radiofrequency-ablated bovine atrial tissue, comprising eight samples imaged at four excitation wavelengths. Experiments are conducted on three deep learning models with different learning paradigms: graph neural network, visual transformer, and a multi-layer perceptron. We evaluate three aggregation techniques: (1) concatenation of all HSI cubes, (2) hard majority voting and (3) soft majority voting across predictions for all HSI cubes. Soft majority voting achieves the best performance for all models. Quantitative and qualitative analyses of all possible excitation combinations with soft majority voting reveal that multi-excitation configurations consistently outperform single-excitation ones, underscoring the complementary contributions of individual cubes.

***Index Terms*—** Hyperspectral imaging, multi-excitation fluorescence, tissue classification, deep learning, majority voting

1. INTRODUCTION

Atrial fibrillation (AF) is the most common cardiac arrhythmia and is strongly linked to serious cardiovascular complications [1]. A standard treatment for AF is radiofrequency (RF) ablation (RFA), a minimally invasive procedure where heat from high-frequency electrical currents is used to destroy arrhythmogenic tissue. A major challenge, however, is determining the exact extent of tissue damage during the procedure. Postoperative AF recurrence is often associated with viability gaps—areas of intact myocardium that continue to conduct electrical signals [2]. These limitations highlight the need for intraoperative, non-invasive imaging techniques that can visualize ablation effects in real time with both high specificity and spatial resolution.

Hyperspectral imaging (HSI) records rich spectral information for each spatial pixel across a wide wavelength range. Unlike conventional RGB imaging, which captures only three color channels, HSI produces a 3D dataset, or hyperspectral cube, with two spatial and one spectral (λ_{em}) dimensions. Biomedical applications of HSI are increasingly combined with machine learning approaches [3, 4, 5]. Its potential to assess lesion formation during RF ablation of atrial tissue was first demonstrated in [6]. More recently, [7] conducted a large-scale comparison of 45 deep learning (DL) methods for classifying ablated versus intact atrial tissue in HSI. Extending beyond conventional methods, [8] introduced multi-excitation HSI (ME-HSI), in which multiple hyperspectral cubes are captured under different excitation wavelengths (λ_{exc}) and combined into a single 4D dataset, improving tissue differentiation. This imaging technique was also utilized for nerve and ligament differentiation, where convolutional neural networks (CNNs) trained on ME-HSI achieved accuracies of up to 99% [9, 10].

This work is an extension of [11] and should be considered in pair with it. In [11] we utilized an ME-HSI dataset of RF-ablated bovine atrial tissue. We showed that ME-HSI configurations with four λ_{exc} allowed for better tissue differentiation than traditional single-excitation HSI. Our experiments with multiple DL approaches showed that ME-HSI setups typically lead to higher tissue classification scores. The first of the two tested ME-HSI aggregation techniques was concatenation of the multiple input HSI cubes along the spectral axis, thus reducing the 4D structure to a 3D structure with more spectral bands. The second technique was to train the same DL model on the four HSI cubes (for each λ_{exc}) sepa-

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

* Equal contribution.

rately and to perform majority voting on the models' outputs.

While [11] focused on benchmarking a range of DL architectures across multiple experimental setups, here we refine the analysis in several directions. Building on our prior finding that multi-excitation setups—concatenated cubes and majority voting—outperformed single-excitation approaches, we further explore this result by introducing soft majority voting as a third contender. More importantly, we investigate the contribution of each λ_{exc} in an ME-HSI setup to DL classification performance. We quantitatively and qualitatively analyze different excitation combinations across all configurations to identify which λ_{exc} contribute most strongly to classification accuracy. In all experiments, we fix training data generation to a 5-click simulated annotation strategy, which was shown in the companion paper [11] to provide the most stable results. For all experiments, we use the three models that achieved the best performance in [11]: GiGCN (a graph neural network), LeViT (a transformer-based architecture), and a standard multi-layer perceptron (MLP).

2. DATASET

Sample collection: Bovine heart samples were provided by a local butcher, immediately after sacrifice of the animals. The hearts were dissected in-place, extracting the atria which were immediately transported on-ice to the lab. The dataset consisted of eight tissue samples from three animals.

RFA execution: Once in the lab, left atria (LA) were separated into pieces for RFA aplication using a *EPT-1000 XP* machine at 8 W power for 15 s, connected to a 4 mm *Blazer* catheter (*Boston Scientific*) which was perpendicularly placed on the tissue during ablation.

HSI acquisition: After ablation, the tissue was imaged using an HSI setup consisting of: (i) 300 W xenon lamp outfitted with four filters centered at 360, 370, 380 and 390 nm (all with 10 nm FWHM; *Thorlabs*) and (ii) a Nuance FX (*Perkin Elmer*) imaging system. Four cubes were obtained, with the above-mentioned λ_{exc} and 31 emission wavelengths (λ_{em} = 420–720 nm with 10 nm step), while spatial dimensions varied depending on tissue size. All imaging was completed within less than six hours after animal sacrifice.

Histological analysis: Upon completion of imaging, blocks of tissue containing the lesions were stained with 1.0% TTC solution, to determine the lateral size of each lesion and to generate the ground truth masks.

Mask generation: The above-mentioned process does not guarantee a completely reliable estimate of the degree of tissue damage. Due to tissue structure heterogeneity (invaginations and folded tissue, collagen thickness differences), for some lesions, the center coordinates were unclear in the obtained image. Also, varying degrees of ablation are obtained using the same settings, leading to different lesion sizes or even TTC-undetected lesions.

To create the ground truth masks, lesion center coordinates were found (as a point in the image) and the TTC obtained lateral sizes considered, assuming the lesions to be circular. The outer 30% of the obtained area was labeled as 'unsure' to consider slicing overestimation, and excluded from DL training/test sets. The inner 70% was labeled as 'ablated'. When the ablation center was unclear, an approximated center was identified, and the whole circle was labeled as 'unsure'. For the TTC undetected ablations, an 'unsure' circle of 0.6 mm diameter was created around the center coordinates. The background pixels were filtered using a signal threshold at the λ_{em} with the maximum signal. The rest of the tissue pixels were labeled as 'unablated'.

'Ablated' and 'unablated' pixel classes were eligible for both training and test groups. The 'unsure' class was excluded from training and test sets, but predictions with the resulting methods were still made for qualitative analysis. Finally, the 'background' class was excluded from the dataset prior to model calculation.

3. METHODS

For the evaluation of the different single- and multi-excitation HSI combinations, we choose the three architectures that showed the best performance in the companion paper [11]: GiGCN [12], LeViT [13] and MLP [14]. These models also represent distinct learning paradigms, allowing us to evaluate the generalizability of different multi-excitation setups.

GiGCN is a graph neural network tailored for hyperspectral imaging. It constructs both internal and external graphs to hierarchically extract features from superpixel regions, enabling rich spatial context modeling through graph-based learning. LeViT is a hybrid vision transformer that combines convolutional layers with self-attention mechanisms. Initially proposed for RGB image classification, LeViT replaces standard linear projections with convolutional embeddings. In our setting, it is adapted to leverage spatial information across the HSI cube. MLP represents a simple four-layer fully connected neural network, used as a baseline model. It lacks any specialized mechanisms for extracting spatial or spectral features.

We evaluate model performance using three statistical metrics. The first metric, overall accuracy (OA), reflects the proportion of correctly classified samples across all predictions. To account for class imbalance, we also report average accuracy (AA), which computes the mean of per-class accuracies, offering a more balanced view of performance. Finally, we include the Cohen's Kappa statistic (κ) to assess the agreement between predicted and true labels beyond chance. This metric ranges from -1 (complete disagreement) to 1 (perfect agreement), with 0 indicating random agreement.

One of the objectives of this work is to further investigate and comparatively analyze strategies for utilizing ME-HSI data in the context of ablated atrial tissue classification.

Specifically, we evaluate three aggregation techniques: concatenation of all input HSI cubes along the emission axis into a single cube, hard majority voting and soft majority voting among outputs for all HSI cubes. In the hard majority voting case, ties are resolved by random selection. Soft majority voting implies averaging probabilistic outputs of the DL architecture for all four HSI cubes *before* binarizing them into classification results.

We evaluate HSI classification using a training pipeline based on simulated user clicks. For each class (ablated or unablated), five clicks are placed, each covering 0.2% of the image area, resulting in a total training coverage of 2%.

The experiments are performed on an Ubuntu 24.04 system with an Intel Core i9-14900K processor and an NVIDIA GeForce RTX 4090 GPU.

4. RESULTS AND DISCUSSION

4.1. Aggregation Techniques

Table 1: Mean OA, AA, and κ scores, $\pm$ standard deviation, for three DL methods and three aggregation techniques: concatenation, hard and soft majority voting (MV). The ME-HSI setup includes all four excitations: 360–390 nm. Per column, the best and second scores are in boldface and underlined, respectively.

	GiGCN			LeViT			MLP		
	OA	AA	κ	OA	AA	κ	OA	AA	κ
Concat	.859	<u>.794</u>	<u>.393</u>	**.802**	.848	<u>.362</u>	.761	<u>.832</u>	.309
	±.087	±.109	±.241	±.119	±.070	±.192	±.129	±.071	±.176
Hard MV	**.887**	.726	.375	.798	<u>.850</u>	.360	<u>.779</u>	.831	<u>.335</u>
	±.077	±.123	±.252	±.115	±.064	±.202	±.117	±.072	±.205
Soft MV	<u>.865</u>	**.805**	**.441**	<u>.800</u>	**.851**	**.371**	**.787**	**.838**	**.352**
	±.097	±.099	±.261	±.121	±.068	±.206	±.118	±.072	±.211

Across the three DL methods, Table 1 presents a comparison of the impact of the three aggregation techniques: concatenation, hard majority voting, and soft majority voting. The AA and κ results clearly indicate that soft majority voting yields the best performance across all three models. For OA, MLP also favors soft majority voting, while GiGCN and LeViT achieve highest OA with hard majority voting and concatenation, respectively. This discrepancy can be attributed to OA being heavily influenced by the majority unablated class. This means that models biased toward this class tend to score higher OA despite poorer class balance.

For all metrics and DL models, there is only marginal variation among the three aggregation techniques, the difference between the best and worst scores mostly staying below 0.05. Still, soft majority voting occasionally achieves a sharp increase of scores, e.g. with a κ score of 0.441 for GiGCN in comparison to 0.375 for hard majority voting and 0.393 for concatenation.

Overall, soft majority voting is the most effective aggregation technique. Unlike hard majority voting, which can neglect minority-class information by favoring the most common predictions, soft majority voting and concatenation better preserve and integrate signals from each individual excitation in the ME-HSI data, reinforcing that these separate excitations contribute complementary information that improves classification performance.

4.2. Single- and Multi-Excitation HSI Combinations

Table 2: Mean OA, AA, and κ scores, $\pm$ standard deviation, for three DL methods across all possible single- and multi-excitation HSI combinations, using soft majority voting. Each column is color-coded with a gradient based on its minimum and maximum values. The encoding for λ_{exc} values is: 1 for 360 nm, 2 for 370 nm, 3 for 380 nm, and 4 for 390 nm.

	GiGCN			LeViT			MLP		
	OA	AA	κ	OA	AA	κ	OA	AA	κ
1	.867	.743	.336	.775	.830	.310	.777	.821	.331
	±.083	±.137	±.243	±.105	±.066	±.172	±.117	±.065	±.208
2	.866	.707	.310	.782	.839	.354	.751	.805	.285
	±.101	±.164	±.308	±.135	±.077	±.211	±.117	±.078	±.187
3	.872	.704	.312	.782	.829	.335	.773	.820	.318
	±.082	±.148	±.271	±.127	±.086	±.211	±.115	±.076	±.194
4	.856	.717	.307	.784	.840	.341	.767	.823	.320
	±.095	±.135	±.254	±.115	±.070	±.205	±.124	±.074	±.199
12	.864	.776	.381	.788	.842	.341	.783	.831	.334
	±.096	±.124	±.261	±.121	±.072	±.178	±.110	±.062	±.199
13	.870	.789	.409	.790	.838	.338	.791	.836	.350
	±.084	±.110	±.245	±.114	±.070	±.187	±.110	±.066	±.210
14	.858	.776	.376	.795	.847	.350	.790	.840	.350
	±.094	±.122	±.259	±.108	±.065	±.190	±.111	±.062	±.207
23	.863	.757	.388	.791	.844	.365	.775	.825	.328
	±.103	±.126	±.282	±.136	±.078	±.213	±.120	±.077	±.202
24	.861	.744	.246	.791	.848	.368	.774	.829	.328
	±.104	±.140	±.841	±.129	±.072	±.213	±.124	±.076	±.198
34	.861	.760	.374	.796	.844	.361	.779	.832	.340
	±.095	±.124	±.270	±.119	±.071	±.214	±.122	±.075	±.209
123	.866	.806	.437	.794	.846	.357	.786	.835	.345
	±.095	±.091	±.251	±.124	±.071	±.195	±.113	±.069	±.209
124	.863	.786	.403	.798	.850	.365	.787	.838	.346
	±.097	±.119	±.265	±.119	±.068	±.199	±.113	±.066	±.205
134	.861	.803	.428	.798	.848	.360	.789	.840	.354
	±.092	±.095	±.249	±.113	±.066	±.201	±.115	±.069	±.213
234	.862	.768	.401	.797	.849	.373	.780	.833	.342
	±.102	±.123	±.284	±.128	±.071	±.215	±.124	±.077	±.207
1234	.865	.805	.441	.800	.851	.371	.787	.838	.352
	±.097	±.099	±.261	±.121	±.068	±.206	±.118	±.072	±.211

Table 2 presents the results for all possible single- and multi-excitation combinations using soft majority voting

across the three DL methods. A clear trend emerges: combinations involving multiple excitations consistently outperform single-excitation setups. The latter tend to yield the lowest scores across models. No single combination dominates all metrics, and performance differences across models are more pronounced than those between the excitation configurations themselves.

Among the double-excitation combinations, across metrics, the top-performing configurations are *13* for GiGCN, *24* for LeViT, and *13* and *14* for MLP. For triple combinations, the strongest performers are *123* for GiGCN, *124* for LeViT, and *134* for MLP. Based on AA, the best-performing configurations are *123* for GiGCN, *1234* for LeViT, and *14* and *134* for MLP. Performance across the remaining metrics further supports the conclusion that these combinations deliver superior overall results.

Fig. 1 presents a qualitative comparison of different single- and multi-excitation combinations across the three DL methods. For GiGCN, an interesting pattern emerges: whenever excitation *1* (λ_{exc} = 360 nm) is included, the model correctly classifies the ablated region. In contrast, when this excitation is absent, the entire sample is misclassified as unablated. This highlights the critical importance of excitation *1* for GiGCN in accurately detecting ablated areas.

For LeViT, some single-excitation setups yield reasonable predictions, but they are accompanied by a notable number of false positives, particularly around the edges of the sample tissue. As the number of excitations increases, the model produces fewer false positives, with *234* and *1234* standing out as the best configurations, matching the quantitative results.

The MLP, serving as the baseline model, does not perform well under any configuration. However, even in this case, a trend is observed: single-excitation setups result in scattered and noisy predictions, whereas increasing the number of excitations leads to slightly more coherent and defined classification results.

Previous work has focused on the physical–chemical factors behind the auto-fluorescence changes in ablated tissue. [15] has found that, after ablation, the main changes in the observed fluorescence spectrum (λ_{exc} = 365 nm) can be attributed to the decrease of the NADH signal, and increase of both the collagen and the yellow pigment signal.

The large impact of cube *1* (λ_{exc} = 360 nm) on the resulting accuracy could be related to the increased collagen autofluorescence, which is greater in the near-UV region. Besides, NADH signal is maximum at λ_{exc} = 365 nm, cube *1* excitation is near that maximum, and the typical fluorescence redshift due to NADH drop is still observable. In this way, this cube may maintain the maximum information from both collagen and NADH signals, leading to its greater effect on prediction performance. On the other hand, the enhanced performance observed with increasing the number of excitations can be due to better combined observations of the NADH drop signal and the yellow pigments' fluorescence (with excitation maxima closer to the visible region). Although current experiments indicate performance improvement for ME-HSI setups with a few λ_{exc}, further experiments are necessary to reveal the impact of more excitations for the specific task, and in general.

5. CONCLUSION

In this work, we evaluated different configurations of ME-HSI for the classification of RF-ablated bovine atrial tissue. Three multi-excitation aggregation techniques were compared: (1) concatenation of all HSI cubes, (2) hard majority voting and (3) soft majority voting across predictions for individual HSI cubes, using three DL models with distinct architectural paradigms: GiGCN, LeViT, and MLP. Among these techniques, soft majority voting achieved the strongest overall performance. Both quantitative and qualitative analyses further demonstrated that multi-excitation configurations consistently outperform single-excitation setups, underscoring the critical contributions of individual excitation channels to accurate classification.

For future work, we aim to investigate the impact of different excitation configurations on additional tasks, as the current results may be highly task-dependent. Additionally, we plan to analyze model performance in the classification of 'unsure' regions, with the goal of identifying which configurations and architectures are best suited for detecting these regions alongside the ablated tissue. Finally, we intend to modify model architectures to inherently support the 4D setup of ME-HSI, rather than limiting our exploration to aggregation-based techniques.

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, 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] Baowei Fei, "Hyperspectral imaging in medical applications," in *Data handling in science and technology*, vol. 32, pp. 523–565. Elsevier, 2019.

[4] Rong Cui et al., "Deep learning in medical hyperspectral images: A review," *Sensors*, vol. 22, no. 24, pp. 9790, 2022.

[5] Uzair Khan et al., "Trends in deep learning for medical hyperspectral image analysis," *IEEE Access*, vol. 9, pp. 79534–79548, 2021.

[6] Luther M Swift et al., "Hyperspectral imaging for label-free in vivo identification of myocardial scars and sites

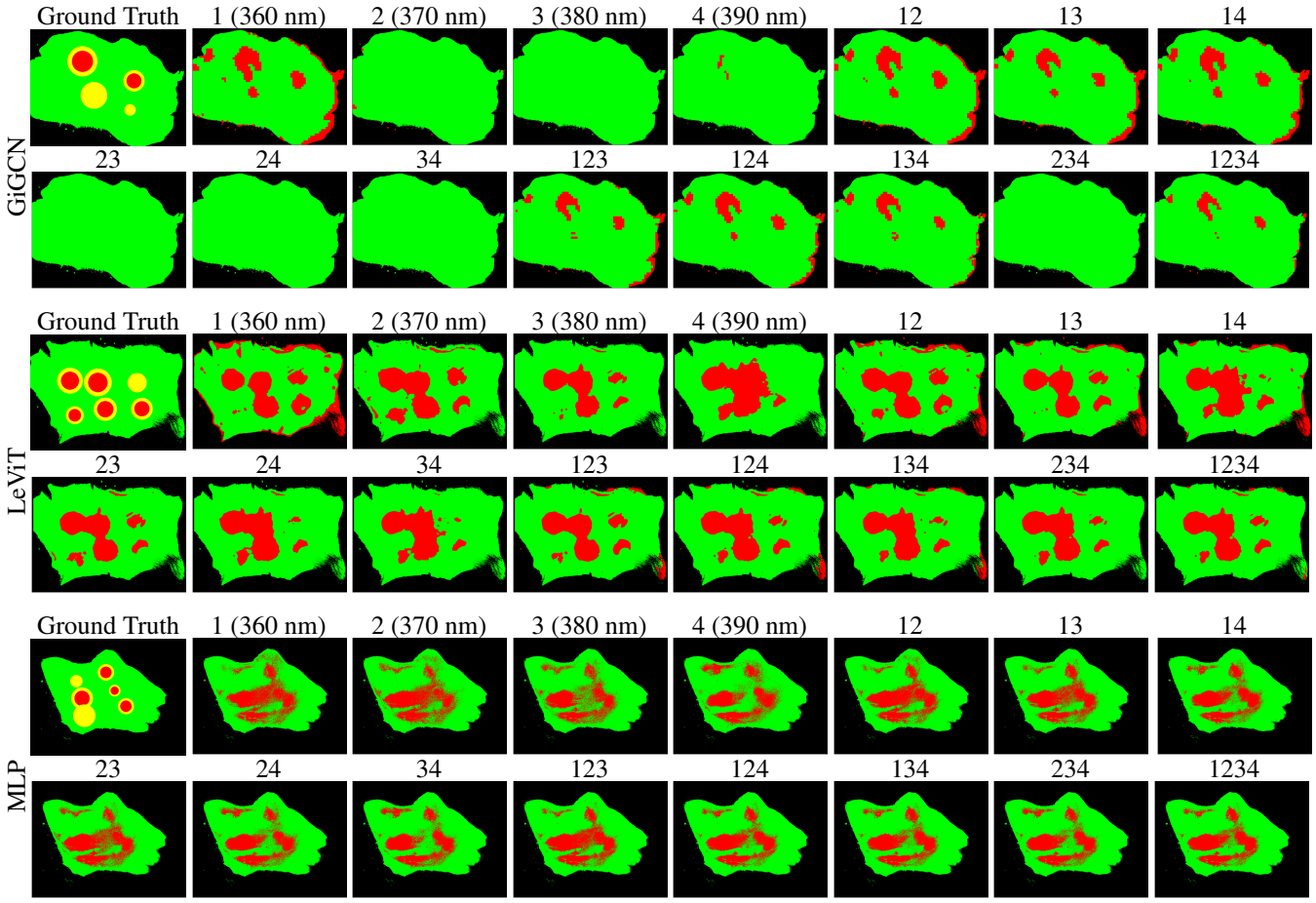

Fig. 1: Visualization of tissue classification results based on different single- and multi-excitation HSI combinations, using soft majority voting, across three DL methods: GiGCN, LeViT, and MLP. The selected samples illustrate peculiar behavioral patterns for each DL model across the various ME-HSI configurations.

of radiofrequency ablation lesions," *Heart Rhythm*, vol. 15, no. 4, pp. 564–575, 2018.

[7] 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.

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

[9] Naira Matosyan et al., "Combining 4D hyperspectral imaging with CNN for nerve and ligament differentiation," in *ISBI*. IEEE, 2025, pp. 1–5.

[10] Naira Matosyan et al., "Spectral pixels as images: CNN-based pixel classification of 4D hyperspectral data for nerve and ligament differentiation," in *MI 2025: IP*. SPIE, 2025, vol. 13406, pp. 565–573.

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

[12] Sen Jia et al., "Graph-in-graph convolutional network for hyperspectral image classification," *IEEE Trans Neural Netw Learn Syst*, vol. 35, no. 1, pp. 1157–1171, 2022.

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

[14] Xiaofei Yang et al., "Extension of deep learning toolbox based on pytorch for hyperspectral data classification," `https://github.com/xiachangxue/DeepHyperX`, Accessed: 18 October 2024.

[15] Narine Muselimyan et al., "Key factors behind autofluorescence changes caused by ablation of cardiac tissue," *Scientific Reports*, vol. 10, no. 1, pp. 15369, 2020.