Pixel-based vegetation mapping at class-level from UAV multispectral imagery: application in an alpine lake ecosystem

Mohammad Elahi[1], Alessandra Spadaro[2], Francesca Matrone[2], Andrea Maria Lingua[2], Chiara Graziani[2], Vittorio Fra[1]

[1] DIST, Interuniversity Department of Regional and Urban Studies and Planning, Politecnico di Torino, Viale Pier Andrea Mattioli, 39, Torino (mohammad.elahi, vittorio.fra)@polito.it

[2] DIATI, Department of Environment, Land and Infrastructure Engineering, Politecnico di Torino, Corso Duca degli Abruzzi 24, Torino (francesca.matrone, andrea.lingua, alessandra.spadaro, chiara.graziani)@polito.it

Keywords: UAV multispectral imagery, vegetation mapping, data-centric approach, feature engineering, XGBoost

Abstract

Vegetation mapping in alpine environments is essential for monitoring ecosystem dynamics and climate change impacts, yet remains challenging when using very high-resolution UAV imagery under limited labeled data. This study proposes a data-centric, pixel-based classification framework for class-level vegetation mapping using multispectral UAV data acquired in an alpine study area. The approach prioritizes improving data representation rather than increasing model complexity. To address label scarcity, a feature-rich dataset was constructed by integrating spectral information, vegetation indices, and lightweight spatial descriptors to enhance class separability. Classification was performed using XGBoost, which is well suited for multispectral tabular data and robust under imbalanced conditions. The results show consistent classification performance across vegetation types and demonstrate the effectiveness of dataset enrichment under limited supervision, highlighting the importance of feature representation in data-scarce scenarios.

1. Introduction

Vegetation mapping is essential for understanding ecosystem dynamics, biodiversity patterns, and climate-change impacts, particularly in sensitive alpine environments. In these systems, strong topographic gradients, short growing seasons, and rapid environmental shifts generate marked ecological heterogeneity, making accurate high-resolution vegetation mapping a key requirement for monitoring and conservation.

In recent years, unmanned aerial vehicles (UAVs) have become increasingly important in plant ecology because they enable flexible acquisitions at centimetre-level spatial resolution, while multispectral sensors provide ecologically meaningful information for vegetation discrimination when radiometric acquisition and preprocessing are carefully controlled (Assmann et al., 2019; Sun et al., 2021). However, the data scarcity of vegetation samples and the complexity of mountainous environments could increase the challenge of vegetation mapping.

1.1 State of the art

Recent studies have shown that UAV imagery can support detailed vegetation and habitat mapping using pixel-based, object-based, and machine-learning workflows, including both classical algorithms and deep learning approaches (Hamylton et al., 2020; Bhatnagar et al., 2020; Belcore et al., 2021). More broadly, deep learning has rapidly expanded across remote-sensing applications and has achieved strong performance in several classification and segmentation tasks (Ma et al., 2019). However, fine-grained vegetation mapping from very high-resolution UAV imagery remains challenging. At centimetre scale, spectral signatures often overlap among classes, within-class variability may be high, and vegetation patches are frequently fragmented and intermixed, reducing separability and complicating robust generalization (Lopatin et al., 2017; Sun et al., 2021).

Despite this progress, an important scientific gap remains. Much of the existing UAV vegetation literature provides application-specific demonstrations or method comparisons, but offers limited guidance for fine-grained classification in ecologically heterogeneous environments when only a small amount of reliable reference data is available. This limitation becomes even more relevant when the objective is not simply broad land-cover discrimination, but ecologically meaningful vegetation mapping at very high thematic detail, including habitat-oriented or species-oriented analyses (Belcore et al., 2021; Demichele et al., 2025). In addition, methodological discussion is still often framed in model-centric terms, whereas recent data-centric literature argues that substantial performance gains may derive from improving the quality, representativeness, and structure of the data and feature space rather than from increasing model complexity alone (Zha et al., 2025). At the same time, work in ecological and geospatial machine learning has shown that non-spatial validation strategies may yield over-optimistic estimates when spatial dependence is ignored, highlighting the need for caution when interpreting pixel-based classification results (Meyer et al., 2019; Stock et al., 2023).

In this study, the term data-centric is used in an operational sense. Rather than prioritizing a more complex model architecture, we focus on improving the informativeness of the input representation through ecologically interpretable feature engineering. Concretely, this means combining multispectral bands with vegetation indices and lightweight local spatial descriptors, while maintaining a classifier that is simple, scalable, and well suited to tabular data. This choice is motivated by the specific conditions of the study: a very large UAV orthomosaic, limited and spatially clustered reference data, and the need to perform efficient full-scene inference.

To answer this question, we apply the proposed workflow to multispectral UAV data acquired in the alpine ecosystem in northwestern Italy of the ACLIMO project. The contribution of this paper is threefold: (1) it proposes a data-centric classification workflow tailored to UAV vegetation mapping under scarce supervision; (2) it evaluates the role of an enriched feature space combining spectral, index-based, and local spatial information; and (3) it discusses the potential and limitations of pixel-based full-scene inference in the presence of limited labels and spatial dependence.

2. The ACLIMO project

This research was developed within the ACLIMO project, funded by the cross-border Interreg VI-A France-Italy ALCOTRA 2021-2027 programme. The project addresses climate-change adaptation in alpine protected areas, with particular attention to water availability, alpine wetlands and aquatic environments to address the resilience of climate-sensitive habitats and species.

The project context is directly relevant to the objectives of this study. In fact, alpine lakes and associated hydric reserve environments are highly responsive to hydrological and climatic variability, and therefore require monitoring approaches capable of capturing fine-scale ecological patterns. In this context, UAV multispectral data provides an effective tool for acquiring high-resolution observations over spatially complex and difficult-to-access mountain environments where data acquisition campaigns could be difficult and limited by accessibility issues.

2.1 Study area

The study area (Lake Vej del Bouc) is illustrated in Figure 1, and it is located on the border of Italy and France (northwestern Italy). Within the whole study, the research is divided across multiple scales of analysis: a) small scale comprising the area of the Protected Areas of Maritime Alps (APAM) to be investigated by satellite images, b) medium scale constituted by a single valley (Gesso valley) to be studied with aerial orthophotos (resolution of 50-30 cm) and c) large scale with the Lake Vej del Bouc at 2042 m a.s.l., to be examined with UAV images. In this contribution, only the large scale is deepened.

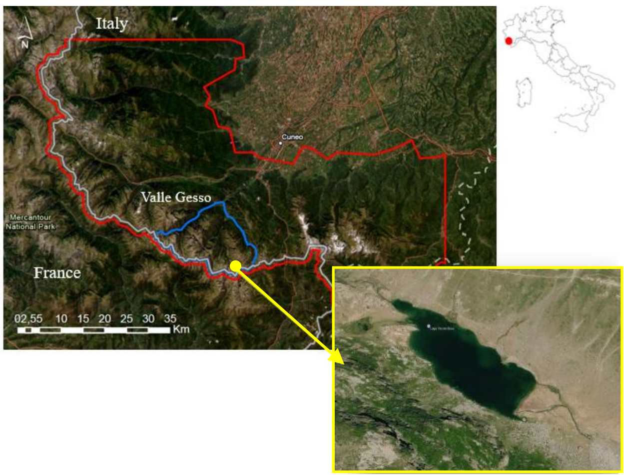

Figure 1. Study area at multiple scales, showing the region of interest of the Protected Areas of Maritime Alps (red), intermediate analysis scale of a single valley (blue), and the Vej del Bouc lake area (yellow).

This alpine environment is characterized by high ecological variability, complex vegetation patterns, and strong topographic gradients. The study area thus includes multiple vegetation communities with fragmented and spatially intermixed distributions, providing a representative and demanding test case for pixel-based classification using UAV multispectral imagery.

3. Materials and methods

3.1 Methodology

The overall methodological workflow of the proposed data-centric, pixel-based classification framework is illustrated in Figure 2. The approach integrates UAV multispectral imagery, image preprocessing, feature construction, and supervised machine learning for class-level vegetation mapping. Under conditions of extreme label scarcity, model-centric approaches such as deep learning are less suitable due to their dependence on large annotated datasets and their susceptibility to overfitting. Therefore, the proposed framework focuses on improving feature representation and data quality to enhance class separability.

The workflow consists of four main stages: (1) data acquisition and preprocessing, including masking of non-vegetated areas, (2) feature construction through spectral bands, vegetation indices, and texture descriptors, (3) reference data preparation and sampling under extreme label scarcity, and (4) model training and full-scene inference using XGBoost.

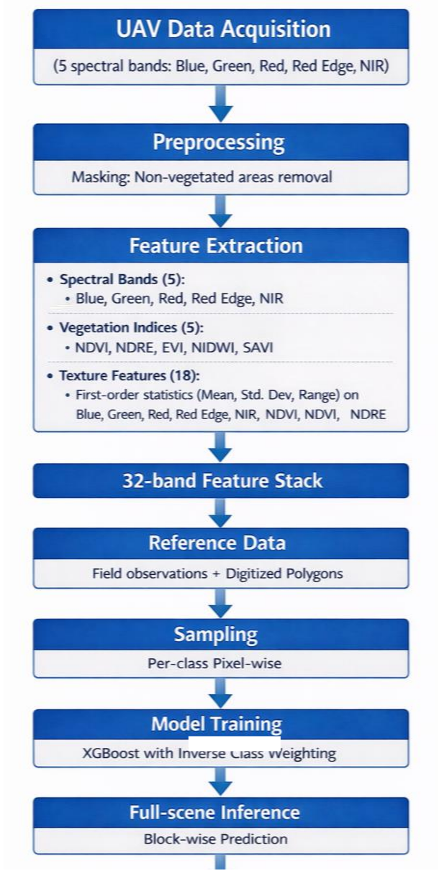

Figure 2. Methodological workflow of the proposed data-centric vegetation classification framework.

3.2 Data acquisition

The UAV data were acquired using the DJI Mavic 3 Multispectral (M3M), a multispectral imaging system for high-resolution environmental monitoring. The acquisition campaign was conducted on July 23rd, 2025, under stable illumination and atmospheric conditions to ensure radiometric consistency across the dataset, also supported by the platform's integrated sunlight sensor for irradiance recording.

https://doi.org/10.5194/isprs-archives-XLIX-B2-2026-657-2026

The sensor captures four spectral bands (Green, Red, Red Edge, and Near-Infrared), enabling the characterization of vegetation properties at very fine spatial scales. In particular, the multi-spectral unit is composed by four 5 MP sensors centered at 560 nm (Green), 650 nm (Red), 730 nm (Red Edge), and 860 nm (Near-Infrared). The platform also integrates a 20 MP RGB camera (with Blue band) with a mechanical shutter and an RTK positioning system. Since these bands have diverse dynamics, sensibility and resolution, we kept only the Blue band for the EVI definition. The flight mission was designed to ensure high image overlap and complete coverage of the study area, supporting accurate photogrammetric reconstruction.

The imagery was processed using a standard Structure-from-Motion photogrammetric workflow, including image alignment, dense point cloud generation, and orthorectification. 1224 images have been processed, acquired with an altitude between 20 and 50 meters during midday, with 8 GCPs (Ground Control Points) evenly distributed around the lake and a final RMSE of 2.9 cm for a total area of about 1.5 Km^2. The radiometric calibration was carried out through the “use sun sensor” option on Agisoft Metashape. A digital surface model (DSM) was derived from the dense reconstruction and used during the orthorectification process to correct geometric distortions caused by terrain variations.

Multispectral orthophotos were generated for each spectral band (RGB image for blue band, green, red, Red Edhe and NIR from radiometric sensors) and mosaicked into a single georeferenced raster dataset with a spatial resolution of 3 cm. Particular attention was given to preserving spatial alignment across spectral bands to avoid co-registration errors.

The resulting orthomosaic serves as the input for feature extraction and subsequent classification.

3.3 Dataset

The final UAV dataset consists of a georeferenced multispectral raster with a spatial resolution of 3 cm, comprising 849,598,608 pixels in total. After applying vegetation masking, 237,882,024 pixels were retained as valid for analysis (28.00% of the total dataset).

A total of 12 vegetation classes were defined based on field observations and image interpretation. Reference polygons were rasterized to generate labeled samples, resulting in 378,748 labeled pixels. This corresponds to 0.16% of the valid dataset and 0.04% of the total image.

These figures highlight the extreme imbalance between labeled and unlabeled data, defining a severe label scarcity scenario. This structural limitation represents a key challenge for supervised classification, as the available training data covers only a very small fraction of the spatial and spectral variability present in the full dataset.

3.4 Data characteristics and challenges

The dataset is characterized by several structural challenges that directly affect the classification task.

First, the data volume is extremely large, consisting of hundreds of millions of pixels, while only a very small fraction is labeled. This results in an extreme label scarcity scenario, where supervised learning must be performed under highly constrained conditions.

Second, the dataset exhibits strong class imbalance. A small number of dominant vegetation classes, such as Paniculate grass, Matgrass, Green Alder, and Cotton Grass, account for a large proportion of labeled samples, whereas several minority classes contribute less than 2.00% of the labeled data. In addition, even the most represented classes occupy only a negligible fraction of the full raster, indicating that the dataset is both globally sparse and locally imbalanced. The class-wise distribution of labeled pixels and their proportion within the valid raster are summarized in Table 1.

To mitigate this issue, inverse class weighting was applied during model training, assigning higher importance to underrepresented classes and improving sensitivity to minority categories.

Third, the centimeter-level spatial resolution introduces high spectral variability and strong intra-class heterogeneity. Variations in illumination, shadows, and vegetation structure further reduce class separability at the pixel level.

Finally, the spatial distribution of labeled samples is highly fragmented, making it difficult to design spatially independent training and testing strategies. This constraint increases the risk of spatial leakage, which may lead to optimistic performance estimates.

ID	Class	% of labelled data	% of valid raster
1	Adenostyles	1.44%	0.0023%
2	Alpine Dock	1.15%	0.0018%
3	Black Sedge	0.98%	0.0016%
4	Blueberry	1.55%	0.0025%
5	Cotton Grass	12.87%	0.0205%
6	Dwarf Juniper	11.68%	0.0186%
7	Goosefoot	2.20%	0.0035%
8	Green Alder	13.29%	0.0212%
9	Male Fern	4.32%	0.0069%
10	Matgrass	16.71%	0.0266%
11	Paniculate grass	20.26%	0.0323%
12	Rhododendron	5.40%	0.0086%
-	Unknown	30,846	0.0130%

Table 1.Class-wise distribution of labeled pixels and their proportion within the valid raster, highlighting severe class imbalance and dataset sparsity.

3.5 Data preprocessing and masking

To restrict the analysis to vegetated areas, a threshold based on NDVI values (NDVI > 0.30) was applied, following established approaches for identifying actively growing vegetation (Walz and Weber, 2021). In addition, water bodies were manually excluded to avoid spectral interference from non-vegetated surfaces.

The resulting masked raster is shown in Figure 3. This preprocessing step reduced the dataset to 237,882,024 valid pixels, corresponding to 28.00% of the total image. By excluding non-vegetated regions, the analysis is focused on ecologically relevant areas, improving class separability and reducing spectral variability unrelated to vegetation.

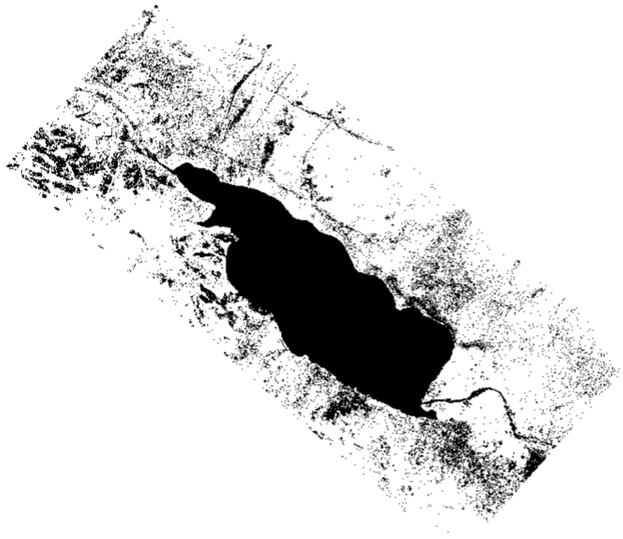

Figure 1. Masked raster showing valid vegetated pixels after NDVI thresholding (NDVI > 0.30) and water removal.

3.6 Feature space design

The feature space was designed to enhance class separability under extreme label scarcity by integrating complementary sources of information, including spectral bands, vegetation indices, and lightweight spatial descriptors.

First, four multispectral bands (Green, Red, Red Edge, and Near-Infrared) were combined with five vegetation indices (NDVI, NDRE, EVI2, NDWI, and SAVI), providing a spectral representation of vegetation properties. NDWI has been computed with the formula by McFeeters (1996), which does not include the SWIR band, which is not available here.

To incorporate local spatial context within the pixel-based framework, first-order statistical features (mean, standard deviation, and range) were computed using a 3×3 moving window on selected spectral bands and indices (Green, Red, Red Edge, NIR, NDVI, and NDRE). These descriptors capture fine-scale spatial variability while preserving local structures. The use of a 3×3 window represents a trade-off between incorporating spatial context and avoiding excessive smoothing.

In total, the feature space consists of four spectral bands, five vegetation indices, and eighteen texture features derived from six base layers using three statistical measures, resulting in a final 27-band feature stack.

Higher-order texture descriptors, such as Gray-Level Co-occurrence Matrix (GLCM) features, were not included due to their high computational cost and limited scalability for datasets of this size. In contrast, first-order statistics provide a computationally efficient alternative that is consistent with the data-centric strategy adopted in this study.

3.7 Reference data and sampling strategy

Reference data were collected through a field survey using a GNSS receiver (accuracy positioning of 4-5 cm) (Figure 4) and expert-based image interpretation. Due to the limited accessibility of the alpine environment, only a small number of ground-truth samples were available, with approximately three reference points per vegetation class. In detail: each species was located and identified in the field using at least three samples. Subsequently, regions of interest (ROIs) were delineated visually on the orthophotos, thereby including multiple pixels for each class. In this case, GNSS accuracy did not affect class annotation, as it is on the order of a few centimeters, whereas the point acquired with the survey pole was always taken at the center of much larger samples (at least 50 cm, as shown in Figure 4).

These samples were manually digitized as polygons to better capture local spatial variability and subsequently rasterized to align with the feature stack. An example of the manually digitized reference polygons and their fragmented spatial distribution is shown in Figure 5.

Given the limited number of labeled samples and their fragmented spatial distribution, a per-class pixel-wise sampling strategy was adopted to ensure adequate representation of both dominant and minority classes.

Figure 4. Samples of vegetation acquired in the field.

Figure 5. Example of manually digitized reference polygons (red) used for training, illustrating the sparse and fragmented distribution of labeled samples.

As mentioned above, this approach introduces a known risk of spatial leakage, as spatially adjacent pixels in the training and testing sets may share similar spectral and spatial characteristics, potentially leading to too positive performance estimates.

However, spatially independent sampling strategies were not feasible in this context, as they would significantly reduce the already limited training data and further exacerbate class imbalance, particularly for minority classes.

Therefore, a trade-off was adopted between statistical independence and sufficient class representation, prioritizing model stability and balanced learning under extreme label scarcity conditions.

3.8 Model training and evaluation

The labeled dataset was split into training (70.00%), validation (10.00%), and testing (20.00%) subsets independently for each class using a fixed random seed (seed = 42). This resulted in 302,994 training pixels, 37,875 validation pixels, and 75,754 testing pixels.

Pixel-level feature vectors were extracted and organized into a tabular dataset. Feature standardization was applied using parameters derived exclusively from the training data.

Classification was performed using XGBoost (version 3.1.3) with GPU acceleration (CUDA) to ensure computational efficiency for large-scale data processing. The model was trained for up to 2,500 boosting iterations, with the optimal performance achieved at iteration 2,499. Inverse class weighting was applied to mitigate the effects of class imbalance.

The training and validation learning curves are shown in Figure 6, illustrating stable convergence and a limited generalization gap. Model selection was based on the Macro F1-score, which provides a balanced evaluation across all classes and is particularly suitable for imbalanced multi-class datasets.

Two configurations were evaluated, comparing a baseline spectral–index feature set with an enriched representation including spatial descriptors.

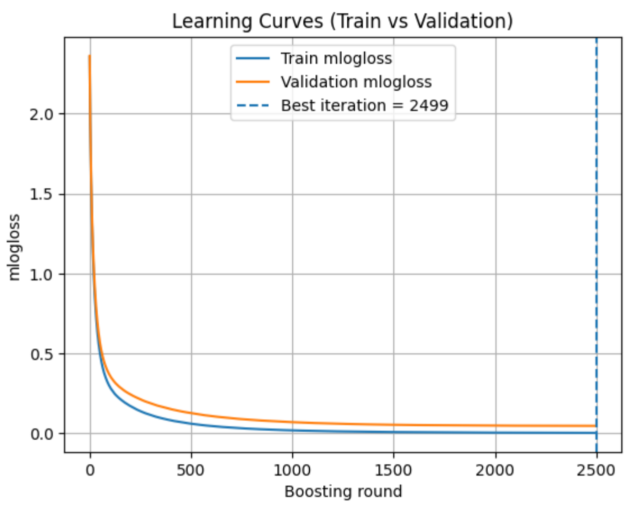

Figure 6. Training and validation learning curves of the XGBoost model (mlogloss), showing stable convergence and limited generalization gap.

3.9 Full-scene inference

Model inference was applied to the full dataset using a tiled (block-wise) processing strategy, enabling efficient handling of large UAV imagery without exceeding memory constraints. In total, 237,882,024 valid pixels were processed across 3,304 blocks, ensuring scalability to large spatial extents.

This approach allows the model to be applied to the entire study area while maintaining computational efficiency and consistent prediction behavior across tiles.

4. Results

4.1 Model performance

The UAV dataset contains 849,598,608 pixels, of which 237,882,024 were retained as valid after masking (28.00% of the total dataset). Rasterization of ground-truth polygons produced 378,748 labeled pixels across 12 vegetation classes, corresponding to 0.16% of the valid dataset, confirming the extreme label scarcity scenario.

A per-class stratified sampling strategy was applied, and the model was trained using pixel-level feature vectors. The training process showed stable convergence, with a training mlogloss of 0.0025 and a validation mlogloss of 0.0464, resulting in a generalization gap of 0.0439. This indicates controlled training behavior without significant overfitting despite the limited and imbalanced training data.

The classification achieved strong performance under these constraints. The final model reached an Overall Accuracy (OA) of 0.99 and a Macro F1-score of 0.98, as summarized in Table 2.

Metric	Value
Overall Accuracy (OA)	0.988
Macro F1-score	0.981
Balanced Accuracy (BA)	0.979
Cohen’s Kappa	0.986
Matthews Correlation Coefficient (MCC)	0.986

Table 2. Overall classification performance metrics.

Performance was consistent across all vegetation classes, including minority categories, with all classes achieving F1-scores above 0.96, as detailed in Table 3. This indicates balanced classification behavior across both dominant and sparsely represented classes, despite the strong class imbalance in the labeled dataset.

Class	Precision	Recall	F1
Adenostyles	0.97	0.97	0.97
Alpine Dock	0.97	0.95	0.96
Black Sedge	0.98	0.97	0.97
Blueberry	0.97	0.97	0.97
Cotton Grass	0.99	0.99	0.99
Dwarf Juniper	0.98	0.98	0.98
Goosefoot	0.98	0.98	0.98
Green Alder	0.99	0.99	0.99
Male Fern	0.99	0.98	0.98
Matgrass	0.99	0.99	0.99
Paniculate grass	1.00	1.00	1.00
Rhododendron	0.98	0.98	0.98

Table 3. Per-class precision, recall, and F1-score.

This contribution has been peer-reviewed.
https://doi.org/10.5194/isprs-archives-XLIX-B2-2026-657-2026

4.2 Comparison with previous classification approach

The proposed pixel-based classification framework was compared with baseline machine learning approaches previously evaluated within the ACLIMO project, as well as the object-based classification workflow presented in (Matrone et al., 2025), with data from July 2024. The comparison has been carried out only on the common classes between the two years. Since the distribution may vary slightly from one year to the next, the comparison has been treated more as a reference than a benchmark, and tests comparing data and methods across only 2024 and 2025 are currently under investigation.

These baseline methods include Bayesian classification, Random Forest, Support Vector Machines (SVM), K-Nearest Neighbors (KNN), and Random Tree models. Among these, the best-performing baseline achieved an Overall Accuracy of 94% (Bayesian classifier), while other methods showed lower performance (Random Forest: 90%, SVM: 89%, kNN: 87%, Random Tree: 85%).

In comparison, the proposed approach achieved an Overall Accuracy (OA) of 0.98 and a Macro F1-score of 98%, representing a substantial improvement over all baseline methods. This improvement can be primarily attributed to the enriched feature representation, which integrates spectral information, vegetation indices, and spatial descriptors. While baseline approaches rely mainly on spectral information, the proposed feature space enhances class separability, particularly under extreme label scarcity conditions.

In addition to quantitative improvements, visual comparison with the previous Bayesian classification results (Figure 7) indicates more consistent spatial structures and reduced fragmentation compared to the OBIA-based approach (Matrone et al., 2025). The proposed framework also offers improved scalability by avoiding the need for segmentation, resulting in a simpler and more efficient workflow.

Overall, the results suggest that, under extreme data scarcity, improving feature representation can have a greater impact on classification performance than increasing model or workflow complexity.

4.3 Spatial patterns and full-scene inference

The trained model was applied to all valid pixels (237,882,024; 28.00% of the total dataset) to generate a full-scene vegetation map (Figure 8). At the landscape scale, the classification reveals coherent spatial patterns, where a limited number of dominant vegetation types occupy large contiguous areas, while less frequent classes appear in fragmented and spatially dispersed distributions. These patterns are consistent with the ecological organization of alpine vegetation.

At finer spatial scales, localized inconsistencies and small fragmented regions are observed, particularly in transition zones and areas characterized by spectral similarity and structural complexity. These effects reflect residual class ambiguity and highlight inherent limitations of pixel-based classification in very high-resolution imagery.

A discrepancy between test-set performance and full-scene inference is also observed. This behavior can be attributed to structural characteristics of the dataset, including limited spatial coverage of labeled samples and spatial leakage introduced by pixel-wise sampling, which may lead to overestimated metrics.

Comparison with object-based image analysis (OBIA) approaches (Matrone et al., 2025) indicates that the proposed framework achieves comparable spatial patterns while offering a simpler and more scalable workflow without requiring segmentation.

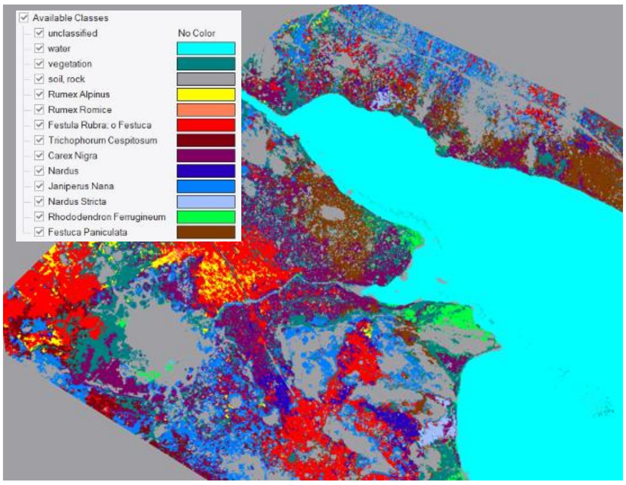

Figure 7. Previous classification of the vegetation classes through the Bayesian algorithm (Matrone et al., 2025).

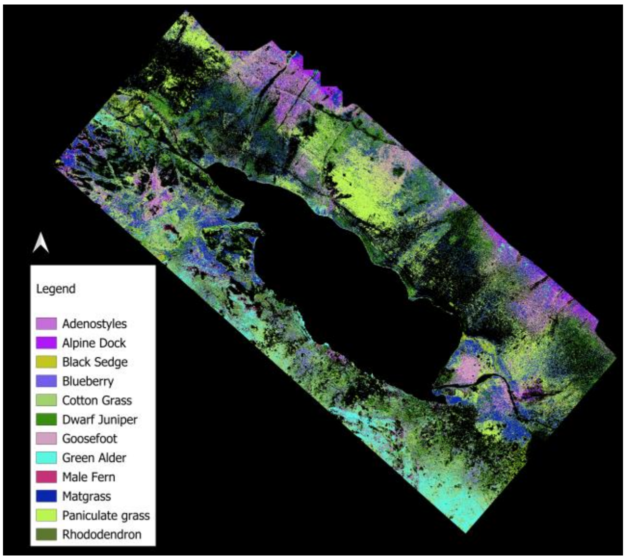

Figure 8. Full-scene vegetation classification map obtained using the XGBoost classifier, showing the spatial distribution of 12 vegetation classes over the study area.

5. Discussion

The results show that, for very high-resolution UAV-based vegetation mapping under extreme label scarcity, the main limitation is structural rather than algorithmic. In other words, performance depends more on how features are represented than on how complex the model is. This highlights the value of data-centric approaches in situations where labeled data are limited.

Even with a very small proportion of labeled samples (0.04% of the total image), the framework achieves stable and balanced performance across all vegetation classes. This suggests that improving the feature space can effectively compensate for limited training data and reduce the impact of class imbalance. By combining spectral information, vegetation indices, and spatial descriptors, the model benefits from complementary sources of information, which improves class separability.

At centimeter-level resolution, spectral overlap and variability within the same class become more pronounced due to fine-scale heterogeneity, such as changes in illumination, canopy structure, and background conditions. Because of this, using only spectral information is often not enough for reliable classification. Adding spatial descriptors helps address this issue by capturing local context, leading to more consistent and stable predictions without increasing computational complexity.

These results are in line with previous studies that emphasize the importance of feature representation in remote sensing (Hamylton et al., 2020). Compared to object-based image analysis (OBIA) approaches (Matrone et al., 2025), which rely on segmentation and classifier selection, the proposed pixel-based framework achieves similar performance while being simpler and more scalable.

However, some limitations remain. Pixel-based classification naturally produces salt-and-pepper noise because it does not include explicit spatial constraints, especially in transition areas and regions with similar spectral characteristics. In addition, there is a noticeable difference between test-set performance and full-scene inference. This is mainly due to limited spatial coverage of labeled samples, spatial leakage caused by pixel-wise sampling, and radiometric variability across the scene.

In terms of generalization, the framework is expected to perform well under similar environmental and acquisition conditions. However, it may be sensitive to changes in sensor properties or preprocessing steps, which highlights the need for consistent data preparation.

Future work should focus on incorporating spatial context more directly, for example by using spatially aware learning methods or hybrid approaches that combine machine learning and deep learning (Bhatnagar et al., 2020). This could improve boundary delineation and reduce local inconsistencies.

6. Conclusions

This study presents a data-centric, pixel-based framework for classes-level vegetation mapping using very high-resolution UAV multispectral imagery in an alpine environment.

The results show that accurate and balanced classification is possible even under extreme label scarcity, where only a very small fraction of pixels is available for training. Combining spectral information, vegetation indices, and lightweight spatial descriptors significantly improves both class separability and spatial consistency, without increasing model complexity.

The proposed framework offers a scalable and efficient solution for large UAV datasets, making it possible to process hundreds of millions of pixels while maintaining stable performance across vegetation classes. This makes it suitable for large-scale ecological monitoring in alpine environments.

However, spatial inconsistencies still exist, especially in the form of salt-and-pepper noise, which reflects the limitations of pixel-based classification without explicit spatial constraints.

Future work should focus on integrating spatial context through spatially aware or hybrid approaches to improve boundary delineation and reduce local fragmentation.

Overall, despite the weak validation of this application (to be further enhanced in future works), the results seem to confirm that under extreme data scarcity, improving feature representation is more effective than increasing model complexity for vegetation mapping tasks.

Acknowledgments

The data has been acquired in collaboration with the operators of APAM (Aree Protette Alpi Marittime, Protected Areas of Maritime Alps) who supported the surveying campaigns.
The study is also partially funded by the project FAIR - Future Artificial Intelligence Research and received funding from the European Union Next-Generation EU (PIANO NAZIONALE DI RIPRESA E RESILIENZA (PNRR FAIR, CUP n. E13C22001800001) – MISSIONE 4 COMPONENTE 2, INVESTIMENTO 1.3 – D.D. 1555 11/10/2022, PE00000013). This manuscript reflects only the authors' views and opinions; neither the European Union nor the European Commission can be considered responsible for them. The authors would also like to acknowledge Eng. Saeed Hemmatianzadeh for the ML classification tests and the student team DIRECT of the Politecnico di Torino for supporting data acquisition activities.

References

Assmann, J.J., Kerby, J.T., Cunliffe, A.M., Myers-Smith, I.H., 2019. Vegetation monitoring using multispectral sensors - best practices and lessons learned from high latitudes. *Journal of Unmanned Vehicle Systems*, 7(1), 54-75. https://doi.org/10.1139/juvs-2018-0018.

Belcore, E., Pittarello, M., Lingua, A.M., Lonati, M., 2021. Mapping Riparian Habitats of Natura 2000 Network (91E0*, 3240) at Individual Tree Level Using UAV Multi-Temporal and Multi-Spectral Data. *Remote Sensing*, 13(9), 1756. https://doi.org/10.3390/rs13091756.

Bhatnagar, S., Gill, L., Ghosh, B., 2020. Drone image segmentation using machine and deep learning for mapping raised bog vegetation communities. Remote Sens., 12(16), 2602. https://doi.org/10.3390/rs12162602.

Demichele, D., Belcore, E., Piras, M., Camporeale, C., 2025. Species-by-Species Pattern Analysis of Coastal Dune Vegetation. *Journal of Geophysical Research: Biogeosciences*, 130(2), e2024JG008419. https://doi.org/10.1029/2024JG008419.

Hamylton, S.M., Morris, R.H., Carvalho, R.C., Roder, N., Barlow, P., Mills, K., Wang, L., 2020. Evaluating techniques for mapping island vegetation from unmanned aerial vehicle (UAV) images: Pixel classification, visual interpretation and machine learning approaches. Int. J. Appl. Earth Obs. Geoinf., 89, 102085. https://doi.org/10.1016/j.jag.2020.102085.

Lopatin, J., Fassnacht, F.E., Kattenborn, T., Schmidtlein, S., 2017. Mapping plant species in mixed grassland communities using close range imaging spectroscopy. *Remote Sensing of Environment*, 201, 12-23. https://doi.org/10.1016/j.rse.2017.08.031.

Ma, L., Liu, Y., Zhang, X., Ye, Y., Yin, G., Johnson, B.A., 2019. Deep learning in remote sensing applications: A meta-analysis and review. *ISPRS Journal of Photogrammetry and Remote Sensing*, 152, 166-177. https://doi.org/10.1016/j.isprsjprs.2019.04.015.

McFeeters, S.K., 1996. The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features. International journal of remote sensing, 17(7), pp.1425-1432.

Matrone, F., Graziani, C., Spadaro, A., Hemmatianzadeh, S., Lingua, A.M., 2025. UAV and AI for large-scale monitoring of the effects of climate change on mountain lake areas: the ACLIMO cross-border cooperation project. Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., XLVIII, 203–210.

Meyer, H., Reudenbach, C., Wöllauer, S., Nauss, T., 2019. Importance of spatial predictor variable selection in machine learning applications - Moving from data reproduction to spatial prediction. *Ecological Modelling*, 411, 108815. https://doi.org/10.1016/j.ecolmodel.2019.108815.

Stock, A., Gregr, E.J., Chan, K.M.A., 2023. Data leakage jeopardizes ecological applications of machine learning. *Nature Ecology & Evolution*, 7(11), 1743-1745. https://doi.org/10.1038/s41559-023-02162-1.

Sun, Z., Wang, X., Wang, Z., Yang, L., Xie, Y., Huang, Y., 2021. UAVs as remote sensing platforms in plant ecology: review of applications and challenges. *Journal of Plant Ecology*, 14(6), 1003-1023. https://doi.org/10.1093/jpe/rtab089.

Walz, J.H., Weber, K.T., 2021. Identifying actively growing vegetation using NDVI threshold values. GIS Training and Research Center, Idaho State University. https://giscenter.isu.edu/Research/Techpg/NASA_ISGC/pdf/AGV_threshold_final.pdf

Zha, D., Bhat, Z.P., Lai, K.-H., Yang, F., Jiang, Z., Zhong, S., Hu, X., 2025. Data-centric Artificial Intelligence: A Survey. *ACM Computing Surveys*, 57(5), Article 129, 1-42. https://doi.org/10.1145/3711118