Evaluation of the IGN FLAIR-HUB Model Transferability Performance for Land Cover Mapping in Iasi, Romania

Ana-Maria Loghin[1], Loredana-Mariana Crenganis[1], Constantin Stoian[2], Ana-Maria Olteanu-Raimond[2], Anatol Garioud[3], Valeria-Ersilia Oniga[1], Bogdan Rusu[1]

[1] Department of Terrestrial Measurements and Cadastre, Faculty of Hydrotechnical, Geodesy and Environmental Engineering, “Gheorghe Asachi” Technical University of Iaşi, 700050 Iasi, Romania – (ana-maria.loghin, loredana-mariana.crenganis, valeria-ersilia.oniga, bogdan.rusu3)@academic.tuiasi.ro

[2] Univ Gustave Eiffel, Géodata Paris, IGN, LASTIG, F-77454 Marne-la-Vallée, France – (constantin.stoian, ana-maria.raimond)@ign.fr

[3] Institut national de l’information géographique et forestière (IGN), France – (anatol.garioud)@ign.fr

Keywords: Semantic segmentation, Land cover, Transferability, Generalization, Spatial heterogeneity, FLAIR-HUB.

Abstract

Accurate land cover mapping at very high spatial resolution is essential for monitoring urban and environmental dynamics, yet the transferability of deep learning models across geographic domains remains a challenge. This study evaluates the transferability and generalization capability of the FLAIR-HUB semantic segmentation model, originally trained on French aerial imagery, when applied to the metropolitan area of Iași, Romania. The model is tested across multiple spatial resolutions (0.084 m, 0.2 m, and 0.5 m) and temporal scenarios (2024 vs. 2019), enabling a comprehensive assessment of cross-resolution and temporal robustness. A novel validation framework is introduced, combining conventional 2D raster-based evaluation with a 3D point-wise assessment using semantically labeled UAV-derived point clouds. The results demonstrate strong performance for dominant classes such as buildings and herbaceous vegetation, with improved accuracy at higher spatial resolution, while stable classes such as buildings and impervious surfaces show a comparatively robust performance, confirming the model’s capability to consistently represent invariant land cover types. However, performance decreases for heterogeneous and vegetation-related classes due to seasonal variability and class complexity. The 3D validation reveals slightly lower but consistent results, highlighting its role as a more rigorous evaluation approach. Overall, the study confirms the potential of transferring pre-trained semantic segmentation models to new geographic contexts, while emphasizing the importance of spatial resolution, temporal consistency, and validation strategy.

1. INTRODUCTION

Accurate Land Cover (LC) mapping using remote sensing technologies is essential for monitoring key anthropogenic development factors such as urban expansion, deforestation, and soil sealing, particularly in regions undergoing rapid and complex structural transformations, as observed in post-socialist Eastern Europe (Garioud et al., 2025, Kucsicsa et al., 2019, Kuemmerle et al., 2016). Reliable LC information supports spatial planning, environmental monitoring, and sustainable land management (Ranatunga et al., 2026, Barman et al., 2026). At continental scale, publicly available products such as the European CORINE Land Cover (CLC) database remain the primary source of harmonized Land Use/Land Cover (LULC) information (Feranec et al., 2016). However, despite continuous updates, CLC data are limited by their relatively coarse spatial resolution-historically 25 m and more recently 10 m—restricting their applicability for detailed urban and peri-urban analyses (Schultz et al., 2025, Rujoiu-Mare and Mihai, 2016). Recent developments aim to overcome these limitations by integrating multi-source data. For instance, the OSMlanduse dataset combines OpenStreetMap information with Sentinel-2 imagery to provide generalized LULC maps at 10 m resolution (Schultz et al., 2025). Similarly, large-scale products derived from satellite time series, such as those based on Sentinel-2 or Landsat data, have improved temporal consistency but still lack the spatial detail required for fine-scale mapping (Manakos and Braun, 2014). In this context, deep learning approaches—particularly convolutional and transformer-based architectures—have demonstrated significant potential for extracting high-resolution land cover information, although challenges remain regarding their transferability across different geographic regions and acquisition conditions (Garioud et al., 2025, Mo et al., 2022). The French

National Institute of Geographical and Forest Information (IGN) has addressed this need by developing the FLAIR-HUB dataset and its deep learning architectures, designed to fuse Very High Resolution (VHR) aerial imagery (0.2 m spatial resolution) with multi-temporal satellite data. However, this model’s capability for generalization remains largely untested in geographically and structurally distinct domains (Garioud et al., 2025, Garioud et al., 2023a). Therefore, our proposed research investigation analyses the ability of the FLAIR-HUB model to generalize to a novel geographic context represented by the Iași metropolitan area and its surroundings in Romania. As input, a UAV-based orthophoto data set from 2024 with 0.084 m spatial resolution is used. To the best of our knowledge, this research produces the first land cover map of Iași city at an unprecedented very-high resolution of 0.084 m for the year 2024. Compared with the existing OSMlanduse dataset (Schultz et al., 2025), our methodology firstly offers significantly higher spatial resolution, and secondly utilizes the fine-grained semantic segmentation capabilities of the FLAIR-HUB architecture to enable a precise focus solely on Land Cover elements. Furthermore, a detailed temporal robustness assessment will be performed by applying the trained network parameters to the Iasi orthophoto from 2019 at 50 cm resolution, thus comparing classification accuracy across significantly different VHR inputs. As a key methodological contribution, this study aims to validate the thematic accuracy of the final land cover map using UAV-based photogrammetrically semantically classified 3D point clouds.

Specifically, our research investigation has the main purpose in finding the answers to the following research questions:

1. What is the spatial generalizability performance of the FLAIR-HUB model, trained on French aerial data, when applied to the geographically and structurally distinct landscape in Iasi, Romania?
2. How to successfully transfer a semantic segmentation model, primarily trained on 0.20 m very-high-resolution aerial data, to datasets featuring significantly different spatial resolutions, such as finer (0.084 m) or coarser (0.5 m)?
3. What is the temporal repeatability of the transferred model, defined as the constancy of LC classification accuracy when moving from the very high resolution 2024 input data to the lower resolution 2019 orthophoto (0.084 m vs. 0.50 m)?

The manuscript is organized as follows: Section 2 describes the study area and datasets; Section 3 presents the methodological framework; Section 4 discusses the results; and Section 5 concludes the work and outlines future research.

2. STUDY AREA AND USED DATASETS

The study focuses on the city of Iași, located in northeastern Romania, and its peri-urban surroundings (Figure 1). The area represents a diverse landscape, combining dense urban zones, industrial areas, and agricultural land, which makes it suitable for testing the transferability and generalization of pretrained deep learning models for land cover mapping. The FLAIR-HUB semantic segmentation model is applied to recent UAV-based orthophotos derived from an aerial photogrammetric flight from 2024 with an Ultracam Eagle Mark 3 from Vexcel. The orthophotos have a spatial resolution of 0.084 m and are georeferenced in the national Stereo 70 coordinate system, with normal heights defined relative to the Black Sea vertical datum (EPSG:3844). For temporal analysis and to test robustness across variable input quality, the analysis also employs aerial-based orthophotos at a coarser 0.5 m spatial resolution, derived from a fotogrammetric flight in 2019.
A rigorous model validation is achieved through Unmanned Aerial Vehicle (UAV) photogrammetrically derived point clouds, which are manually labeled to match the Land Cover semantic classes defined by the pre-trained FLAIR-HUB model. The reference UAV dataset was acquired on May 23 2025 with a Matrice 350 RTK UAV system equipped with a SHARE penta camera model (PSDK 102S V3 445), flown at a height of 60 meters above ground level, covering an area of $0.2 km^2$ on the ground. The subsequent photogrammetric processing was performed using Agisoft Metashape, incorporating 28 Ground Control Points (GCPs) to ensure high spatial accuracy.

3. METHODOLOGICAL FRAMEWORK

The proposed methodological framework focuses on evaluating the cross-domain transfer of the pre-trained FLAIR-HUB semantic segmentation model to very high and coarser resolutions of orthophotos of 0.084 m and 0.50 m, in Iasi metropolitan area (Garioud et al., 2025, Garioud et al., 2023a, Garioud et al., 2023b). The $swin\_base\_patch4\_window12\_384$ encoder

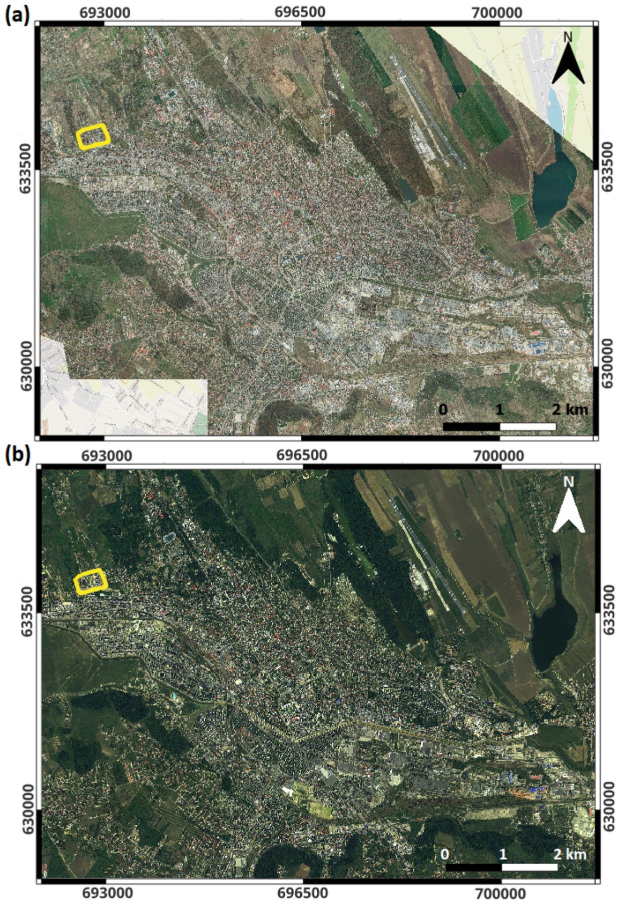

Figure 1. Overview of the orthophoto from 2024 (0.02 m) over the Iași metropolitan area, with marked location of the validation UAV dataset (yellow).

with $Aerial\_RGBI$ modality and the $upernet$ deocoder with $AERIAL\_LABEL - COSIA$ modality were considered in the FLAIR-HUB model architecture to detect the following 15 land cover classes: building, greenhouse, swimming pool, impervious surface, pervious surface, bare soil, water, snow, herbaceous vegetation, agricultural land, plowed land, vineyard, deciduous, coniferous, and brushwood. The model inference was performed on tiles of 2048 pixels with 512 pixel overlapping margins each to avoid edge artifacts.
Corresponding to each experimental scenario, specific input and output spatial resolutions were defined according to the characteristics of the respective orthophotos. For the 2024 dataset, two configurations were evaluated: (i) a native high-resolution setup with both input and output at 0.084 m, and (ii) a resampled configuration with input and output at 0.2 m. For the 2019 dataset, a multi-resolution approach was adopted, using an input resolution of 0.5 m and generating predictions at an output resolution of 0.2 m.

3.1 Proposed Workflow

The proposed workflow for evaluating the FLAIR-HUB model transferability in Iasi, Romania (Figure 2) comprises the following main steps:
(1) Data preparation and harmonization: this initial step ensures that both the high-resolution (0.084 m) and coarse-resolution (0.50 m) orthophotos from 2024 and 2019, respectively are well georeferenced in the same reference coordinate frame.

https://doi.org/10.5194/isprs-archives-XLIX-B1-2026-589-2026 |

(2) Model inference: the pre-trained FLAIR-HUB semantic segmentation model is applied to both input orthophotos at different spatial resolutions for generating land cover prediction maps.
(3) Validation: semantic segmentation results are quantitatively assessed against high-precision reference data derived from semantically labeled UAV 3D point clouds.

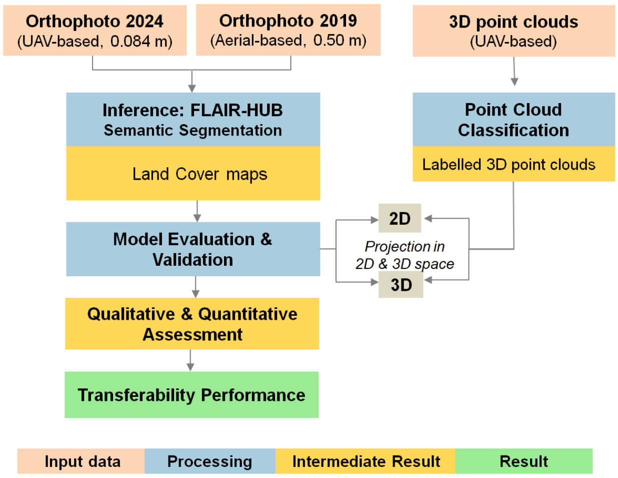

Figure 2. Overview of the workflow developed to evaluate the transferability performance of the IGN FLAIR-HUB model for land cover mapping in the Iași metropolitan area, Romania.

The comparison between predicted land cover maps and reference UAV-derived 3D semantic labels enables a rigorous quantitative assessment of classification performance and supports the evaluation of model transferability across varying spatial resolutions and temporal domains. The UAV-derived data supports two complementary validation approaches:

(1) a 2D approach, when the labeled 3D point clouds are rasterized to create a spatially aligned reference map. This will enable a direct pixel-level comparison against the predicted land cover maps for quantitative accuracy metrics.
(2) a 3D approach, when the 2D semantic labels produced by the segmentation model are projected onto the 3D point cloud, facilitating a point-wise evaluation of classification consistency in a complex, three-dimensional space.

3.2 Class Harmonization and Thematic Compatibility

Prior to comparative accuracy assessment, to ensure a consistent and statistically valid comparison between the classified orthophoto and the classified point cloud, a harmonized classification scheme was established . As the two datasets differ in their geometric properties, spatial resolution, and object detectability, certain object classes present in the point cloud were not explicitly represented in the orthophoto classification. Therefore, a semantic harmonization of classes was required.
In line with the labeling strategy adopted in the FLAIR model, movable and small-scale objects were not treated as independent land cover classes. Instead, they were assigned to the class of the underlying surface on which they are located. This approach ensures conceptual consistency between datasets and avoids introducing artificial discrepancies due to differences in object-level detectability. The original point cloud classification included additional object-level categories not explicitly defined in the orthophoto-based land cover schema, including: Vehicles (cars), Wall fences, Sealed grave surfaces within the cemetery area, Pathways with stairs, and Electrical towers. Since these objects do not represent independent land cover categories in the orthophoto classification, they were reassigned to semantically and functionally equivalent land cover classes based on the characteristics of the underlying surface.

The reclassification strategy was defined as follows:

- Vehicles (cars) were reassigned to impervious surfaces, as they represent transient objects whose labeling follows the surface on which they are positioned (e.g., asphalt roads), consistent with the adopted labeling strategy.
- Wall fences were aggregated into impervious surfaces or buildings, depending on their structural continuity and physical association with adjacent built structures, and their underlying ground context.
- Sealed grave surfaces, pathways, and stairs were reclassified as impervious surfaces, reflecting their artificial and non-permeable material characteristics.
- Electrical towers were reassigned to impervious surfaces, as their footprint corresponds to artificial ground surfaces and their limited spatial extent does not justify a separate class at the given raster resolution.

This aggregation strategy, based on the assignment of objects to their underlying land cover class, reduces artificial disagreement arising from differences in object detectability between 2D and 3D datasets and ensures thematic compatibility between the compared classifications.

3.3 Validation approaches

Before the quality and validation assessment step, all raster datasets were ensured to have identical spatial characteristics, including extent, spatial resolution, and coordinate reference system (projected national coordinate system Stereo 70, EPSG code 3844), so that each pixel corresponds spatially to the same ground location.

(1) 2D approach (pixel-based): This approach enabled a direct qualitative and quantitative comparison between the classified model output and the UAV reference dataset. For this, the UAV-based point attribute labels were projected into a regular two-dimensional grid representation using the Rasterize module of the OPALS software (Pfeifer et al., 2014). Therefore, the classification attributes were written to the output raster grid according to the planar (2D) location of each point. In cases where multiple points fall within a single grid cell, the dominant or representative class value was assigned according to the rasterization settings. To ensure consistency with the spatial resolution of the classified model output, the resulting raster dataset was generated at cell sizes of 0.2 m and 0.084 m. The raster layers were then compared on a pixel-by-pixel basis. NoData values were excluded from the analysis to avoid introducing bias from undefined areas.

(2) 3D approach (point cloud-based): The 2D semantic labels from the segmentation model are projected onto the 3D point cloud, enabling a point-wise evaluation in a three-dimensional space. The reference point cloud is manually labeled, and, depending on the respective input resolution of the models (0.2 m, 0.084 m, and 0.5 m), the corresponding predicted class labels are assigned to each point. The final evaluation is then performed by comparing the projected model-derived labels with the reference labels at the individual point level.

As not all classes defined in the classification scheme are present in the evaluated scene, the analysis was restricted to the classes occurring in the input datasets. This prevented bias in the confusion matrix and ensured that the reported metrics reflect only relevant classes. To evaluate the prediction performance of the model and counter class imbalance, metrics such as Intersection over Union (IoU), precision, recall, and F1-score were derived for each class. Global metrics for Overall Accuracy, mean and weighted IoU and F1-score were also computed.

4. RESULTS AND DISCUSSION

4.1 Reference data

The UAV dataset acquired in May 2025 was used to assess the semantic segmentation accuracy and transferability of the FLAIR-HUB model. From the SHARE Penta camera system, only nadir images were selected for further processing. The UAV imagery was processed following a standard photogrammetric workflow, beginning with the measurement of ground control points (GCPs) in the images, followed by image orientation and subsequent 3D point cloud reconstruction. The main acquisition parameters, results regarding ground control point (GCP) accuracy and photogrammetric processing are summarised in Table 1.
The resulting photogrammetrically derived 3D UAV point cloud was manually labeled into the corresponding land cover segmentation classes, matching the FLAIR-HUB model's output (Figure 3).

No. img	Nadir GSD [cm/px]	GCPs Accuracy Mean residual X[m]	Y[m]	Z[m]	Reproj. error [px]	3D point cloud [mil]
1310	1.24	0.02	0.02	0.16	0.20	4.34

Table 1. Reference UAV dataset: georeferencing accuracy and photogrammetric processing results.

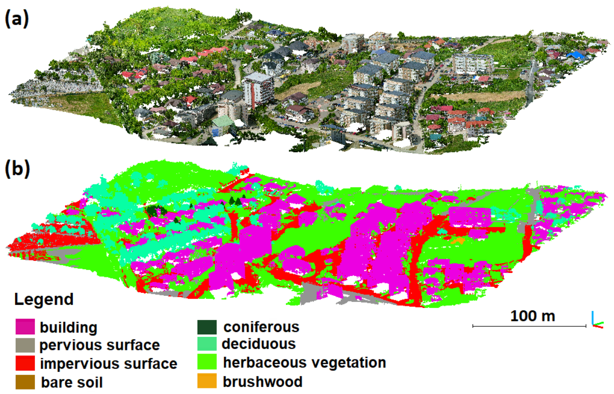

Figure 3. 3D-view of the UAV-photogrammetrically derived 3D point cloud: (a) RGB colors; (b) semantic class labels

4.2 FLAIR model semantic segmentation inference

The FLAIR-HUB semantic segmentation model was applied to the orthophotos over Iasi city at different dates and resolutions (Figure 4). Overall, the results for the two orthophotos acquired in 2024 (with spatial resolutions of 0.2 m and 0.084 m) show a high degree of consistency, particularly in the delineation of buildings, impervious surfaces, water bodies, and herbaceous vegetation. However, notable discrepancies are observed for coniferous areas as well as agricultural and plowed land land surfaces. In contrast, an improved representation of vineyards, deciduous forests, and agricultural land can be observed in the 2019 dataset. This can primarily be attributed to the acquisition period in summer, characterized by full foliage, whereas the 2024 dataset was acquired in April under leaf-off conditions, affecting the spectral and structural appearance of vegetation classes.

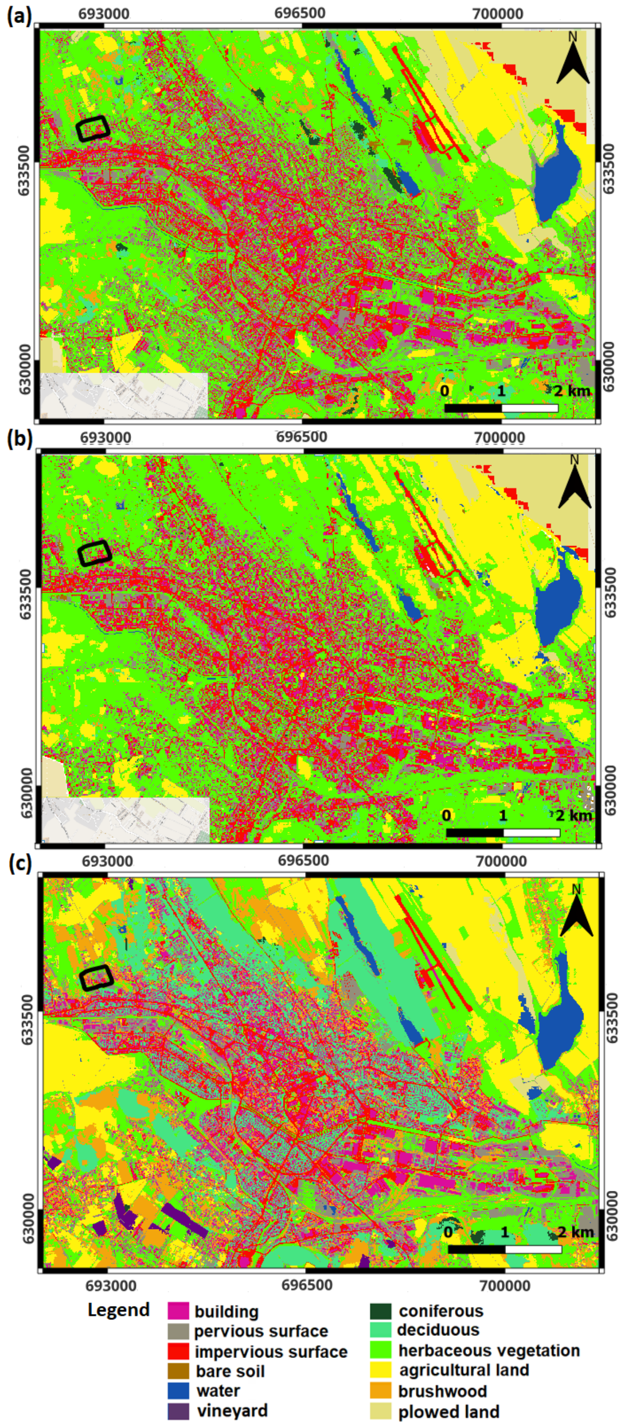

Figure 4. Results of the FLAIR-HUB model inference over the Iași metropolitan area on: (a) Orthophoto 2024 (0.2m); (b) Orthophoto 2024 (0.084m); (c) Orthophoto 2019 (0.5m).

The inference results for the UAV reference area are presented in Figure 5. Owing to the temporal discrepancies between the datasets, elements of change detection are inherently present; however, the analysis focuses primarily on temporally stable classes, particularly buildings and impervious surfaces. The detection of impervious paved graves varies across scenarios. These features are not identified in the 2019 dataset (input resolution 0.5 m) nor in the 2024 dataset at 0.2 m resolution, where they are either omitted or insufficiently represented. In contrast, at the higher spatial resolution (0.084 m), some of these structures are partially captured, although they are occasionally misclassified as buildings (Figure 6(a)). Overall, buildings and impervious surfaces are consistently well detected across all scenarios, with only minor deviations (Figure 6(b)), indicating robust model performance for structurally distinct and stable land cover classes. However, classification inconsistencies are observed in the 2019 dataset, likely due to differing radiometric properties. In this case, deciduous vegetation is frequently misclassified as brushwood, while certain surfaces, such as roads, are correctly identified as pervious, reflecting their unpaved condition at the time of acquisition. A clear influence of spatial resolution is observed for impervious surfaces, which are not fully captured at 0.2 m resolution but are more accurately delineated at 0.084 m (Figure 6(c)). Similarly, coniferous vegetation is primarily detected in the higher-resolution dataset, further demonstrating the benefit of finer spatial detail for class discrimination. Seasonal effects significantly impact vegetation classification. As the 2024 dataset was acquired under leaf-off conditions (April), many deciduous areas are misclassified as herbaceous vegetation, highlighting the sensitivity of spectral-based classification to phenological variation.

4.3 2D approach validation results

The validation results show that the FLAIR model reproduces the overall class distribution reasonably well for dominant classes, especially buildings and herbaceous vegetation (Figure 7, Table 2). Herbaceous vegetation is consistently overestimated, while impervious surfaces are systematically underestimated, indicating confusion with adjacent classes such as pervious or vegetation surfaces (Figure 8).
The higher-resolution dataset (8.4 cm) improves the agreement, especially for impervious surfaces and buildings, confirming the positive effect of finer spatial resolution on class discrimination. However, vegetation subclasses (deciduous) remain strongly underestimated, mainly due to seasonal differences.

Class	Reference UAV 0.2m	Predicted DOP 0.2m	Predicted DOP 0.084m
Herbaceous	47.01	52.97	51.42
Impervious	24.30	18.24	20.24
Buildings	14.60	14.39	14.68
Pervious	9.00	13.25	11.20
Deciduous	4.89	0.73	1.02
Coniferous	0.12	0.00	0.35
Brushwood	0.08	0.34	0.64

Table 2. Class distribution comparison between reference and predicted datasets (values given in %).

The resulting metrics indicate an overall robust performance of the model for well-defined and dominant classes, particularly buildings and herbaceous vegetation (Table 3). At 0.2 m resolution, buildings achieve high and balanced scores (F1 = 0.904, IoU = 0.824), demonstrating reliable detection of structurally distinct objects. Similarly, herbaceous vegetation shows strong performance (F1 = 0.876). The higher-resolution dataset (0.084 m) further supports these findings, showing improved agreement for most major classes, with the highest scores (F1 = 0.91, IoU = 0.835) for buildings class. Across all scenarios, minor and spectrally ambiguous classes (deciduous, brushwood, coniferous) show low performance, which is expected due to their limited spatial extent, class imbalance, and sensitivity to seasonal and phenological variation.

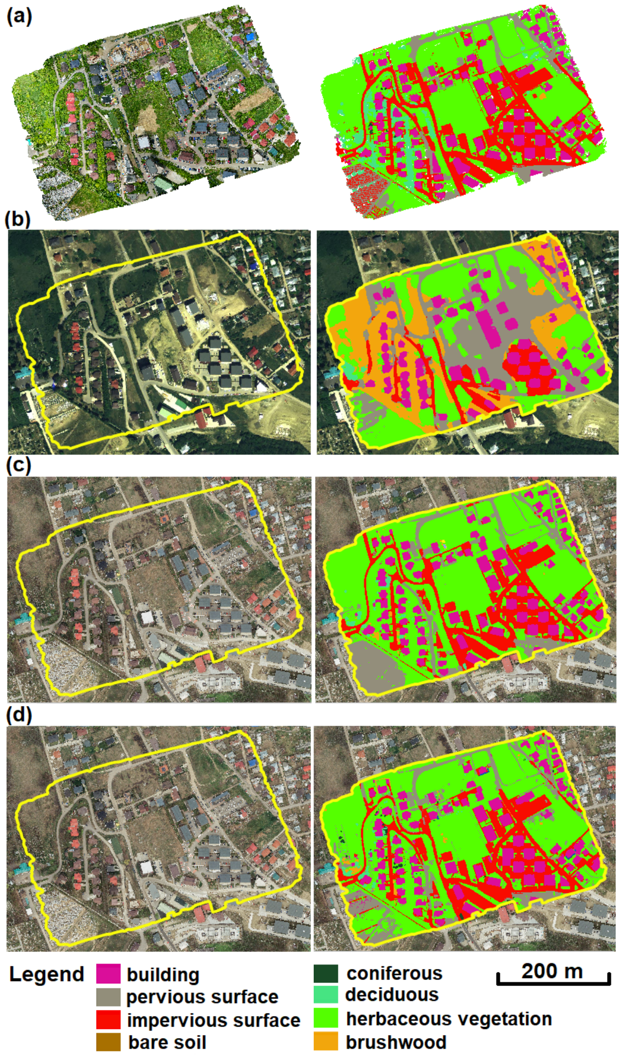

Figure 5. Results of the FLAIR-HUB model inference for the validation area: (a) UAV-based 3D point cloud 2025 with reference class labels; (b) Orthophoto 2019 (0.5 m); (c) Orthophoto 2024 (0.2 m); (d) Orthophoto 2024 (0.084 m).

Due to the five-year temporal discrepancy between the datasets, the analysis of the 2019 orthophoto was limited to stable land cover classes, namely buildings and impervious surfaces. All areas subject to temporal changes were excluded, including building construction or demolition, modifications of street surfaces (e.g., unpaved to paved) and land cover changes within the cemetery associated with its expansion.
A comparison of area differences (in %) between the predicted classification and the reference UAV labeled projections for the two stable classes (buildings and impervious) across all scenarios is shown in Figure 9. The results show that impervious surfaces are consistently underestimated across all scenarios, with larger deviations at coarser resolution (-6.06%) and improved agreement at finer resolution (-1.60%). This behavior can be partly attributed to the thematic harmonization, as the UAV-derived reference raster includes additional elements such as sealed grave surfaces, walls, and stairs within the impervious class, increasing its complexity and extent. In contrast, the building class shows only minor deviations (¡+/-0.3%), indicat-

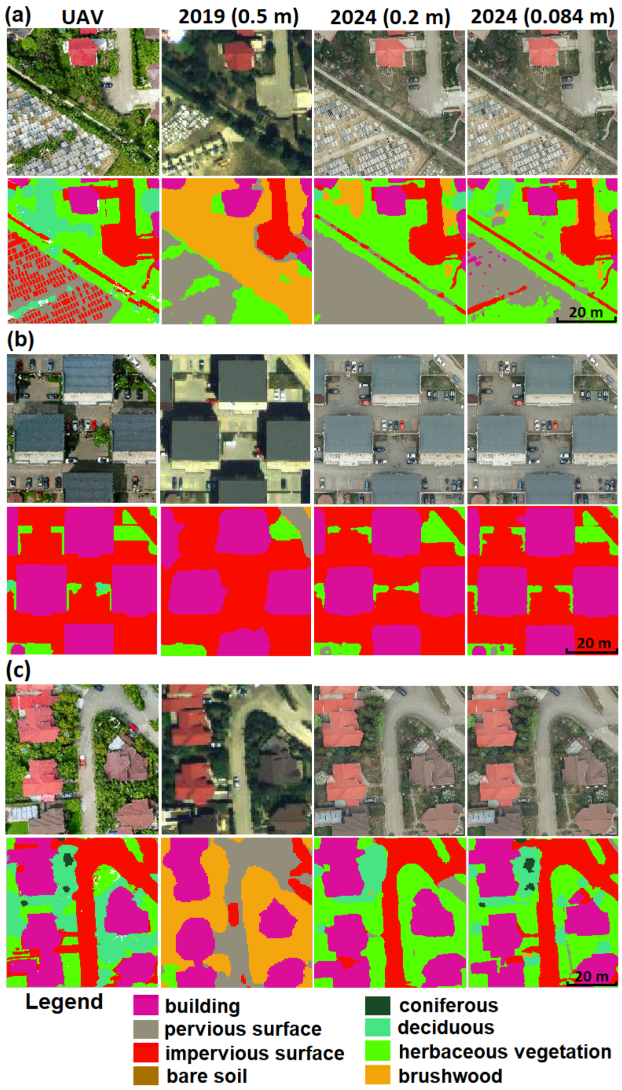

Figure 6. Results of the FLAIR-HUB model inference - detail view: (a) Cemetery; (b) Built-up area; (c) Street, houses.

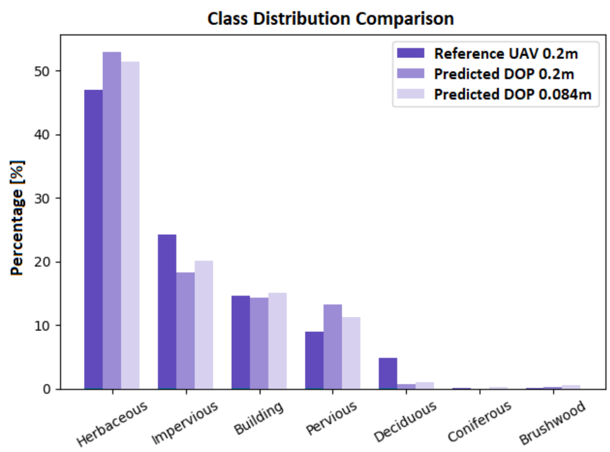

Figure 7. Class distribution comparison between UAV-reference and model predictions at 0.20 m and at 0.084 m resolutions.

ing a more robust and stable classification performance. Overall, the results highlight the influence of spatial resolution and

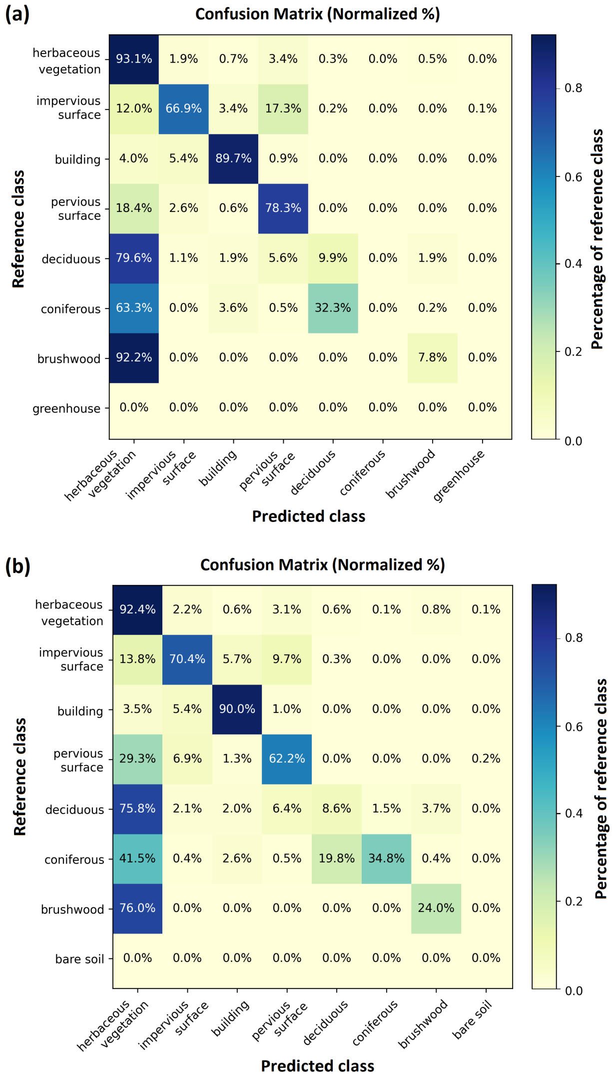

Figure 8. Normalized confusion matrices for: (a) FLAIR prediction 0.2 m resolution; (b) FLAIR prediction 0.084 m resolution. Values represent pixel frequencies (%) per reference class, with diagonal elements indicating classification accuracy and off-diagonal elements misclassification patterns.

class definition on classification accuracy, particularly for heterogeneous surface types such as impervious areas.

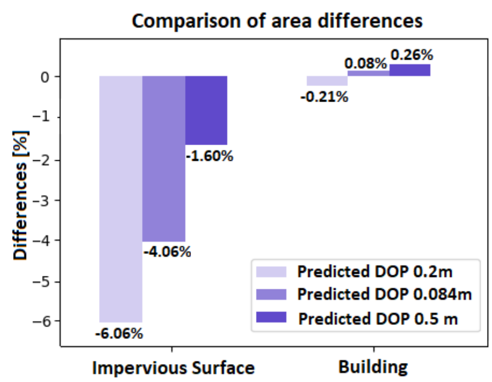

Figure 9. Comparison of area differences (values in %).

Class	Prec	Recall	F1	IoU	IoU-Baseline
DOP 2024 0.2m					
Buildings	91.0	89.7	**90.4**	**82.4**	**83.9**
Impervious	89.2	66.9	76.5	61.9	**75.7**
Herbaceous	82.6	93.1	87.6	**77.9**	**54.3**
Pervious	53.2	78.3	63.3	46.3	**57.2**
Coniferous	0.0	0.0	0.0	0.0	**62.6**
Deciduous	66.7	9.90	17.3	9.50	**71.7**
Brushwood	1.9	7.8	3.1	1.6	**30.2**
DOP 2024 0.084m					
Buildings	90.8	91.3	**91.0**	**83.5**	**83.9**
Impervious	87.9	73.1	79.8	**66.5**	**75.7**
Herbaceous	82.5	89.9	86.1	75.5	**54.3**
Pervious	50.5	66.4	57.4	40.2	**57.2**
Coniferous	15.2	39.9	22.0	12.3	**62.6**
Deciduous	54.1	11.3	18.6	10.3	**71.7**
Brushwood	2.7	20.2	4.8	2.4	**30.2**

Table 3. Validation results for selected classes using DOP 2024 orthophotos at different spatial resolutions (values given in %).

4.4 3D approach validation results

The 3D validation results confirm the performance observed in the 2D analysis, with consistent class distributions for buildings and herbaceous vegetation (Figure 10, Table 4) and improved performance for the higher spatial resolution scenario. Class buildings achieves the highest and most consistent scores across both resolutions (F1 = 90.5–91.1, IoU up to 83.7), indicating robust detection of structurally well-defined objects (Table 5). Herbaceous vegetation also shows high recall and stable performance (IoU up to 64.4), while impervious surfaces benefit from higher resolution, improving to IoU = 60.3 at 0.084 m. Particular improvement is pronounced at the Coniferous class, with IoU increasing from near-zero to 16.0%, demonstrating that higher spatial detail enables better discrimination of fine-scale canopy structures. Similar to the 2D results, minor and spectrally ambiguous classes remain poorly detected, reflecting class imbalance and sensitivity to structural and radiometric variability.

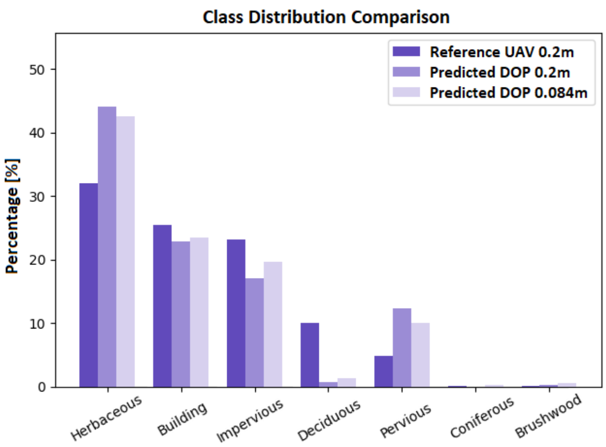

Figure 10. Class distribution comparison between UAV-reference and model predictions at 0.20 m and at 0.084 m resolutions.

4.5 Discussion

While the 3D validation results are broadly consistent with the 2D evaluation, they reveal a slightly reduced performance in the global metrics: overall accuracy, mean and weighted scores

Class	Reference UAV 0.2m	Predicted DOP 0.2m	Predicted DOP 0.084m
Herbaceous	32.91	45.40	42.89
Buildings	25.74	23.16	24.03
Impervious	23.21	17.37	19.97
Deciduous	10.88	1.17	1.69
Pervious	6.72	12.43	10.12
Coniferous	0.35	0.03	0.58
Brushwood	0.18	0.44	0.70

Table 4. Class distribution comparison between reference labels and projected predictions in the 3D point cloud (values in %).

Class	Prec	Recall	F1	IoU	IoU-Baseline
DOP 2024 0.2m					
Buildings	95.5	86.0	**90.5**	**82.7**	**83.9**
Impervious	82.6	61.8	70.7	54.6	**75.7**
Herbaceous	67.6	93.2	78.3	**64.4**	**54.3**
Pervious	43.1	79.8	55.9	38.8	**57.2**
Coniferous	1.1	0.1	0.2	0.1	**62.6**
Deciduous	81.2	8.7	15.8	8.6	**71.7**
Brushwood	2.1	5.2	3.0	1.5	**30.2**
DOP 2024 0.084m					
Buildings	94.0	88.3	**91.1**	**83.7**	**83.9**
Impervious	82.9	68.8	75.2	**60.3**	**75.7**
Herbaceous	67.4	89.1	76.7	62.2	**54.3**
Pervious	42.6	64.8	51.4	34.6	**57.2**
Coniferous	21.9	37.0	327.5	**16.0**	**62.6**
Deciduous	71.1	11.3	19.6	10.9	**71.7**
Brushwood	4.3	17.8	6.9	3.6	**30.2**

Table 5. Validation results for selected classes using DOP 2024 orthophotos at different spatial resolutions (values given in %).

((Table 6)). Lower efficiency can also be observed for several classes, particularly impervious and pervious surfaces, indicating a higher sensitivity to projection uncertainties at the point level. In contrast, buildings remain robust across both approaches, confirming that structurally distinct classes are less affected by the transition from 2D to 3D evaluation. These differences highlight that, although the overall trends are preserved, the 3D assessment provides a more rigorous evaluation framework, especially for heterogeneous and less well-defined classes.
The comparison between 0.2 m and 0.084 m spatial resolutions shows that increasing spatial resolution leads to a consistent improvement in segmentation performance. Improvement can be observed at the individual class level, especially for the stable classes such as buildings (IoU increase from 82.4% to 83.5% and from 82.7% to 83.7% for the 2D and 3D validation approaches, respectively) and impervious surfaces (IoU increase from 61.9% to 66.5% and from 54.6% to 60.3% for the 2D and 3D validation approaches, respectively). Moreover, the global metrics show a substantially improvement of the mean IoU (from 39.94% to 41.53% and from 35.80% to 38.84% for the 2D and 3D validation approaches, respectively) and mean F1-score (from 48.31% to 51.38% and from 44.91% to 49.77% for the 2D and 3D validation approaches, respectively), indicating a better class-wise balance (Table 6).

For the most representative classes in the validation UAV dataset, namely Herbaceous, Buildings, and Impervious, the IoU values show mixed agreement with the high-quality benchmarks provided by the FLAIR-HUB dataset, which is based on very-high-resolution (20 cm) annotations (Garioud et al., 2025). Our results for the Building class (up to 83.7%) are very closely aligned with the reported baseline (83.9%), confirming the effectiveness of transformer-based architectures in capturing structural textures. However, a noticeable discrepancy could be ob-

Approach		O.A.	mIoU	wIoU	mF1	wF1
	Baseline	**77.50**	**64.10**			
2D	**0.2m**	80.68	39.94	70.62	48.31	79.48
	0.084m	80.10	41.53	69.58	51.38	79.41
3D	**0.2m**	73.47	35.80	58.71	44.91	71.20
	0.084m	73.82	38.84	62.14	49.77	72.46

Table 6. Global evaluation for Land Cover Segmentation - Comparison of validation approaches (values given in % for: Overall Accuracy, mean and weighted Intersection Over Union, mean and weighted F1-score).

served for the Impervious Surface class, where our results (66.5%) are with 9.2% lower than the reported value (75.7%). This difference is largely attributable to the semantic harmonization required for our validation dataset, assigning additional objects such as vehicles (cars), wall fences, sealed grave surfaces, pathways with stairs, and electrical towers to the impervious class. In our case, the model struggled to recognize these specific integrated objects — particularly the sealed grave surfaces, which were not identified at all — leading to a lower IoU for the broader impervious category. In contrast, our findings for Herbaceous Vegetation show a significant improvement over the FLAIR-HUB baseline, reaching 77.9% compared to the reported IoU value of 54.3%. Finally, our Overall Accuracy (80.68%) demonstrates strong generalization capabilities, exceeding the 78.2% reported by the most complex multimodal configuration which integrated aerial, elevation, and satellite data(Garioud et al., 2025).

5. CONCLUSIONS AND FUTURE WORK

This study investigated the cross-domain transferability of the FLAIR-HUB semantic segmentation model to a geographically distinct environment in Iași, Romania, while evaluating its robustness across varying spatial and temporal conditions. The results demonstrate that the model generalizes well to new geographic contexts, particularly for dominant and structurally well-defined classes such as buildings and herbaceous vegetation. Despite the domain shift between Western and Eastern European landscapes, the model maintains high classification performance, confirming that learned geometric and textural features can be effectively transferred beyond the original training domain.

Regarding the transfer across spatial resolutions, the analysis shows that the model can be successfully applied to both finer (0.084 m) and coarser (0.5 m) datasets, although with varying performance. An increase in spatial resolution consistently improves both class-wise and global metrics, enhancing the delineation of complex surfaces such as impervious areas. Conversely, coarser resolution data leads to a noticeable degradation in accuracy due to reduced spatial detail and increased class mixing. At the same time, very high resolution introduces additional intra-class variability, which may result in over-segmentation and misclassification of fine-scale objects.

The temporal transferability assessment highlights that model performance decreases when applied to datasets acquired at different times, particularly due to seasonal and radiometric differences. Vegetation classes are especially affected, with deciduous areas frequently confused with herbaceous or brushwood categories under leaf-off conditions. Nevertheless, stable classes such as buildings and impervious surfaces remain comparatively robust, confirming the model's ability to maintain consistent performance for invariant land cover types.

A key contribution of this work is the introduction of a dual 2D–3D validation framework based on semantically labeled UAV point clouds. While the 3D evaluation yields slightly lower performance metrics due to increased sensitivity to spatial inconsistencies, it provides a more rigorous and realistic assessment of classification quality in complex environments. This demonstrates that combining 2D and 3D validation approaches enhances the reliability and interpretability of semantic segmentation results.

Future work should focus on improving model adaptation to heterogeneous classes and temporal variability by incorporating multi-seasonal training data and refining class definitions to better capture complex land cover types. Additionally, integrating 3D information directly into the training process, rather than only in validation, could further enhance classification robustness. Finally, extending the approach to larger areas and additional geographic regions would provide further insights into the scalability and generalization capabilities of deep learning-based land cover mapping frameworks.

References

Barman, P., Mustak, S., Shammas, M. I., Pradhan, B., Chakraborty, S., Maitra, S., 2026. A Multi-scale Object-based Framework for Hierarchical Urban Land Use and Land Cover Classification Using Ensemble Machine Learning and Multi-source Geospatial Data. *Earth Systems and Environment*, 1–29.

Feranec, J., Soukup, T., Hazeu, G., Jaffrain, G., 2016. *European landscape dynamics: CORINE land cover data*. CRC Press.

Garioud, A., De Wit, A., Poupée, M., Valette, M., Giordano, S., Wattrelos, B., 2023a. FLAIR# 2: textural and temporal information for semantic segmentation from multi-source optical imagery. *arXiv preprint arXiv:2305.14467*.

Garioud, A., Giordano, S., David, N., Gonthier, N., 2025. FLAIR-HUB: Large-scale Multimodal Dataset for Land Cover and Crop Mapping. *arXiv preprint arXiv:2506.07080*.

Garioud, A., Gonthier, N., Landrieu, L., De Wit, A., Valette, M., Poupée, M., Giordano, S., Wattrelos, b., 2023b. Flair : a country-scale land cover semantic segmentation dataset from multi-source optical imagery. A. Oh, T. Naumann, A. Globerson, K. Saenko, M. Hardt, S. Levine (eds), *Advances in Neural Information Processing Systems*, 36, Curran Associates, Inc., 16456–16482.

Kucsicsa, G., Popovici, E.-A., Bălteanu, D., Grigorescu, I., Dumitraşcu, M., Mitrică, B., 2019. Future land use/cover changes in Romania: regional simulations based on CLUE-S model and CORINE land cover database. *Landscape and ecological engineering*, 15(1), 75–90.

Kuemmerle, T., Levers, C., Erb, K., Estel, S., Jepsen, M. R., Müller, D., Plutzar, C., Stürck, J., Verkerk, P. J., Verburg, P. H. et al., 2016. Hotspots of land use change in Europe. *Environmental research letters*, 11(6), 064020.

Manakos, I., Braun, M., 2014. *Land use and land cover mapping in Europe*. 18, Springer.

Mo, Y., Wu, Y., Yang, X., Liu, F., Liao, Y., 2022. Review the state-of-the-art technologies of semantic segmentation based on deep learning. *Neurocomputing*, 493, 626–646.

Pfeifer, N., Mandlburger, G., Otepka, J., Karel, W., 2014. OPALS–A framework for Airborne Laser Scanning data analysis. *Computers, Environment and Urban Systems*, 45, 125–136.

Ranatunga, S., Strand Ødegård, R., Jetlund, K., Godager, B. A., Storie, C., Storie, J., 2026. From remote sensing to decision-making: a web-based platform for near-real-time land cover classification. *International Journal of Digital Earth*, 19(1), 2639785.

Rujoiu-Mare, M.-R., Mihai, B.-A., 2016. Mapping land cover using remote sensing data and GIS techniques: A case study of Prahova Subcarpathians. *Procedia Environmental Sciences*, 32, 244–255.

Schultz, M., Li, H., Wu, Z., Wiell, D., Auer, M., Alexander, Z., 2025. OSMlanduse a dataset of European Union land use at 10 m resolution derived from OpenStreetMap and Sentinel-2. *Scientific Data*, 12(1), 750.