Optimizing Deep Learning for Satellite Hyperspectral Data: an xAI-Driven Approach to Hyperparameter Selection

Michele Linardi
CY Cergy Paris Université
ETIS UMR-8051, Paris, France
michele.linardi@cyu.fr

Sékou Dabo
CY Cergy Paris Université
ETIS UMR-8051, Paris, France
dabo.sekou@cyu.fr

Claudia Paris
University of Twente, ITC
7522NH Enschede, The Netherlands
c.paris@utwente.nl

***Abstract*—The growing availability of spaceborne Hyperspectral Imaging (HSI) missions combined with advancements in Deep Learning (DL), offers significant potential for global environmental mapping. However, most DL methods are tailored to airborne HSI, making it challenging to adapt them for optimal performance with satellite data. To solve this issue, this study explores the use of explainable Artificial Intelligence (xAI) to adapt existing DL architectures to spaceborne HSI. In particular, the best hyperparameter selection is carried out by evaluating the consistency of the explanations across different model instances using xAI Integrated Gradients method. Experiments were conducted using two cutting-edge DL models commonly used for airborne HSI: an attention-based Vision Transformer (ViT) and a standard 2D-Convolutional Neural Network (CNN). Results from a crop-type mapping task using Hyperspectral Precursor and Application Mission (PRISMA) satellite data demonstrate the effectiveness of the proposed approach in facilitating the optimal hyperparameter selection, i.e., the one able to maximize the macro Fscore obtained on the test set.**

***Index Terms*—explainable Artificial Intelligence (xAI), Hyperspectral Imaging (HSI), Deep Learning (DL), Hyperspectral Precursor and Application Mission (PRISMA) data.**

I. Introduction

Hyperspectral Imaging (HSI), with their ability to capture spectral information in hundreds of narrow contiguous bands, are highly valuable for environmental monitoring [1]. Recent advancements in Deep Learning (DL), particularly Convolutional Neural Network (CNN) and attention-based models, have demonstrated to be effective in addressing the high dimensionality of Hyperspectral Imaging (HSI), leading to accurate classification results [2]. In [3], Yang *et al.* proposed a method that extracts global and local features from HSI using both Transformer and CNNs. Similarly, in [4], Ahmad *et al.* introduced a hybrid approach that integrates multiscale CNNs and transformers for improved HSI classification performance. However, most existing methods are tailored to the spatial and spectral characteristics of airborne HSI data, while few works have been defined to work with spaceborne HSI [2, 5]. While airborne HSI typically offer high spatial resolution (e.g., $\leq$5m), resulting in pixels with nearly pure spectral signatures, spaceborne HSI generally have coarser spatial resolution (e.g., 30m), leading to pixels with mixed spectral signatures that cover larger areas [2, 6]. In this context, it is crucial to understand how to automatically adapt DL methods developed for airborne HSI to effectively handle spaceborne data.

Recently, explainable Artificial Intelligence (xAI) approaches have been increasingly applied to remote sensing data to interpret the outputs of DL architectures [7]. These methods are used according to the application context, specific explainability goals, data properties, and end-user expectations [8]. Indeed, different xAI metrics can offer different perspectives on the quality of explanations [9]. To address this variability, Duan *et al.* introduces Meta-Rank [9], a framework that quantitatively evaluates xAI methods by calculating a score that reflects their stability across various datasets, models, and evaluation protocols, enabling a ranking of xAI methods from most to least stable.

In this work, we aim to assess the robustness of xAI methods to perform hyperparameter analysis and optimization. Specifically, we focus on two state-of-the-art DL models commonly applied to airborne HSI, an attention-based Vision Transformer (ViT) [10] and a standard 2D-CNN [11]. To adapt these models to the unique properties of spaceborne HSI, we analyzed the consistency [12] of the explanation across different model instances derived using the Integrated Gradients xAI method. Assuming that effective DL architectures produce more confined and interpretable results, the proposed method computes a consistency metric to automatically identify the best hyperparameter setup without requiring additional labeled validation data. Experiments conducted with spaceborne Hyperspectral Precursor and Application Mission (PRISMA) data to address a downstream crop type mapping task demonstrate the effectiveness of the proposed approach.

II. xAI-driven hyperparameter selection

This work aims to investigate the use of xAI for DL hyperparameters analysis and optimization, specifically targeting spaceborne HSI. Figure 1 shows the block-scheme

This work is supported by the "EDEM: Explaining Deep Learning Models of Satellite Time Series for the Agritech domain" project funded by l'Agence Nationale de la Recherche (ANR), EDEM project ANR-24-CE23-6449.

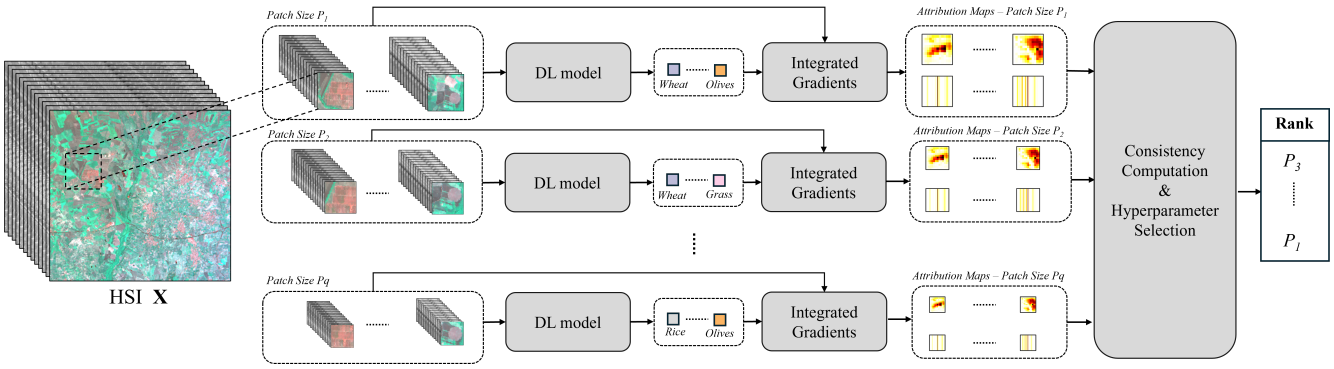

Fig. 1: Block-scheme of the proposed approach. The HSI is divided into patches of different sizes. To determine the optimal patch size, the method evaluates the consistency of the spatial and spectral attribution maps generated across different patches.

of the proposed approach. In the preliminary version of the method, we focused on the optimization of the patch size hyperparameter, which sharply affects the performance of the DL models. This parameter could be drastically different when moving from airborne and spaceborne data due to the fundamental differences in the spatial characteristics of these data types, which affect how spatial context and spectral features are captured and learned by the DL models. First, the DL models utilized in the study are described. Then, the employed Integrated Gradients xAI method is presented. Finally, the approach for automatically determining the optimal patch size hyperparameter is presented.

A. DL classification models

To assess the proposed approach, we considered two state-of-the-art approaches widely used for airborne HSI data classification, an attention-based ViT [13] approach and 2D-CNN. In our study, we employed a pixel-by-pixel classification approach while incorporating a neighborhood context, thereby forming a patch. Each patch of size P is obtained by extracting the central pixel along with a neighborhood of radius $P//2$ (where P is an odd integer). For an input $X \in \mathbb{R}^{H\times W\times B}$, where $H \times W$ represents the spatial dimensions and B the number of spectral bands, we generate $H \times W$ patches with dimensions $P \times P \times B$. Given the high dimensionality of the data, we added a dimensionality reduction block to the input of our models, consisting of a 1×1 kernel and N filters. This allowed us to reduce the number of bands B to N. As a result, our patches are now of size $P \times P \times N$.

The 2D-CNN used in our approach is a lightweight CNN with approximately 0.5 million parameters. Its core architecture consists of four convolutional blocks (Conv–BatchNorm–ReLU), followed by global average pooling and a Multi-Layer Perceptron (MLP) equipped with dropout and layer normalization for regularization. The convolutional layers operate on HSI patches to extract spatial–spectral features, as follows:

$$\sigma\left(\sum_{b=0}^{N-1}\sum_{m=0}^{P-1}\sum_{n=0}^{P-1}\mathbf{W}(m,n)\mathbf{X}(x+m,y+n,b)+b_k\right), \quad (1)$$

where $\mathbf{W} \in \mathbb{R}^{U\times U}$ is the kernel, b_k is the bias for the k-th filter, σ is the activation function and U is the size of the of the convolutional kernel. The obtained features map $\in \mathbb{R}^{P'\times P'\times N'}$. The ViT model is based on an encoder-only Transformer architecture with 6 encoder layers, each employing 4 attention heads, and a total of approximately 0.81 million parameters. It processes non-overlapping image patches that are linearly projected into token embeddings, augmented with positional encodings and a classification token for global representation learning.

$$\mathbf{Z}_0 = [\mathbf{x}_{\text{class}}; \mathbf{x}_1\mathbf{P}_E; \mathbf{x}_2\mathbf{P}_E; \ldots; \mathbf{x}_N\mathbf{P}_E] + \mathbf{E}_{\text{pos}}, \quad (2)$$

where $\mathbf{P}_E \in \mathbb{R}^{(P^2\cdot C)\times D}$ is the patch embedding projection, with C representing the number of reduced channels obtained from the dimensionality reduction step, $\mathbf{E}_{\text{pos}} \in \mathbb{R}^{(N+1)\times D}$ is the positional encoding, and $\mathbf{Z}_0$ is the input sequence to the Transformer encoder. The Transformer encoder then applies self-attention and a feedforward network across L layers:

$$\mathbf{Z}'_\ell = \text{MSA}(\text{LN}(\mathbf{Z}_{\ell-1})) + \mathbf{Z}_{\ell-1}, \quad \ell = 1,\ldots,L, \quad (3)$$

$$\mathbf{Z}_\ell = \text{MLP}(\text{LN}(\mathbf{Z}'_\ell)) + \mathbf{Z}'_\ell, \quad \ell = 1,\ldots,L, \quad (4)$$

$$\mathbf{y} = \text{LN}(\mathbf{Z}_0^L), \quad (5)$$

where MSA denotes multi-head self-attention, LN is layer normalization, and MLP is a position-wise feedforward network. The final output $\mathbf{y}$ is derived from the transformed class token, with $\mathbf{Q}, \mathbf{K}, \mathbf{V}$ being the query, key, and value matrices derived from $\mathbf{Z}_0$.

B. xAI approach

In the considered study, we apply the Integrated Gradients xAI agnostic method [14]. This approach is chosen among various xAI methods due to its adaptability to HSI data, relatively low computational cost, and effectiveness in model evaluation. Let $F : \mathbb{R}^{P\times P\times B} \mapsto \Omega^V$ be the DL model used to address the considered classification task, where $\Omega = \{\omega_n\}_{n=1}^V$ denotes the set of V classes. The Integrated Gradients method attributes a model's prediction to its input features by integrating the model's gradients along a linear path from a baseline input $\mathbf{X}_{\text{base}}$ to the actual input $\mathbf{X}$. For

a HSI image input $\mathbf{X}$ and a baseline input $\mathbf{X}_{\text{base}}$ (typically a zero tensor), the attribution tensor $\mathbf{A} \in \mathbb{R}^{P\times P\times B}$ quantifies the contribution of each feature. The attribution for the l-th feature is computed as the path integral of the gradients along the straight-line path between $\mathbf{X}_{\text{base}}$ and $\mathbf{X}$:

$$\mathrm{A}_l = (X_l - X_{\text{base},l}) \int_{\alpha=0}^{1} \frac{\partial F(\mathbf{X}_{\text{base}} + \alpha(\mathbf{X} - \mathbf{X}_{\text{base}}))}{\partial x_l} d\alpha, \quad (6)$$

As represented in Figure 1, we generate for each patch an attribution map in both the spatial and spectral domains. The integral component of Eq 6 is approximated by a Riemann sum over m arbitrary steps. Hence, computing explanations for an instance X_l takes $O(mC)$ time, where C is the cost of the model inference.

C. Hyperparameter Selection

Unlike conventional methods in the literature, our hyperparameter selection is not based on maximizing the accuracy of a validation set of labeled samples. As selection criteria, we consider the consistency of the explanations across different model instances to understand how to adapt the existing DL model to spaceborne HSI data. This can be extremely relevant in scenarios where reference data is scarce. Moreover, leveraging xAI for hyperparameter selection not only optimizes model performance but also ensures the model is interpretable and robust, thus moving from purely performance-driven optimization to a more holistic and informed model development strategy.

xAI Consistency Let $X \in \mathbb{R}^{P\times P\times B}$ denote an input sample, $F : \mathbb{R}^{P\times P\times B} \mapsto \Omega^V$ a classifier, and $\phi : \mathbb{R}^{P\times P\times B} \to \mathbb{R}^{P\times P\times B}$ an attribution method. For a given input X, $\phi(X)$ produces an attribution map that assigns importance scores to each spectral feature. Here, for simplicity, we average the contribution of each single band over the entire spatial dimension. Hence, the final attribution map becomes $F_s = (f_1, \ldots, f_B) \in \mathbb{R}^B$. The computed explanation consistency is based on Jensen-Shannon divergence (JSdiv) [15], which quantifies how concentrated the attribution is on a subset of spectral bands. The metric is computed as follows:

$$\text{JSdiv} = \frac{1}{2} D_{\text{KL}}(F_s \parallel M) + \frac{1}{2} D_{\text{KL}}(U \parallel M), \quad (7)$$

where $U = (\frac{1}{N}, \ldots, \frac{1}{N})$ is the uniform distribution and $M = \frac{1}{2}(F_s + U)$ is the average distribution, and D_{KL} is the standard Kullback-Leibler Divergence distance [15]. A high JSdiv indicates that the model relies on a small subset of features (high simplicity), while a low value implies more uniform usage across all features (higher complexity). Based on the reasonable assumption that effective DL architectures produce more confined and interpretable results, the proposed xAI consistency metric is defined as follows:

$$Consistency^{JSdiv} = 1 - \text{JSdiv}, \quad (8)$$

where the lowest value represents the most compact model focusing on the smallest important feature set.

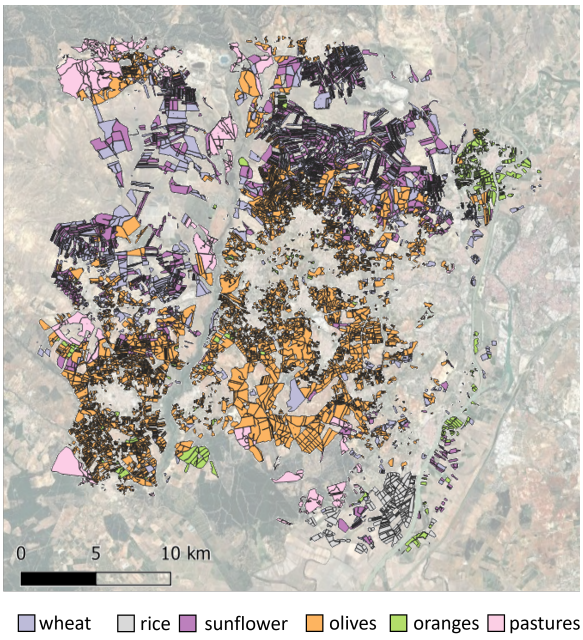

Fig. 2: Visual representation of the considered study area. The main crops are highlighted with different colors.

TABLE I: Number of reference samples per crop type available in the considered study area.

Classes	# Samples	# Training	# Validation	# Test
Durum Wheat	64601	18806	12544	33251
Olives	210996	84760	56834	69402
Oranges	8836	3847	2576	2413
Pastures	30324	6121	4084	20119
Rice	1717	807	542	368
Sunflower	67669	21034	14113	32522

III. Dataset Description and Experimental Setup

To favor reproducibility, the code and all the details of our implementations are online [16]. To test the proposed approach, we considered a crop type mapping downstream task using a satellite PRISMA Level-1 image (top-of-atmosphere radiometrically and geometrically calibrated radiance) acquired on the 27th of May 2020[1]. Before performing the classification, we discarded the noisy and water absorption spectral bands according to [17], reducing the original 234 bands to a final HSI data having 174 bands. The PRISMA data are characterized by a spatial resolution of 30m, resulting in mixed spectral signatures, unlike airborne HSI data, which typically feature a spatial resolution of less than 5 meters.

The study area is an agricultural region located in Western Seville, Spain, with publicly available agricultural reference data[2]. Figure 2 reports the considered study area, by representing the main crops present in the scene in 2020, which

[1]obtained from https://prisma.asi.it/ according to the General conditions for the provision of PRISMA products and the PRISMA License.

[2]https://www.juntadeandalucia.es/organismos/agriculturapescaaguaydesarrollorural/servicios/sigpac/visor/paginas/sigpac-descarga-informacion-geografica-shapes-provincias.html

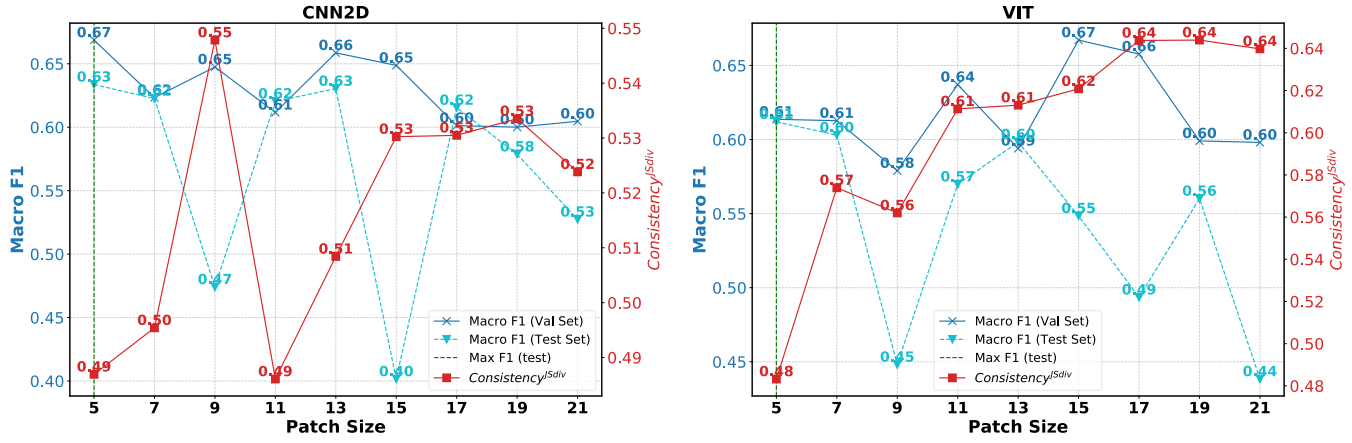

Fig. 3: Macro Fscore (F1) computed on the test and validation dataset (in dark plain blue and light dashed blue, respectively) and $Consistency^{JSdiv}$ metric (in red) when considering different patch sizes using: (a) the 2D-CNN, and (b) the ViT models.

are "Olives", "Oranges", "Pastures", "Rice", "Sunflower" and "Wheat". Table I reports the number of samples per crop used for training the DL models (training data) and assesses the accuracy (test data). To ensure statistical independence, the training and test datasets are extracted from spatially distinct areas, with test data obtained from the left side of the river and training, along with validation data from the right side.

As an initial evaluation of the proposed approach, we concentrated on one of the most critical hyperparameters influencing model performance: the patch size, which varies from 5 to 21 in increments of 2. Both models were trained considering a number of epochs of 30, a learning rate of $3e{-}4$, and a batch size of 256.

IV. Experimental Results

Figure 3(a) and Figure 3(b) show the macro Fscore (F1) computed on the test and validation sets when considering different patch sizes with the 2D-CNN and the ViT models, respectively. The considered $Consistency^{JSdiv}$ metric is reported in red. The results indicate that, as expected, this hyperparameter plays a critical role in influencing the classification performance of both DL models due to its direct impact on the contextual information available to the model. Specifically, the macro F-score on the test set for the 2D-CNN ranges from a minimum of 0.40 (at patch size of 15) to a maximum of 0.63 (at patch size of 13 and 5). Figure 3(a) highlights a correlation between the consistency metric and and F-score behavior on the test set, with the maximum F-score occurring at a patch size of 13 and 5, closely followed by the lowest $Consistency^{JSdiv}$ values at a patch size of 11 and patch size 5. Compared to the standard approach of selecting hyperparameters by maximizing validation accuracy, the selected patch size is equal to 5.

A similar pattern is observed with the ViT model, where the macro F-score on the test set varies from a minimum of 0.44 (with a patch size of 21) to a maximum of 0.61 (with a patch size of 5). Notably, both DL models identified the optimal patch sizes equal to 5 for both the ViT and the 2D-CNN model, according to the highest accuracy achieved on the test set (upper bound). In this case, the proposed $Consistency^{JSdiv}$ metric reaches its absolute minimum at a patch size of 5, whereas selecting the patch size based on maximum validation accuracy (i.e., the baseline approach) would result in choosing a patch size of 15. It is worth noting that the maximum value of JSdiv (0.64) are associated with the patch sizes that lead to the minimum F-score on the test set (i.e., 17 and 21).

In both models, we also note that high $Consistency^{JSdiv}$, which reveals higher uniformity in relevant features, strongly denotes a divergence between validation and test performance. Such an observation confirms that a model focusing on a larger set of salient features (when $Consistency^{JSdiv}$ is high) is more sensitive to data distribution shift (occurring across validation and test datasets).

V. Conclusion

This paper presents an xAI-driven approach for hyperparameter selection to optimize existing DL models tailored from airborne HSI data to spaceborne data. Classification results on PRISMA data for an agricultural downstream task demonstrate the potential of integrating xAI in the selection of the optimal hyperparameters. The results, computed on two state-of-the-art DL models, i.e., the 2D-CNN and the ViT, show that there is a correlation between the $Consistency^{JSdiv}$ metric and the macro F-score achieved on the test set. In future work, we plan to further investigate the relationship between the proposed $Consistency^{JSdiv}$ metric and the F-score achieved on the test set, with a detailed analysis of results at the class level. Additionally, we aim to apply the approach to optimize the selection of multiple hyperparameters using a grid-based method. Finally, we will test the method across a broader range of classification tasks and study areas to comprehensively assess its effectiveness and generalizability.

REFERENCES

[1] José M Bioucas-Dias, Antonio Plaza, Gustavo Camps-Valls, Paul Scheunders, Nasser Nasrabadi, and Jocelyn Chanussot, "Hyperspectral remote sensing data analysis and future challenges," *IEEE Geoscience and remote sensing magazine*, vol. 1, no. 2, pp. 6–36, 2013.

[2] Nassim Ait Ali Braham, Conrad M Albrecht, Julien Mairal, Jocelyn Chanussot, Yi Wang, and Xiao Xiang Zhu, "Spectralearth: Training hyperspectral foundation models at scale," *arXiv preprint arXiv:2408.08447*, 2024.

[3] Hao Yang, Haoyang Yu, Ke Zheng, Jiaochan Hu, Tingting Tao, and Qiang Zhang, "Hyperspectral image classification based on interactive transformer and cnn with multilevel feature fusion network," *IEEE Geoscience and Remote Sensing Letters*, 2023.

[4] Irfan Ahmad, Ghulam Farooque, Qichao Liu, Fazal Hadi, and Liang Xiao, "Mstsenet: Multiscale spectral–spatial transformer with squeeze and excitation network for hyperspectral image classification," *Engineering Applications of Artificial Intelligence*, vol. 134, pp. 108669, 2024.

[5] Linus Scheibenreif, Michael Mommert, and Damian Borth, "Masked vision transformers for hyperspectral image classification," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2023, pp. 2166–2176.

[6] Stefano Pignatti, Angelo Palombo, Simone Pascucci, Filomena Romano, Federico Santini, Tiziana Simoniello, Amato Umberto, Cuomo Vincenzo, Nicola Acito, Marco Diani, et al., "The prisma hyperspectral mission: Science activities and opportunities for agriculture and land monitoring," in *2013 IEEE International Geoscience and Remote Sensing Symposium-IGARSS*. IEEE, 2013, pp. 4558–4561.

[7] Emrullah ŞAHiN, Naciye Nur Arslan, and Durmuş Özdemir, "Unlocking the black box: an in-depth review on interpretability, explainability, and reliability in deep learning," *Neural Computing and Applications*, pp. 1–107, 2024.

[8] Caroline M Gevaert, "Explainable ai for earth observation: A review including societal and regulatory perspectives," *International Journal of Applied Earth Observation and Geoinformation*, vol. 112, pp. 102869, 2022.

[9] Jiarui Duan, Haoling Li, Haofei Zhang, Hao Jiang, Mengqi Xue, Li Sun, Mingli Song, and Jie Song, "On the evaluation consistency of attribution-based explanations," in *European Conference on Computer Vision*. Springer, 2025, pp. 206–224.

[10] Jinghui Yang, Anqi Li, Jinxi Qian, Jia Qin, and Liguo Wang, "A cross-attention-based multi-information fusion transformer for hyperspectral image classification," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2024.

[11] Wei Hu, Yangyu Huang, Li Wei, Fan Zhang, and Hengchao Li, "Deep convolutional neural networks for hyperspectral image classification," *Journal of Sensors*, vol. 2015, no. 1, pp. 258619, 2015.

[12] Etienne Vareille, Adel Abbas, Michele Linardi, and Vassilis Christophides, "Evaluating explanation methods of multivariate time series classification through causal lenses," in *10th IEEE International Conference on Data Science and Advanced Analytics, DSAA 2023, Thessaloniki, Greece, October 9-13, 2023*. 2023, pp. 1–10, IEEE.

[13] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby, "An image is worth 16x16 words: Transformers for image recognition at scale," 2021.

[14] Mukund Sundararajan, Ankur Taly, and Qiqi Yan, "Axiomatic attribution for deep networks," in *International conference on machine learning*. PMLR, 2017, pp. 3319–3328.

[15] Cristian Munoz, Kleyton da Costa, Bernardo Modenesi, and Adriano Koshiyama, "* Evaluating Explainability in Machine Learning Predictions through Explainer-Agnostic Metrics," Nov. 2024, arXiv:2302.12094 [cs].

[16] Sekou Dabo, Michele Linardi, and Claudia Paris, "https://github.com/ds4xai/xAIDrivenHyperOpt," 2025.

[17] Arthur Vandenhoeke, Lennert Antson, Guillem Ballesteros, Jonathan Crabbé, and Michal Shimoni, "Explaining the absorption features of deep learning hyperspectral classification models," in *IGARSS 2023-2023 IEEE International Geoscience and Remote Sensing Symposium*. IEEE, 2023, pp. 958–961.