

Politecnico di Torino

Master Degree course in Digital Skills for Sustainable Societal Transitions
Master Degree Thesis

High-Resolution UAV-Based Vegetation classes Mapping Using Data-Centric Machine Learning Under Extreme Label Scarcity

Supervisors:

Francesca Matrone
Vittorio Fra
Alessandra Spadaro

Candidate:

Mohammad Elahi

Academic Year 2025-2026

Acknowledgements

I would like to express my sincere gratitude to my supervisors, Prof. Francesca Matrone, Prof. Vittorio Fra, and Prof. Alessandra Spadaro, whose guidance, expertise, and continuous support were essential throughout the development of this thesis. I am also deeply thankful to the professors of my program at Politecnico di Torino, whose dedication and knowledge have significantly contributed to my academic and professional growth.

Above all, I extend my heartfelt thanks to my parents, my brother, and my sister for their unconditional love and endless patience throughout my entire educational journey. Their belief in me has always been my greatest source of strength.

I also want to warmly thank my friends and classmates who stood by me through every challenge and every achievement along this path. You made this experience feel less like a long road and more like a shared journey and for that, I am truly grateful.

Finally, I wish to thank all the professors who have taught me over the years, both before and during my studies at Politecnico di Torino, whose commitment and passion for knowledge have profoundly shaped who I am today.

Abstract

Accurate vegetation classification from very high-resolution UAV imagery remains a fundamental challenge in environmental monitoring and the management of aquatic ecosystems. Although multispectral UAV data provide detailed spectral information, factors such as spectral overlap among vegetation species, intra-class structural heterogeneity, and the severe scarcity of reference data particularly under pixel-based approaches can significantly limit classification reliability. These challenges are especially pronounced in complex alpine environments, where vegetation patterns are highly heterogeneous and dynamic.

This study develops a pixel-based machine learning framework for vegetation classification using multispectral UAV imagery acquired over an alpine lake environment. The proposed workflow follows a data-centric approach, including preprocessing and vegetation masking, feature engineering based on spectral bands and vegetation indices, and supervised classification using a gradient boosting model. To enhance the descriptive power of the input data, first-order statistical texture features derived from local neighborhoods are incorporated into the feature space, complementing the spectral representation with spatial information.

The integration of spatial texture information enables the model to better capture structural differences between vegetation types, particularly in areas characterized by strong spectral similarity. Beyond numerical performance, the approach contributes to improved spatial consistency of the classification outputs, reducing fragmentation and enhancing the interpretability of the resulting maps.

Overall, the study highlights the importance of feature-space design and data representation in high-resolution UAV-based vegetation mapping, particularly under conditions of limited reference data. The proposed framework demonstrates that combining spectral features, vegetation indices, and computationally efficient texture descriptors within a lightweight machine learning pipeline provides a practical and scalable solution for vegetation classification in complex and data-scarce environments.

Keywords: UAV multispectral imagery; vegetation species mapping; pixel-based classification; data-centric machine learning; extreme label scarcity; XGBoost; texture feature enrichment

Table of Contents

1. Introduction 1
1.1 Background and motivation 1
1.2 Problem statement 2
1.3 Research objectives 3
1.4 Novel contributions 4
1.5 Thesis structure 5
2. Literature review 5
2.1 UAV-based vegetation mapping 6
2.2 Spectral bands and vegetation indices 7
2.3 Texture features in high-resolution remote sensing 8
2.4 Texture-based vegetation classification 9
2.4.1 Previous work on the same study areas 10
2.5 Machine learning for vegetation classification 11
2.6 Research gap 12
2.7 Data-centric approaches in remote sensing 12
3. Methodology and data 13
3.1 Overview of the proposed methodology 13
3.2 Data acquisition 15
3.2.1 Study area description 15
3.2.2 Environmental and ecological context 16
3.2.3 UAV image acquisition 17
3.2.4 Multispectral data characteristics 17
3.2.5 Reference data and ground truth 19
3.2.6 Independent validation Dataset (Lake Vej del Bouc) 21
3.3 Data processing 23
3.3.1 Image preprocessing and mask generation 23
3.3.2 Feature engineering 24
3.3.3 Training sample selection and data split strategy 26
3.3.4 XGBoost model and class weighting 28
3.3.5 Experimental design 28
3.3.6 Accuracy assessment metrics 29
3.3.7 Reproducibility notes 29

3.3.8 Data-centric enhancement strategy 30
3.4 Post-processing strategy 30
3.4.1 Majority filtering 31
3.4.2 Small patch removal 31
4 Results 32
4.1 Baseline classification results (9-band Dataset) 32
4.1.1 Model convergence behavior 32
4.2 Impact of texture features 33
4.3 Transferability results 34
4.3.1 Classification performance on independent validation dataset (27-band Model) 35
4.4 Spatial analysis and salt-and-pepper noise 38
4.4.1 Spatial behavior in case Study 1 (9-band vs 27-band) 39
4.4.2 Spatial behavior in independent validation dataset (27-band) 39
4.4.3 Spatial leakage analysis 42
4.4.4 Discrepancy between test metrics and full-scene inference 43
4.5 Post-processing for spatial refinement 45
4.5.1 Majority filtering (3×3 and 5×5) 45
4.5.2 Small patch removal (<250 px and <500 px) 46
4.5.3 Comparative impact of post-processing methods 48
4.6 Summary of results 49
4.6.1 Case study 1 (Baseline vs Texture) 49
4.6.2 Independent validation dataset (27-band + spatial refinement) 50
5. Discussion 51
5.1 Interpretation of main findings 51
5.2 Role of texture features 52
5.3 Comparison with similar studies 53
5.4 Robustness, class imbalance and generalization 53
5.5 Strengths and limitations 54
5.6 Data-centric vs. Model-centric perspective 54
5.7 Implications for UAV-based vegetation mapping 55
6. Conclusions and outlook 56
6.1 Conclusions and final methodological insight 56
6.2 Scientific and practical contributions 57

6.3 Limitations and future research directions 57
6.4 Generalization across case studies 58
References 60

1. Introduction

1.1 Background and motivation

Environmental monitoring and ecosystem assessment have become central topics in Earth and environmental sciences, driven by the increasing impacts of climate change, land-use transformation, and anthropogenic pressure (Zhu et al., 2018; Matrone et al., 2025). Among ecosystem components, vegetation cover is widely recognized as a sensitive indicator of ecological conditions, as its spatial distribution and temporal dynamics reflect underlying climatic and environmental processes (Persello and Bruzzone, 2016). Consequently, vegetation monitoring plays an essential role in the sustainable management and conservation of terrestrial ecosystems and is closely aligned with **Sustainable Development Goal 15 (Life on Land)**.

Remote sensing has emerged as a key tool for vegetation monitoring across multiple spatial scales, particularly with the rapid development of unmanned aerial vehicle (UAV) platforms. Multispectral UAV imagery provides very high spatial resolution data, enabling detailed observation of fine-scale spatial patterns and spectral variability (Gini et al., 2018). These capabilities make UAV-based observations especially suitable for complex and heterogeneous environments and support **Sustainable Development Goal 9 (Industry, Innovation and Infrastructure)**, which promotes innovative technologies for environmental monitoring (Feng et al., 2015).

Despite these advantages, the analysis of very high-resolution UAV imagery presents several challenges. Such datasets often contain hundreds of millions of pixels, making pixel-based classification computationally demanding. In addition, spectral overlap between vegetation species, intra-class variability, and environmental factors such as shadows, moisture differences, and illumination variability can reduce the effectiveness of purely spectral, per-pixel approaches (Persello and Bruzzone, 2016; Zhu et al., 2018).

These challenges are particularly evident in alpine aquatic and riparian ecosystems. Mountain lake environments are highly dynamic and sensitive to climatic variations, including temperature, precipitation, and water-level changes (Matrone et al., 2025). Vegetation in these areas plays a key role in ecosystem stability, water quality, and biodiversity, and its spatial patterns can provide early signals of environmental change. In this context, the present study also contributes to **Sustainable Development Goal 13 (Climate Action)**, by supporting improved monitoring of vulnerable alpine ecosystems.



Figure 1 - Sustainable Development Goals (SDGs) relevant to the study

1.2 Problem statement

Despite substantial advances in the availability of very high-resolution multispectral UAV imagery, accurate vegetation classification in mountainous aquatic ecosystems remains a challenging task. A primary difficulty arises from spectral overlap among vegetation species, where pixels belonging to different classes exhibit similar reflectance responses across spectral bands. This phenomenon, widely reported in high-resolution remote sensing studies, leads to increased class confusion and limits the discriminative power of purely spectral features (Gini et al., 2018; Feng et al., 2015). Under pixel-based approaches, where each pixel is analyzed independently, this limitation is further amplified, resulting in reduced interpretability and fragmented classification outputs, particularly in heterogeneous vegetation environments (Persello and Bruzzone, 2016).

A second major challenge is the scarcity of reliable reference data, often accompanied by pronounced class imbalance. In many very high-resolution UAV datasets, only a small fraction of the available pixels is associated with definitive labels. Under such conditions, model training becomes highly constrained, and the reliability of performance evaluation is reduced. Previous investigations have demonstrated that achieving high overall accuracy does not necessarily translate into reliable full-scene predictions or spatially stable classification outcomes (Persello and Bruzzone, 2016; Zhu et al., 2018). This limitation is particularly critical in sensitive aquatic and riparian ecosystems, where even small spatial misclassifications may lead to substantial ecological misinterpretations (Matrone et al., 2025).

Accordingly, the central problem addressed in this study is not merely the selection of a specific classification algorithm, but rather the design of an effective feature representation capable of reducing spectral ambiguity and improving spatial coherence. The objective is to minimize the gap

between numerical performance metrics and the actual spatial quality and interpretability of the final vegetation map. This problem framing is consistent with recent data-centric perspectives in remote sensing, which emphasize feature-space enrichment over increasing model complexity under conditions of limited labeled data (Zhu et al., 2018; Fan et al., 2025).

1.3 Research objectives

The main objective of this research is to develop and evaluate a pixel-based machine learning framework for vegetation classification in multispectral UAV imagery under conditions of severe reference-data scarcity. The proposed framework follows a data-centric perspective, prioritizing feature-space enrichment while deliberately avoiding complex and computationally intensive models that are not well suited for data-limited scenarios (Persello and Bruzzone, 2016; Zhu et al., 2018).

The specific objectives of the study are as follows:

- To design and implement a baseline pixel-level vegetation classification approach based on original multispectral bands and derived vegetation indices, serving as a reference configuration for subsequent comparisons (Gini et al., 2018);
- To evaluate the performance of XGBoost as a lightweight, stable, and scalable machine learning model under conditions of limited labeled data and pronounced class imbalance (Fan et al., 2025);
- To investigate the contribution of neighborhood-based spatial texture information in improving class separability, particularly for vegetation types exhibiting strong spectral overlap in very high-resolution UAV imagery (Gini et al., 2018; Feng et al., 2015);
- To develop an enriched feature space by combining spectral features, vegetation indices, and local first-order statistical texture descriptors, and to systematically compare its performance against the baseline configuration (Fan et al., 2025);
- To quantitatively assess classification performance using appropriate accuracy metrics, with particular attention to spectrally confounded and under-represented vegetation classes (Persello and Bruzzone, 2016);
- To evaluate the robustness and generalization capability of the proposed framework through its application to an independent dataset with different spatial characteristics and levels of label availability;
- To assess the practical applicability of the proposed framework for real-world environmental monitoring and ecosystem management scenarios characterized by severe reference-data scarcity (Fan et al., 2025).

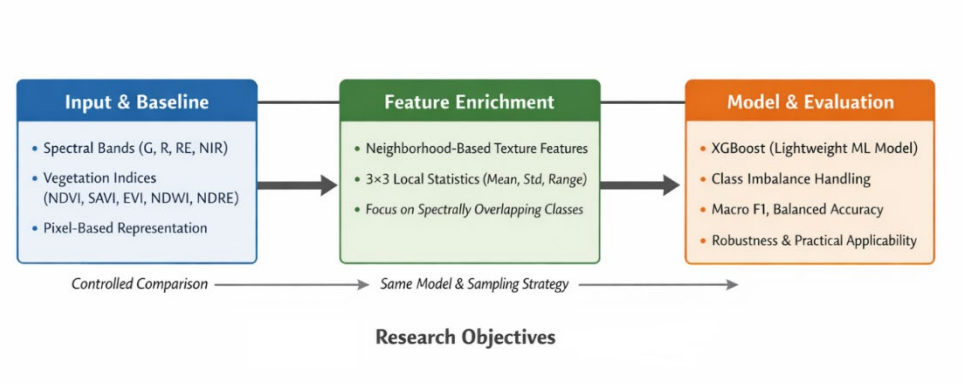


Figure 2 - Conceptual framework and research workflow

1.4 Novel contributions

The novelty of this research lies in the development and evaluation of an efficient pixel-based vegetation classification framework specifically tailored to conditions of extremely limited reference data. In contrast to model-centric approaches that rely on complex and data-hungry algorithms, this study adopts a data-centric perspective, emphasizing feature-space design and the use of a lightweight and interpretable learning model (Persello and Bruzzone, 2016; Zhu et al., 2018).

The main contributions of this research can be summarized as follows:

- Proposing a pixel-level vegetation classification framework based on XGBoost for multispectral UAV imagery, explicitly designed to operate under conditions of severe reference-data scarcity (Fan et al., 2025);
- Demonstrating that targeted feature-space enrichment can significantly improve classification performance without increasing model complexity or computational cost, supporting a data-centric approach to high-resolution vegetation mapping (Fan et al., 2025; Persello and Bruzzone, 2016);
- Introducing first-order statistical texture features computed over local 3 × 3 neighborhoods as an efficient and scalable alternative to more computationally demanding texture methods, thereby improving class separability in very high-resolution imagery (Gini et al., 2018);
- Providing a systematic and quantitative comparison between a baseline 9-band feature configuration and an enriched 27-band feature space, highlighting clear performance gains, particularly in terms of Macro F1 score for spectrally overlapping and under-represented vegetation classes (Fan et al., 2025);

- Validating the robustness and generalization capability of the proposed framework through its application to an independent dataset with different spatial characteristics and levels of label availability;
- Demonstrating that a lightweight, data-centric approach can achieve both high classification accuracy and improved spatial consistency, supporting its practical applicability in real-world environmental monitoring and ecosystem management scenarios characterized by severe data limitations.

1.5 Thesis structure

This thesis is organized into six main sections:

- The Introduction presents the research context and motivation, outlines the significance of the problem, and defines the main challenges, objectives, and contributions of the study.
- The Literature Review examines previous work on UAV-based vegetation mapping, spectral bands and vegetation indices, texture feature extraction, and machine learning approaches, and identifies the key research gaps addressed in this thesis.
- The Methodology and Data section provides a detailed description of the study area, the characteristics of the multispectral UAV imagery, the reference data, and the complete processing workflow, including feature engineering, experimental design, and the XGBoost-based classification framework.
- The Results section presents the experimental findings, including the comparison between the baseline spectral approach and the texture-enhanced feature space, as well as the evaluation of the proposed framework on an independent dataset.
- The Discussion interprets the results in relation to existing studies, with particular attention to robustness under class imbalance, spatial behavior of the predictions, and the implications of adopting a data-centric rather than a model-centric perspective.
- Finally, the Conclusions and Outlook summarize the main findings, highlight the scientific and practical contributions of the work, and outline the limitations and directions for future research.

2. Literature review

The increasing availability of very high-resolution UAV imagery has significantly enhanced the potential for detailed vegetation and species mapping in complex natural environments. Multispectral UAV data allow the observation of fine-scale spatial patterns and spectral variability that are not adequately captured by conventional satellite platforms. Recent studies have

demonstrated the effectiveness of UAV-based multispectral imagery for detailed habitat and vegetation mapping, particularly in ecologically heterogeneous environments where fine-scale spatial information is essential for accurate classification and monitoring (Belcore et al., 2021). However, despite these technological advances, vegetation classification under such conditions continues to face fundamental methodological challenges, most notably the limited availability of reliable reference data. Field-based data collection in heterogeneous and hard-to-access environments is inherently time-consuming, costly, and logistically demanding, often resulting in training datasets that represent only a small fraction of the total image content (Bazzo et al., 2024).

In UAV-based studies characterized by large image volumes and limited labeled data, achieving high overall accuracy does not necessarily guarantee spatially coherent or generalizable classification maps at the scale of the full scene. Previous research has shown that pixel-based classification approaches are particularly sensitive to spatial instability and salt-and-pepper noise, even when numerical performance metrics appear satisfactory, especially when spatial context is not explicitly considered during model design and validation (Ploton et al., 2020). These effects are further amplified in very high-resolution imagery, where local spectral variability and structural heterogeneity within vegetation classes lead to fragmented and noisy prediction patterns (Chen et al., 2022).

These limitations highlight a critical gap in the existing literature: the lack of classification frameworks capable of simultaneously improving statistical performance and spatial quality under conditions of large data volumes and limited reference data. Recent studies suggest that, in such scenarios, increasing model complexity alone is often insufficient, and that greater emphasis should instead be placed on feature representation, data quality, and appropriate validation strategies (Maxwell et al., 2018; Roberts et al., 2017). Similar observations have been reported in UAV-based vegetation mapping studies, where the integration of informative spectral features and carefully designed data processing workflows has been shown to substantially improve classification performance without relying on increasingly complex modelling approaches (Belcore et al., 2021). Addressing this gap requires the development of efficient and interpretable approaches based on data-centric feature enrichment, which forms the methodological foundation of the present study.

2.1 UAV-based vegetation mapping

In recent decades, imagery acquired by unmanned aerial vehicles (UAVs) has become one of the most important data sources in very high-resolution remote sensing. The combination of centimetric spatial resolution, flexible temporal acquisition, and the ability to operate in hard-to-access areas makes UAV platforms particularly suitable for vegetation mapping and monitoring at local scales, especially in heterogeneous and environmentally sensitive ecosystems (Bazzo et al., 2024).

Compared to satellite-based observations, UAV imagery enables the capture of fine-scale structural details of vegetation canopies, which are essential for species-level discrimination, accurate delineation of vegetation boundaries, and the analysis of spatial patterns. The integration of multispectral sensors further allows the acquisition of spectral information related to plant physiological conditions, supporting the computation of vegetation indices and the detection of subtle variations in vegetation status (Singhal et al., 2022). Recent research has demonstrated that UAV multispectral imagery can effectively support detailed habitat and vegetation mapping in complex ecological environments, enabling the discrimination of vegetation communities and habitat types at fine spatial scales that are difficult to achieve using conventional remote sensing platforms (Belcore et al., 2021).

Despite these advantages, UAV-based vegetation mapping presents several methodological challenges. Very high-resolution data are highly sensitive to illumination conditions, shadows, local moisture variations, and structural heterogeneity within vegetation covers. These factors increase spectral overlap among classes and complicate the classification process (Chen et al., 2022). Such limitations are particularly evident in pixel-based approaches, where each pixel is treated independently and local spatial context is not explicitly incorporated.

In addition, UAV imagery typically consists of extremely large data volumes, often reaching tens or hundreds of millions of pixels. Under these conditions, the limited availability of reliable reference data becomes a critical bottleneck, constraining both model training and the generalization capability of classification results. As a result, the selection of appropriate preprocessing strategies, feature extraction methods, and machine learning approaches becomes a key methodological consideration in UAV-based vegetation studies, particularly when operating under conditions of limited labeled data (Belgiu and Drăguţ, 2016).

2.2 Spectral bands and vegetation indices

Spectral information recorded in remote sensing imagery represents a fundamental component of vegetation analysis and classification. In multispectral UAV imagery, each spectral band captures a specific aspect of plant reflectance behavior, which is closely related to vegetation structure, biochemical properties, and physiological processes. In particular, the green, red, red-edge, and near-infrared bands play a central role in vegetation studies and form the basis for the computation of the most widely used vegetation indices (Huete et al., 2002).

The strong absorption of chlorophyll in the red band, combined with the high reflectance of healthy vegetation in the near-infrared region, enables an initial separation between vegetated and non-vegetated surfaces. The green band provides complementary information related to overall vegetation condition, while the red-edge band is especially sensitive to variations in chlorophyll

content and plant stress. This makes it particularly useful for discriminating vegetation types with similar spectral signatures (Gitelson et al., 1996).

Based on these spectral properties, a large number of vegetation indices, such as NDVI, EVI, NDRE, SAVI, and NDWI, have been developed. By combining spectral bands mathematically, these indices aim to reduce the effects of illumination variability, soil background, and geometric distortions, thereby improving vegetation discrimination (Huete, 1988). However, several studies have shown that relying solely on spectral information and vegetation indices, particularly in very high-resolution imagery, does not always lead to satisfactory classification results. This limitation highlights the need for complementary sources of information to improve class separability. Recent UAV-based studies have further demonstrated that the combination of multispectral bands with robust feature engineering strategies can significantly enhance vegetation classification performance, especially in heterogeneous environments where spectral similarity between classes is high (Belcore et al., 2021). (Pan et al., 2022)

2.3 Texture features in high-resolution remote sensing

With the continuous increase in the spatial resolution of remote sensing imagery, the limitations of approaches relying exclusively on spectral information have become increasingly evident. In very high-resolution images, each pixel represents a small portion of the Earth's surface, and local variations in vegetation density, structure, and spatial arrangement appear in a highly heterogeneous and fragmented manner. Under such conditions, pixel-wise analysis that ignores spatial relationships fails to capture a significant portion of the available information and often leads to increased class confusion and spatial instability in classification outputs (Ploton et al., 2020).

To address these limitations, texture features have been introduced to describe spatial patterns and local variations in pixel values beyond what can be captured by spectral information alone. In remote sensing, texture refers to the spatial distribution of reflectance values, surface uniformity, roughness, and variability, all of which are closely related to vegetation structure and growth patterns. These characteristics are particularly useful for distinguishing vegetation classes that exhibit similar spectral responses but differ in spatial organization or canopy structure (Jintasuttisak et al., 2022).

Texture extraction methods reported in the literature are generally categorized into statistical, structural, and spatial dependency-based approaches. Among these, statistical texture descriptors computed over local neighborhoods have received considerable attention in UAV-based studies due to their computational efficiency, limited parameter requirements, and scalability to very large datasets (Zou et al., 2021). Such features provide complementary information on vegetation

heterogeneity and spatial structure, and have been shown to improve class separability and spatial coherence in classification results without requiring more complex modeling strategies.

2.4 Texture-based vegetation classification

Numerous studies have demonstrated that the incorporation of texture features in high spatial resolution imagery can significantly enhance the discriminative power of vegetation classification models, particularly in heterogeneous environments where vegetation species exhibit strong spectral similarity. In many cases, texture features are used as a complement to spectral bands and vegetation indices, resulting in improved classification accuracy and reduced inter-class confusion (Gini et al., 2018; Zou et al., 2023).

These improvements are especially evident for vegetation classes that, despite similar spectral responses, differ in terms of spatial structure, canopy density, or growth patterns. For example, (Gini et al. 2018) showed that texture descriptors extracted from small local neighbourhoods in multispectral UAV imagery can substantially improve species separability and the spatial coherence of classification maps. Similar findings were reported by Zou et al. (2023), highlighting the effectiveness of texture-based feature enrichment without significantly increasing model complexity.

Moreover, recent studies conducted within the ACLIMO project emphasize the importance of integrating spatial and structural information when mapping vegetation in alpine and lake-shore environments, where reliance on spectral information alone often leads to pixel-level noise and spatial instability (Matrone et al., 2025). Overall, the literature indicates that the selection of texture extraction methods should be aligned with image scale, spatial resolution, and classification objectives. In very high-resolution UAV imagery, texture descriptors derived from small local neighbourhoods are particularly effective in capturing fine-scale spatial variation while enriching the feature space without imposing significant computational complexity.

To visually support the concepts discussed in this section, Figure 3 from (Gini et al. 2018) and is included solely for illustrative purposes. The figure highlights the effect of texture feature integration on the spatial coherence of vegetation classification results and does not represent outcomes of the present study.

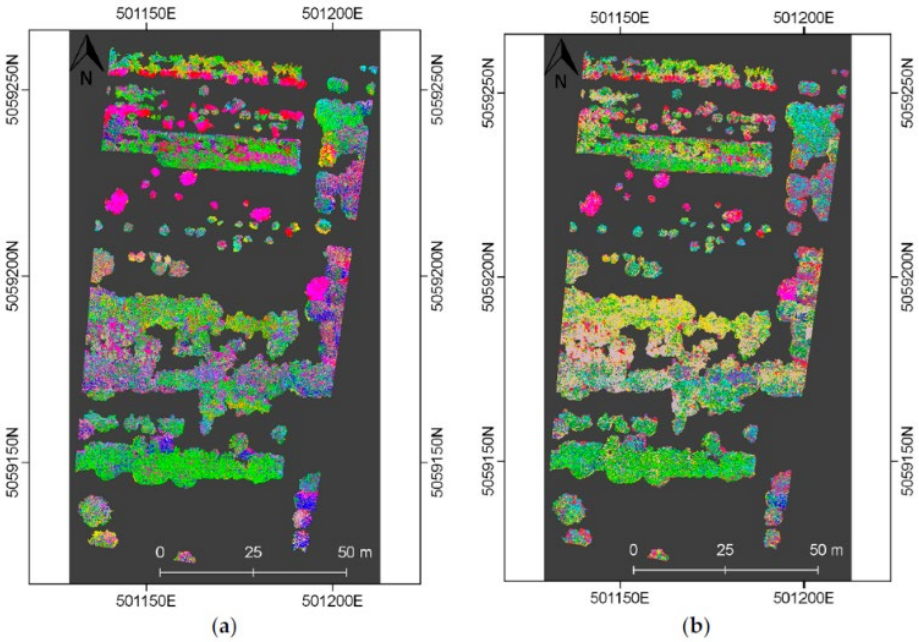


Figure 3 - conceptual illustration of the effect of texture feature integration on vegetation classification (Gini et al. 2018)

2.4.1 Previous work on the same study areas

Previous research conducted in the same alpine lake environments has mainly been carried out within the ACLIMO project, which focuses on monitoring climate change effects on mountain lake ecosystems using UAV-based remote sensing and artificial intelligence techniques (Matrone et al., 2025). These studies provide a comprehensive framework for mapping vegetation patterns in high-altitude aquatic and riparian zones using very high-resolution multispectral imagery.

Within this framework, vegetation classification is typically based on object-based image analysis (OBIA), where image segmentation is used to delineate homogeneous regions, followed by classification using spectral, spatial, and contextual features. This approach allows spatial structure to be incorporated into the analysis, reducing pixel-level noise and improving the consistency of mapped vegetation units.

Multispectral UAV data are combined with vegetation indices and spatial information to characterize different vegetation types and ecological zones. These studies emphasize the importance of integrating structural and contextual information when working with very high-resolution imagery, particularly in heterogeneous alpine environments where spectral similarity among vegetation classes is common (Matrone et al., 2025).

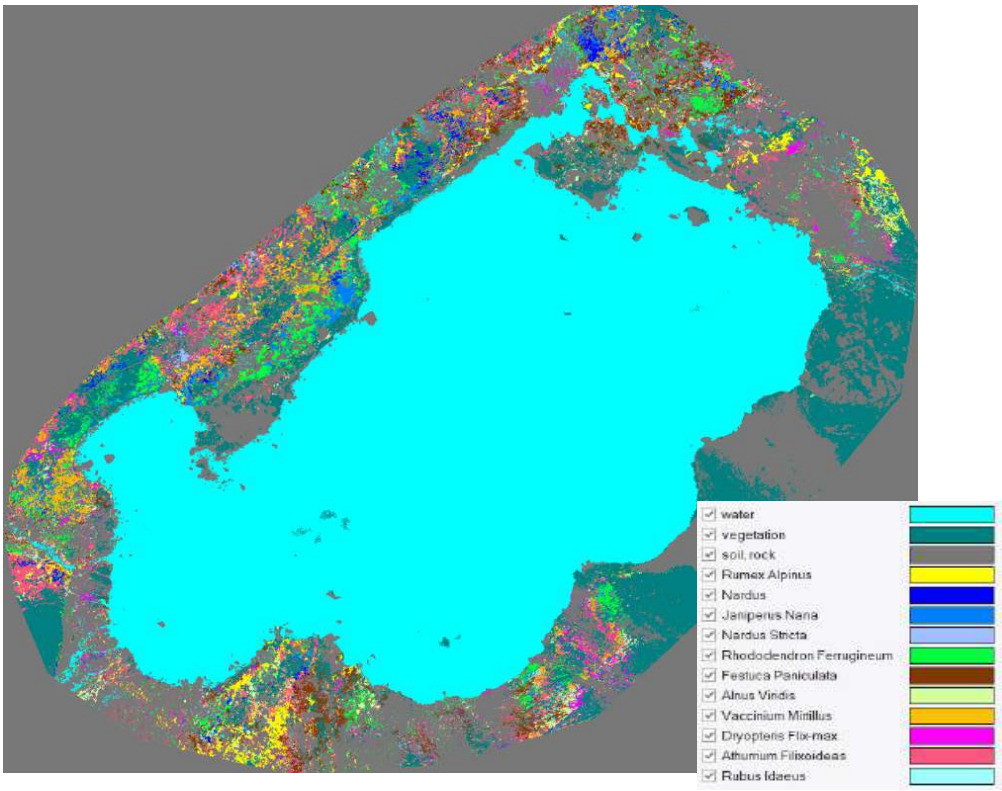


Figure 4 - Example of vegetation classification in the study area ACLIMO project. (Matrone et al., 2025)

2.5 Machine learning for vegetation classification

With the increasing volume and complexity of remote sensing data, machine learning has become one of the primary approaches for vegetation classification. By modeling nonlinear relationships between input features and class labels, these methods enable more effective use of spectral, spatial, and texture information, offering greater flexibility than traditional rule-based techniques (Mountrakis et al., 2011; Rodriguez-Galiano et al., 2012).

The literature reports extensive use of algorithms such as Support Vector Machines, Random Forests, k-Nearest Neighbors, and decision tree-based methods, many of which show strong performance under conditions of limited but high-quality training data. In contrast, deep learning models, despite their high representational capacity, typically require large and balanced labeled datasets, as well as substantial computational resources. These requirements often limit their applicability in UAV-based environmental monitoring and data-scarce scenarios (Safonova et al., 2023).

In this context, several studies suggest that the design and richness of the feature set can be more influential than increasing model complexity alone. Lightweight tree-based and boosting-oriented models are therefore considered suitable for practical applications due to their robustness to class

imbalance, interpretability, and relatively low data requirements. Such approaches enable effective use of enriched feature spaces while maintaining stable model behavior and computational efficiency (He and Garcia, 2009; Samat et al., 2020; Łoś et al., 2021).

2.6 Research gap

The reviewed literature indicates that, despite significant methodological advances, important gaps remain in the practical application of vegetation classification at local scales and in sensitive natural ecosystems. Many existing studies focus either on satellite imagery with moderate spatial resolution or on experimental settings with relatively large and balanced training datasets. Such conditions do not reflect the operational constraints of UAV-based applications, where image volumes are extremely large and labeled data are often scarce and highly imbalanced (Safonova et al., 2023; Li et al., 2025).

In UAV-based scenarios characterized by limited reference data, it has been repeatedly observed that satisfactory numerical performance does not necessarily translate into classification maps with adequate spatial coherence and semantic reliability at the full-scene scale. Pixel-based approaches, in particular, tend to exhibit spatial instability and structural inconsistencies when applied to very high-resolution imagery, even when standard accuracy metrics suggest acceptable performance.

Consequently, a key research gap lies in the lack of classification frameworks capable of simultaneously improving statistical accuracy and spatial quality under conditions of large data volumes and limited reference information. Addressing this gap requires approaches that emphasize lightweight and controllable models combined with targeted feature-space enrichment, rather than increasing model complexity alone. The present research is designed to address this methodological challenge.

2.7 Data-centric approaches in remote sensing

In recent years, increasing attention has been directed toward data-centric approaches in remote sensing, particularly in applications with limited labeled data and high data complexity. In contrast to model-centric strategies, which focus on improving performance through more sophisticated algorithms, data-centric approaches emphasize the quality and representation of input data (Zhu et al., 2018).

Within this perspective, feature-space design plays a central role. Instead of relying on algorithmic complexity, data-centric methods aim to improve classification performance by enhancing the descriptive power of input features, for example through the integration of spectral, spatial, and texture information, as well as appropriate preprocessing of training data.

This approach is particularly relevant in high-resolution UAV-based remote sensing, where datasets are large and labeled data are often limited and imbalanced. Under such conditions, increasing model complexity alone does not necessarily improve generalization and may lead to unstable predictions. Several studies suggest that improving data representation can be more effective, especially in scenarios with limited ground-truth information. In this context, UAV-based vegetation mapping studies have demonstrated that carefully designed feature engineering strategies, particularly those combining multispectral information with derived vegetation indices, can substantially improve classification robustness without increasing model complexity (Belcore et al., 2021). (Fan et al., 2025)

Data-centric strategies therefore provide a practical framework for addressing key challenges in vegetation classification, including spectral ambiguity, spatial heterogeneity, and limited reference data, by focusing on feature-space enrichment and data quality.

3. Methodology and data

3.1 Overview of the proposed methodology

This research focuses on the classification of vegetation cover and species in a mountainous alpine environment using very high-resolution UAV imagery. The proposed framework is developed under two main constraints: the limited availability of labeled reference data and the inherent complexity of high-resolution imagery, characterized by strong intra-class heterogeneity, complex spatial boundaries, and significant spectral overlap among vegetation species (Persello and Bruzzone, 2016).

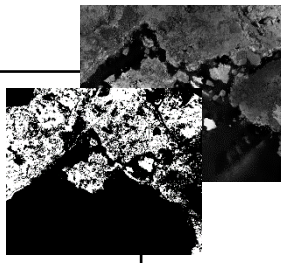
Phase 1

Data Acquisition

UAV multispectral (4 bands)

Green, Red, Red Edge, Nir





Phase 2

Preprocessing & Masking

Orthophoto & point cloud (3 cm)

Vegetation indices

non-vegetation removal

Phase 3

Feature Engineering

Texture (3×3) → 27 bands

Spectral + indices (9 bands)

Indices: NDVI, SAVI, EVI, NDRE, NWDI

Phase 4

Model Training

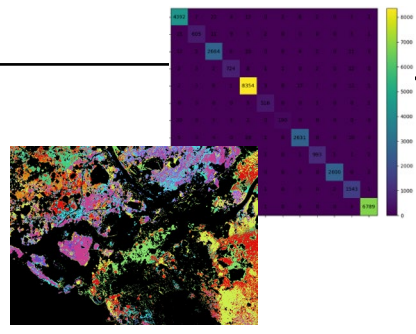
Supervised learning (XGBoost)

Baseline (9-band) vs. texture-enhanced (27-band)

Phase 5

Evaluation & Output

Model evaluation Classification maps



Such data conditions reflect typical real-world environmental monitoring scenarios, where image volumes are large while labeled data remain scarce and unevenly distributed. Under these circumstances, increasing model complexity alone is often ineffective, highlighting the importance of designing informative and well-structured feature representations (Safonova et al., 2023).

Accordingly, the proposed framework follows a data-centric paradigm, placing emphasis on improving the quality and descriptive power of the input data rather than relying on more complex models. A supervised, pixel-based classification strategy is adopted, in which spectral information is progressively complemented with additional features to enhance the representation of vegetation characteristics.

To evaluate the impact of feature-space design, a set of aligned experimental configurations is defined, allowing a controlled comparison between a baseline representation and an enriched feature space. This design enables a clear assessment of how the integration of additional information influences classification performance.

Across all configurations, the classifier, sampling strategy, class weighting scheme, and evaluation metrics are kept consistent to ensure that observed differences are attributable to variations in the feature space. The overall workflow includes preprocessing and masking, feature engineering, sample extraction and splitting, model training within a gradient-boosting framework with class weighting, and evaluation using metrics suitable for imbalanced datasets (He and Garcia, 2009).

3.2 Data acquisition

This study is based on two multispectral UAV datasets acquired over alpine lake ecosystems in the southwestern Italian Alps. The primary dataset, collected over Lago Brocan, was used for the development and evaluation of the proposed vegetation classification framework, while an independent dataset from Lake Vej del Bouc was employed to assess its robustness and transferability. Both datasets consist of very high spatial resolution UAV imagery acquired using a DJI Mavic 3M platform and are characterized by complex vegetation patterns, strong spatial heterogeneity, and severe scarcity of labeled reference data. To support subsequent classification analyses, georeferenced multispectral imagery, manually delineated ground-truth polygons, and derived vegetation features were prepared following a consistent data acquisition and preprocessing workflow.

3.2.1 Study area description

The study area encompasses Lago Brocan and its surrounding landscape within a mountainous alpine environment in Italy (southwestern Alps). Owing to its specific climatic conditions,

seasonal variability, and direct dependence on water resources, this region represents a sensitive ecosystem for environmental monitoring. Alpine lakes are widely recognized as effective natural indicators for assessing climatic and environmental changes at local and regional scales (Matrone et al., 2025). The second study area is located around Lake Vej del Bouc in the southwestern Italian Alps and was used as an independent validation site for evaluating the robustness and transferability of the proposed framework.

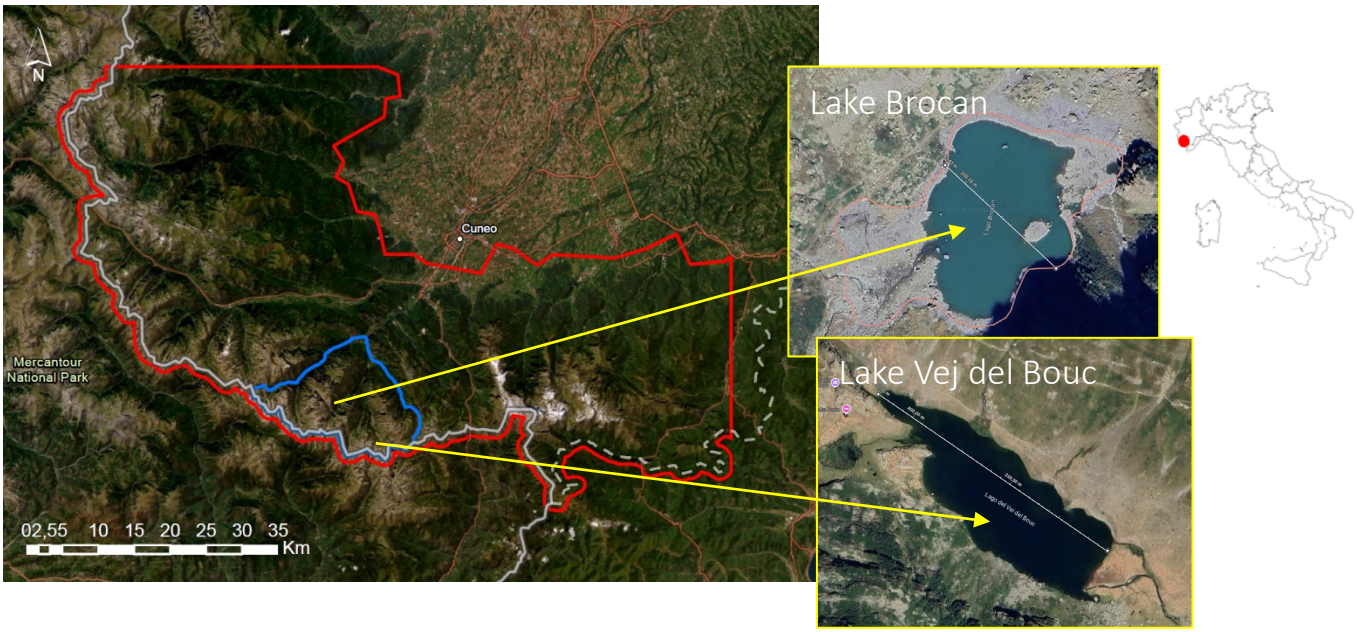


Figure 5 - Study area at multiple scales, showing the region of interest (red), intermediate scale (blue), and the lake area (yellow), adapted from (Matrone et al. 2025).

3.2.2 Environmental and ecological context

Alpine aquatic ecosystems are simultaneously influenced by climatic, hydrological, and biological factors. Variations in temperature, precipitation patterns, snowmelt dynamics, and water-level fluctuations directly affect the structure and spatial distribution of vegetation. From a spatial perspective, vegetation in these regions is typically patchy and highly heterogeneous; species alternate over short distances in response to localized changes in moisture, substrate type, and solar exposure. This results in complex and gradual class boundaries, which pose significant challenges for pixel-based classification approaches (Persello and Bruzzone, 2016).

From a temporal perspective, alpine vegetation exhibits dynamic behavior due to its phenology and short growing season. In this study, UAV imagery was acquired in 2024, and the results therefore correspond to the phenological conditions of that specific acquisition period. Collectively, these characteristics indicate that vegetation classification in this area is not purely a

spectral problem but is strongly dependent on spatial structural information, which underpins the feature space design adopted in this research (Zou et al., 2021).

3.2.3 UAV image acquisition

The data were acquired through UAV-based aerial imaging as part of environmental monitoring activities in a mountainous alpine region. The primary objective of UAV deployment was to obtain very high spatial resolution imagery and capture detailed vegetation patterns at a local scale. Image acquisition was conducted in 2024 during an appropriate phase of the growing season, in order to maximize both spectral and structural separability among vegetation types (Matrone et al., 2025).

The UAV platform used in this study was a DJI Mavic 3M equipped with a GNSS-RTK positioning system. High positional accuracy is essential for pixel-based analysis and spatial feature extraction, as geometric inaccuracies can directly affect feature quality and classification results, particularly in large-scale UAV datasets (Roberts et al., 2017).

The multispectral sensor captures four spectral bands—Green, Red, Red Edge, and Near-Infrared (NIR)—which form the basis for vegetation index computation and subsequent feature extraction.

3.2.4 Multispectral data characteristics

The primary dataset consists of a georeferenced UAV orthomosaic with very high spatial resolution. The four spectral bands Green, Red, Red Edge, and Near-Infrared (NIR) serve as the base input features for subsequent analysis. The spatial resolution of the imagery is 3 cm (0.03 m), and the raster dimensions are 23,409 × 24,174 pixels.

The orthomosaic is stored in GeoTIFF format with LZW compression, and all spectral bands are co-registered within a common coordinate reference system (EPSG:32632), ensuring spatial consistency across the dataset. The NoData value is set to -9999, and the number of valid pixels after masking is 131,865,369.

All data are projected in EPSG:32632 (WGS84 / UTM Zone 32N), providing a consistent spatial reference for further processing and analysis.

In addition to the four multispectral bands, a blue channel derived from the RGB orthomosaic of the same study area was used exclusively for the computation of the Enhanced Vegetation Index (EVI). The RGB data were generated at the same spatial resolution and are spatially aligned with the multispectral imagery, allowing consistent integration of the blue band within the feature extraction process.

To investigate the impact of feature-space design, two feature configurations were defined for classification. The first configuration is based on spectral bands combined with vegetation indices, while the second extends this representation by incorporating additional spatial texture features derived from local neighbourhoods (Gini et al., 2018).

Attribute	Description
Platform	DJI Mavic 3M (GNSS-RTK enabled)
Acquisition year	2024
Study area	Lago Brocan, Alpine environment, Italy
Sensor type	Multispectral
Spectral bands	Green, Red, Red Edge, Near-Infrared (NIR)
Additional band (derived)	Blue channel from RGB orthomosaic (used for EVI computation)
Spatial resolution	0.03 m (3 cm)
Raster dimensions	23,409 × 24,174 pixels
Total number of pixels (raw image)	≈ 566,000,000 pixels
Coordinate reference system	EPSG:32632 (WGS84 / UTM Zone 32N)
Data format	GeoTIFF (LZW compression)
NoData value	-9999
Valid pixels after vegetation masking	131,865,369 pixels
Vegetation mask criterion	NDVI > 0.30

Table 1 - Summary of the UAV multispectral dataset, including sensor, spatial resolution, spectral bands, and vegetation masking criteria. (Brocan Lake)

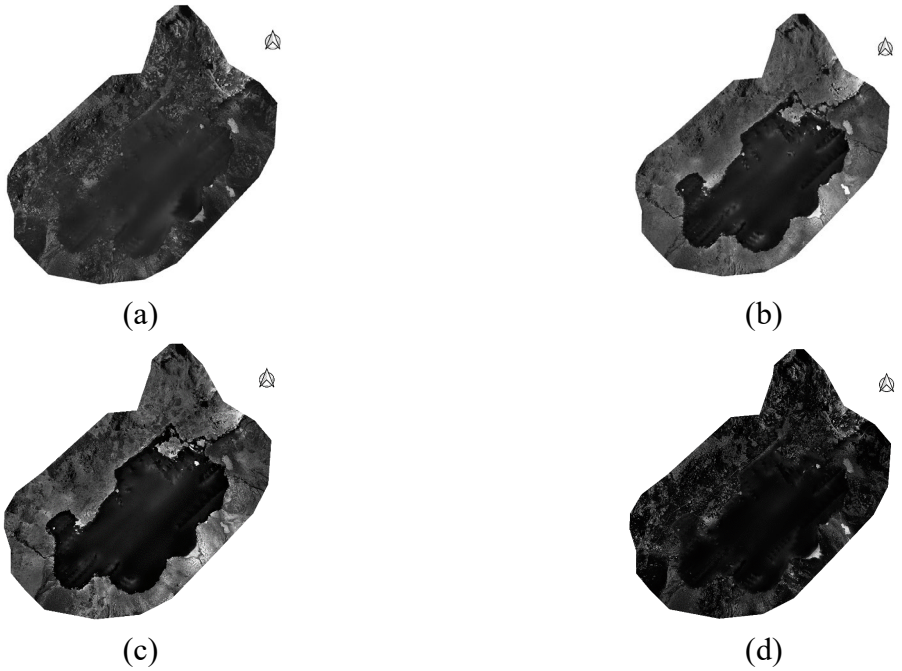


Figure 6 - Individual multispectral bands of the study area: (a) Green, (b) Red, (c) Red Edge, and (d) Near-Infrared (NIR).

3.2.5 Reference data and ground truth

The reference data consist of manually delineated polygons used for model training and evaluation, defined within the same coordinate reference system as the raster data (EPSG:32632). Due to practical constraints and the ecological sensitivity of the alpine environment, the volume of reference data is extremely limited. A total of 45 polygons across 11 vegetation classes were defined, yielding 162,156 labeled pixels, corresponding to approximately 0.12% of the valid raster pixels.

At the scale of the full image, this represents only about 0.03% of all pixels, placing the problem in a severely label-scarce regime. In addition, class imbalance is pronounced, with some classes represented by substantially more samples than others.

The ground truth data were generated through a combined field-based and image-based procedure. During field surveys, vegetation species of interest were identified and marked by domain experts, and representative points were recorded for each instance. These points were positioned as close as possible to the central and most homogeneous part of each vegetation patch in order to minimize boundary-related uncertainty. The geographic coordinates of these points were then imported into a GIS environment and used as reference locations.

Based on these reference points and the high-resolution UAV orthomosaic, regions of interest (ROIs) were manually delineated to define the training polygons. This approach ensures that each polygon corresponds to a homogeneous vegetation area associated with expert-identified field samples.

To mitigate these limitations, polygons were selected to capture intra-class variability as effectively as possible. Only interior pixels were used for training and evaluation in order to reduce label uncertainty near class boundaries. The spatial accuracy of the ground truth is supported by the high positional precision of the UAV data and the use of field-based point collection. While some degree of uncertainty is inherent in manual delineation, the adopted strategy provides a consistent and reliable reference dataset under conditions of limited labeled data.

Attribute	Value
Reference data type	Polygon-based ground truth
Number of vegetation classes	11
Number of reference polygons	45
Total labeled pixels	162,156 pixels
Percentage of labeled pixels (of valid pixels)	≈ 0.123%
Percentage of labeled pixels (of full image)	≈ 0.03%
Total valid pixels	131,865,369
Labeling strategy	Manual expert interpretation
Class balance	Highly imbalanced
Smallest class	Red Raspberry
Largest class	Dwarf Juniper

Table 2 - Summary of the reference (ground truth) data, including the number of polygons, labeled pixels, class distribution, and data split strategy under conditions of severe label scarcity.



Figure 7 - Representative samples of vegetation classes in the Brocan dataset: Green Alder (a), Tussock Fescue (b), Matgrass (c), Lady Fern (d), Blueberry (e), Male Fern (f), Upright Matgrass (g), Rusty-Leaf Rhododendron (h), Dwarf Juniper (i), Red Raspberry (j), and Rumex Alpinus *(k).*

ID	Class	Labeled Pixels	% of Labeled Pixels	% of Total Raster Pixels
1	Alpine Dock	22,292	13.75%	0.0169%
2	Lady Fern	3,253	2.01%	0.0025%
3	Green Alder	13,601	8.39%	0.0103%
4	Male Fern	3,774	2.33%	0.0029%
5	Dwarf Juniper	42,038	25.92%	0.0319%
6	Matgrass	2,625	1.62%	0.0020%
7	Red Raspberry	1,086	0.67%	0.0008%
8	Rusty-Leaf Rhododendron	13,393	8.26%	0.0102%
9	Tussock Fescue	5,109	3.15%	0.0039%
10	Upright Matgrass	13,014	8.03%	0.0099%
11	Blueberry	7,950	4.90%	0.0060%
-	Unknown	34,021	20.98%	0.0258%
-	Total	162,156	100%	0.123%*

Table 3 - Distribution of labeled pixels across vegetation classes in the Brocan dataset, highlighting class imbalance and the proportion of each class relative to both the labeled data and the full raster.

3.2.6 Independent validation Dataset (Lake Vej del Bouc)

To assess the robustness and generalization capability of the proposed framework, an independent dataset from Lake Vej del Bouc is considered as a second case study. This dataset represents a larger and more complex alpine environment, providing a suitable test scenario for evaluating the transferability of the methodology under different spatial and data conditions.

The dataset consists of very high-resolution multispectral UAV imagery acquired over a mountainous lake environment similar in ecological context to Lago Brocan, but characterized by a substantially larger spatial extent and a different distribution of vegetation classes. As in the primary dataset, vegetation areas are affected by strong spectral overlap, high spatial heterogeneity, and complex ecological gradients.

Despite the increased spatial coverage, the availability of labeled reference data remains limited, resulting in a label-scarce setting comparable to the primary study area. This combination of large data volume and limited ground truth makes the dataset particularly suitable for testing the performance of data-centric classification approaches.

The same data preparation principles are applied, including masking of non-vegetated areas and extraction of multispectral and derived features. The dataset is used exclusively for independent validation and is not involved in the design or tuning of the proposed framework.

Attribute	Description
Study area	Lake Vej del Bouc, Alpine environment, Italy
Data type	Multispectral UAV imagery
Spatial resolution	Very high resolution (≈ 3 cm)
Total number of pixels	≈ 849,598,608 pixels
Valid pixels after masking	≈ 237,882,024 pixels
Number of vegetation classes	12
Number of reference polygons	46
Total labeled pixels	378,748 pixels
Percentage of labeled pixels (of valid pixels)	≈ 0.16%
Percentage of labeled pixels (of full image)	≈ 0.045%
Labeling strategy	Manual expert interpretation
Class balance	Imbalanced

Table 4 - Summary of the independent validation dataset from Lake Vej del Bouc, including spatial properties, labeled data availability, and class characteristics under a label-scarce setting.

ID	Class	Pixels	% of Labeled Data	% of Valid Raster
1	Adenostyles	5,463	1.44%	0.0023%
2	Alpine Dock	4,363	1.15%	0.0018%
3	Black Sedge	3,703	0.98%	0.0016%
4	Blueberry	5,888	1.55%	0.0025%
5	Cotton Grass	48,730	12.87%	0.0205%
6	Dwarf Juniper	44,224	11.68%	0.0186%
7	Goosefoot	8,335	2.20%	0.0035%
8	Green Alder	50,331	13.29%	0.0212%
9	Male Fern	16,360	4.32%	0.0069%
10	Matgrass	63,307	16.71%	0.0266%
11	Paniculate grass	76,749	20.26%	0.0323%
12	Rhododendron	20,449	5.40%	0.0086%
-	Unknown	30,846	8.14%	0.0130%

Table 5 - Distribution of labeled pixels across vegetation classes in the Lake Vej del Bouc dataset, highlighting class imbalance and the proportion of each class relative to the labeled data and valid raster.

3.3 Data processing

The data processing workflow was designed to support vegetation classification under conditions of very high spatial resolution and limited labeled data. The workflow includes image preprocessing and vegetation masking, feature engineering based on spectral, vegetation index, and texture information, training sample preparation, and supervised classification using a class-weighted XGBoost model. To evaluate the contribution of feature-space enrichment, a controlled experimental design was adopted in which different feature configurations were compared while maintaining consistent sampling, model settings, and evaluation procedures. The overall methodology follows a data-centric approach, emphasizing improvements in feature representation and data quality to enhance classification performance and generalization capability.

3.3.1 Image preprocessing and mask generation

To reduce data complexity and exclude irrelevant areas such as open water, bare rock or soil, and non-vegetated surfaces, a vegetation mask was generated based on the Normalized Difference Vegetation Index (NDVI). A conservative threshold of NDVI > 0.30 was selected to identify actively vegetated areas (Walz and Weber, 2009).

The objective of this thresholding was not to perform phenological analysis, but to spatially constrain the analysis domain prior to feature extraction and model training. By focusing only on vegetated regions, the approach reduces the influence of non-informative pixels and limits noise introduced by heterogeneous background elements.

The resulting mask was applied consistently across all spectral bands and derived indices. This step helps reduce boundary-related noise and improves the stability of pixel-based classification, particularly in areas with gradual transitions between vegetation and non-vegetation classes.



(a) (b)

Figure 8 - NDVI-based vegetation masking in Brocan Lake: (a) binary mask derived using NDVI > 0.30, and (b) corresponding multispectral representation, illustrating the exclusion of non-vegetated areas.

3.3.2 Feature engineering

Feature engineering was performed to enhance the descriptive power of the input data by integrating complementary spectral and spatial information. Two feature representations were constructed: a baseline feature set consisting of multispectral bands and vegetation indices, and an extended feature set that additionally incorporates texture descriptors derived from local neighborhoods. This design enables the evaluation of the contribution of spatial context to vegetation classification under limited labeled data conditions.

3.3.2.1 Spectral and index features

The primary spectral features consist of four bands Green, Red, Red Edge, and Near-Infrared (NIR)extracted from multispectral UAV imagery. These bands provide fundamental information on vegetation reflectance behavior and form the basis for computing vegetation indices and analyzing plant physiological conditions (Gitelson et al., 1996).

To enhance spectral discriminability and reduce overlap among vegetation classes, particularly in very high-resolution UAV imagery, five vegetation indices were computed on a pixel-wise basis and incorporated into the feature space (Pan et al., 2022). To ensure numerical stability, a small constant ε was added to the denominator of each index formulation, and all values were constrained within the standard range of [-1, 1].

The Normalized Difference Vegetation Index (NDVI) was used as a core indicator of vegetation activity and as the basis for generating the vegetation mask (Walz and Weber, 2009):

$$NDVI = \frac{NIR - Red}{NIR + Red + \varepsilon}$$

The Soil Adjusted Vegetation Index (SAVI) was applied to reduce soil background effects, particularly in sparsely vegetated areas and transitional zones. In this study, the soil adjustment parameter was set to L = 0.5 (Huete, 1988):

$$SAVI = (L + 1)\left(\frac{NIR - Red}{NIR + Red + L + \varepsilon}\right)$$

The Enhanced Vegetation Index (EVI) was used to improve robustness against illumination variations, reduce saturation effects in dense vegetation, and increase sensitivity to canopy structure (Huete et al., 2002). For the computation of EVI, in addition to the multispectral bands, a blue channel was required. In this study, the blue band was derived from the RGB orthomosaic of the same study area, which is spatially aligned with the multispectral imagery and shares the same spatial resolution.

$$EVI = 2.5 \times \frac{NIR - Red}{NIR + 6\,Red - 7.5\,Blue + 1 + \varepsilon}$$

The Normalized Difference Water Index (NDWI) was included to improve the discrimination of moist areas and reduce classification errors near vegetation–water boundaries (Pan et al., 2022):

$$NDWI = \frac{Green - NIR}{Green + NIR + \varepsilon}$$

Finally, the Normalized Difference Red Edge Index (NDRE), derived from the Red Edge band, provides increased sensitivity to chlorophyll content and plays an important role in distinguishing spectrally similar vegetation classes (Gini et al., 2018):

$$NDRE = \frac{NIR - RedEdge}{NIR + RedEdge + \varepsilon}$$

Together, the four spectral bands and the five vegetation indices form a 9-band feature representation. Compared to using spectral bands alone, this feature set provides richer and more stable information on vegetation condition, surface moisture, and subtle spectral differences. This is particularly important under conditions of limited reference data, where only a small fraction of pixels are reliably labeled (Persello and Bruzzone, 2016; Fan et al., 2025).

3.3.2.2 Texture feature extraction

To overcome the limitations of single-pixel information and reduce ambiguity among classes with similar spectral responses**, first-order statistical texture features** derived from local neighborhoods were incorporated. This approach is widely used in very high-resolution UAV-

based vegetation studies, where spatial context plays a critical role in improving class separability (Gini et al., 2018; Feng et al., 2015).

Instead of using computationally intensive co-occurrence matrix-based methods such as **GLCM**, three simple yet effective statistical descriptors were extracted from a 3×3 window centered on each pixel: Mean, Standard Deviation (Std), and Range (Max − Min). These features were selected due to their computational efficiency and robustness when applied to large-scale UAV datasets (Feng et al., 2015).

The Mean feature represents the local reflectance trend and contributes to smoothing small-scale variations. The Std feature captures local heterogeneity and surface roughness, while the Range feature describes local contrast and abrupt changes, making it particularly useful for identifying transitions between classes.

Texture features were computed for six selected layers: **Green, Red, Red Edge, NIR, NDVI, and NDRE**. This selection combines spectral information with vegetation indices that are sensitive to plant physiological conditions such as chlorophyll content and vegetation vigor (Gini et al., 2018; Zou et al., 2021).

In heterogeneous alpine and aquatic environments, fine-scale variations caused by moisture conditions, plant structure, and interactions with surrounding surfaces limit the effectiveness of purely spectral features. Texture descriptors enable the representation of spatial patterns such as patch homogeneity, surface roughness, and boundary structure, which are not captured at the pixel level.

To maintain a balance between feature richness and computational efficiency, texture extraction was restricted to these selected layers, avoiding unnecessary feature expansion and redundancy under limited labeled data conditions (Persello and Bruzzone, 2016).

The integration of these features results in a structured expansion of the feature space, where 4 spectral bands and 5 vegetation indices (9-band representation) are complemented by 18 texture features, leading to a total of 27 input features. This design enables a controlled evaluation of the contribution of spatial information to vegetation classification without modifying the underlying model or learning strategy.

3.3.3 Training sample selection and data split strategy

The selection of training samples and the data splitting strategy play a critical role in the evaluation of pixel-based classification models, particularly in very high-resolution remote sensing imagery (Foody, 2002; Persello and Bruzzone, 2016). In UAV data, strong spatial autocorrelation among neighboring pixels may lead to spatial leakage, where highly similar pixels appear in both training

and testing sets, resulting in overly optimistic accuracy estimates (Roberts et al., 2017; Ploton et al., 2020). Therefore, a transparent and well-justified sampling strategy is essential.

Samples were extracted from reference polygons defined through field observations and visual interpretation of UAV imagery. In line with the ground truth generation process, only pixels located in the central and homogeneous areas of the annotated vegetation patches were considered, in order to minimize boundary-related uncertainty and ensure high label reliability. Due to the limited availability of reference data, some classes were represented by only a small number of polygons, and in some cases by a single polygon, reflecting common constraints in UAV-based vegetation mapping (Persello and Bruzzone, 2016; Safonova et al., 2023). To better capture intra-class variability and reduce bias caused by local conditions such as illumination, moisture, and phenological stage, all available polygons for each class were included, following recommended practices for small and imbalanced datasets (He and Garcia, 2009).

Data splitting was performed at the pixel level using a class-wise strategy. For each class, approximately 80% of the pixels were assigned to the training set and 20% to the test set. The validation subset (10%) was subsequently derived from the training data only, ensuring that the test set remained completely independent during model development. All random operations were conducted using a fixed seed (seed = 42) to ensure reproducibility, in line with standard practices in remote sensing machine learning (Maxwell et al., 2018).

Training and test samples were generated as disjoint subsets of pixel coordinates using class-wise random permutation. Therefore, no identical pixel location was shared between the training and test sets, ensuring strict separation of samples at the pixel level.

Given the pixel-based nature of the data and the limited number of reference polygons, splitting was performed at the pixel level rather than at the polygon or spatial-block level. This choice was made with full awareness of the potential for spatial leakage. However, in this context, leakage does not arise from duplicated samples, but rather from spatial autocorrelation between neighboring pixels. Since the primary objective of this study is within-scene mapping rather than cross-scene generalization, this approach represents a practical and defensible compromise under limited data conditions, as also reported in similar studies (Persello and Bruzzone, 2016; Ploton et al., 2020).

Finally, the same sampling strategy, data splitting procedure, class weighting approach, and evaluation metrics were applied consistently across all experimental configurations. This ensures a fair comparison of feature-space representations and isolates the effect of feature enrichment on classification performance, while minimizing confounding factors related to data handling and evaluation (Foody, 2002; Maxwell et al., 2018).

3.3.4 XGBoost model and class weighting

The classifier used in this study is XGBoost, configured for multi-class classification and trained using GPU acceleration (CUDA). This model is widely adopted in remote sensing applications due to its scalability, robustness, and computational efficiency when handling large and high-dimensional datasets (Chen et al., 2022; Łoś et al., 2021). Its suitability is particularly evident in UAV-based vegetation mapping, where very high-resolution imagery results in large sample sizes and complex feature spaces (Maxwell et al., 2018).

To address the pronounced class imbalance in the reference data, an inverse class weighting strategy was applied during training. This approach assigns higher importance to under-represented classes and reduces the influence of dominant classes, and is commonly used in imbalanced classification problems (He and Garcia, 2009; Wang et al., 2022). Class weights were computed exclusively from the training data to avoid information leakage and ensure a fair learning process (King and Zeng, 2001). The weighting scheme is defined as:

$$w_c = \left(max_{k \in 1,\ldots,K} n_k\right)/n_c$$

Model training was monitored using a validation set, and learning behavior was assessed across boosting iterations to ensure stable convergence and to reduce the risk of overfitting (Liu et al., 2021; Safonova et al., 2023). The analysis of training and validation trends provides insight into model optimization and supports appropriate control of training duration and model complexity.

In practice, under conditions of limited labeled data, performance improvements were found to be more strongly associated with enhancements in feature representation than with extensive hyperparameter tuning. This observation is consistent with recent studies in high-resolution UAV-based classification and supports the data-centric design adopted in this research (Zhang et al., 2022; Ma et al., 2023).

3.3.5 Experimental design

To systematically evaluate the effect of feature-space design, three experimental scenarios were defined:

- Baseline scenario: classification based on a 9-band feature stack composed of spectral bands and vegetation indices.
- Texture-enhanced scenario: classification based on an expanded 27-band feature stack, obtained by incorporating spatial texture features into the baseline representation.

- Transferability scenario: evaluation of the proposed framework on an independent dataset to assess its generalization capability under different spatial and data conditions.

Across all scenarios, the classifier, sampling strategy, class weighting scheme, and evaluation metrics were kept consistent. This controlled experimental design ensures that any observed differences in performance can be attributed to variations in the feature space, rather than to changes in model configuration or data handling procedures.

3.3.6 Accuracy assessment metrics

Given the presence of class imbalance and limited reference data, a set of complementary evaluation metrics was used to ensure a reliable and balanced assessment of model performance (Persello and Bruzzone, 2016; Foody, 2002). The Macro F1-score was selected as the primary evaluation metric, as it assigns equal importance to all classes and provides a meaningful measure of performance for minority classes (Fan et al., 2025; Foody, 2023).

Overall Accuracy (OA) was also reported for descriptive purposes; however, it is not considered a reliable indicator under imbalanced conditions, as it can be dominated by majority classes (He and Garcia, 2009). To further characterize model performance, the Weighted **F1-score** was used to reflect the actual class distribution, while Balanced Accuracy was included to ensure equal sensitivity across all classes (Foody, 2002; Kramarczyk and Hejmanowska, 2025).

In addition, **Cohen's Kappa** and the **Matthews Correlation Coefficient (MCC)** were employed as robust agreement measures that are less sensitive to class imbalance and provide a more comprehensive evaluation of multi-class classification performance (Foody, 2023; Kramarczyk and Hejmanowska, 2025).

Finally, considering the pixel-based nature of the classification and the very high spatial resolution of UAV imagery, particular attention was given to the spatial behavior of the predictions, including the presence of salt-and-pepper noise (Gini et al., 2018). These aspects are further analyzed in the Results section.

3.3.7 Reproducibility notes

All preprocessing, feature extraction, model training, and evaluation steps were implemented within a script-based workflow to ensure full reproducibility (Zhu et al., 2018; Matrone et al., 2025). A fixed random seed (seed = 42) was used for all stochastic processes, including pixel sampling and data splitting, following reproducible machine learning practices in remote sensing (Maxwell et al., 2018; Persello and Bruzzone, 2016).

To prevent information leakage and ensure a fair evaluation, feature standardization parameters were computed exclusively from the training data and subsequently applied to the validation and test sets, in accordance with established guidelines for robust model evaluation under limited data conditions (Roberts et al., 2017; Maxwell et al., 2018). In addition, the validation subset was derived exclusively from the training data, while the test set remained completely independent and was not used during model training, parameter tuning, or early stopping.

3.3.8 Data-centric enhancement strategy

The methodological framework of this research follows a data-centric perspective, where emphasis is placed on improving the quality and representativeness of the input data rather than increasing model complexity. This choice is motivated by the characteristics of very high-resolution UAV imagery, where data volumes are large while labeled reference data remain limited and imbalanced (Zhu et al., 2018).

Within this context, the main strategy is the progressive enrichment of the feature space. Instead of relying on more complex models, the approach enhances the descriptive power of the input data by integrating complementary information from spectral bands, vegetation indices, and spatial texture descriptors.

A key component of this strategy is the use of a controlled experimental design, in which the classifier, sampling strategy, class weighting, and evaluation metrics are kept constant across all configurations. This allows a direct assessment of the effect of feature-space enrichment on classification performance.

This approach is particularly suitable under limited data conditions, where increasing model complexity may lead to overfitting and unstable predictions, while improved feature representation provides a more reliable basis for learning.

3.4 Post-processing strategy

To improve the spatial quality and reduce noise in the classification results, a set of post-processing techniques was applied. These methods aim to enhance spatial coherence and reduce artifacts introduced by pixel-based classification.

3.4.1 Majority filtering

To reduce salt-and-pepper noise and improve the spatial coherence of the classification results, a majority filtering approach was applied as a post-processing step. This method operates by reassigning the class label of each pixel based on the most frequent class within a local neighborhood, thereby enforcing spatial consistency in the classification map (Gonzalez and Woods, 2008).

Two neighborhood configurations were considered: a **3×3** window and a **5×5** window. The smaller window preserves finer spatial details while reducing isolated misclassified pixels, whereas the larger window provides stronger smoothing by emphasizing dominant local patterns. This allows for a controlled balance between noise reduction and preservation of spatial structures.

The filtering process was applied exclusively to valid pixels, while NoData values were preserved to maintain the integrity of masked areas. By incorporating local spatial context after classification, majority filtering improves the visual consistency of the output map and reduces fragmentation caused by independent pixel-wise predictions.

3.4.2 Small patch removal

In addition to majority filtering, a small patch removal strategy was applied to further improve the spatial structure of the classification results. This approach targets isolated regions of pixels that form small connected components, which are often associated with classification noise rather than meaningful vegetation patterns (Soille, 2003).

The method is based on connected-component analysis, where groups of adjacent pixels belonging to the same class are identified as individual patches. Patches below a predefined size threshold are considered unreliable and are reassigned to the dominant class in their surrounding neighborhood.

Two threshold levels were considered to evaluate the effect of this operation at different spatial scales, specifically patches smaller than **250** pixels and **500** pixels. The smaller threshold allows the preservation of finer spatial details, while the larger threshold results in stronger generalization of the classification map by removing a greater number of small and fragmented regions.

This post-processing step enhances the structural consistency of the classification output and improves its interpretability, particularly in heterogeneous environments where pixel-based classification tends to produce fragmented patterns.

4 Results

4.1 Baseline classification results (9-band Dataset)

The baseline model, based on the 9-band feature representation, demonstrated satisfactory classification performance. The model achieved an Overall Accuracy (OA) of 0.9547 and a Macro F1-score of 0.9339, indicating that spectral bands combined with vegetation indices provide a strong foundation for vegetation discrimination.

However, the Balanced Accuracy of 0.9314 suggests that performance was not uniform across all classes, with reduced accuracy observed for minority classes and those exhibiting strong spectral similarity.

Quantitative evaluation was conducted on a test set of 32,436 pixels, obtained through a class-wise 80/20 split. Therefore, the reported metrics reflect model performance within the labeled data domain.

Class-wise analysis indicates that structurally distinct and well-represented classes, such as Upright Matgrass, achieved near-perfect performance. In contrast, classes such as Blueberry and Red Raspberry showed lower precision or recall, reflecting the limitations of purely spectral, single-pixel information in separating spectrally similar vegetation types.

From a spatial perspective, these limitations are also evident in the classification outputs. Despite high overall metrics, localized errors and scattered misclassifications are observed, particularly in transition zones and along class boundaries. This highlights that overall accuracy alone is insufficient for evaluating the quality of high-resolution classification maps.

4.1.1 Model convergence behavior

The training process of the baseline model exhibited stable and consistent convergence behavior. The model was trained for a total of 2,500 boosting iterations, and early stopping was not triggered. The optimal model was obtained at iteration 2,499, indicating that performance continued to improve throughout the training process without signs of premature convergence.

The evaluation of the loss function further confirms the stability of the training process. The training mlogloss reached a value of 0.0025, while the validation mlogloss stabilized at 0.0464, resulting in a generalization gap of 0.0439.

This relatively small gap between training and validation loss suggests that the model maintained good generalization capability and did not suffer from significant overfitting, despite the limited and imbalanced nature of the labeled dataset.

Metric	Value
Total boosting iterations	2,500
Best iteration	2,499
Early stopping	Not triggered
Train mlogloss	0.0025
Validation mlogloss	0.0464
Generalization gap	0.0439

Table 6 - Summary of the training convergence behavior of the baseline model, including boosting iterations, loss values, and generalization gap, indicating stable learning without significant overfitting.

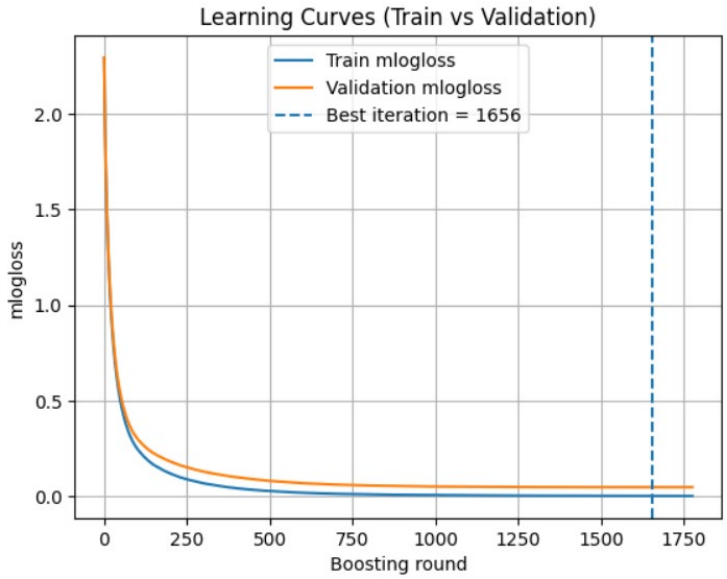


Figure 9 - Learning curves showing training and validation mlogloss over boosting iterations, indicating stable convergence and a limited generalization gap.

4.2 Impact of texture features

The incorporation of spatial texture features and the resulting 27-band feature representation led to a clear improvement across all evaluation metrics (Table 7). The Macro F1-score increased from 0.9339 to 0.9742, corresponding to an absolute improvement of approximately 4%, while Balanced Accuracy improved from 0.9314 to 0.9684. These results indicate that the main performance gains were achieved for minority and spectrally challenging classes.

Additional metrics confirm this trend. Overall Accuracy increased from 0.9547 to 0.9866, while both Cohen's Kappa and the Matthews Correlation Coefficient (MCC) improved from 0.9468 to 0.9841. Similarly, the Weighted F1-score increased from 0.9551 to 0.9865, demonstrating consistent improvements across both balanced and imbalanced evaluation perspectives.

Class-wise analysis (Table 8) shows that all classes benefited from the inclusion of texture features. The largest improvement was observed for Blueberry (+0.10), followed by Red Raspberry (+0.05)

and Tussock Fescue (+0.05). These classes exhibit strong spectral similarity to neighboring vegetation types but differ in spatial structure, which becomes more distinguishable when local texture information is incorporated.

The effect of feature enrichment on the spatial characteristics of the classification outputs is further examined in Section 4.4. Overall, the observed improvements support the data-centric design of the proposed framework, where performance gains are primarily driven by feature-space enrichment rather than increased model complexity.

Metric	9-band (Baseline)	27-band (Texture enhanced)
Overall Accuracy (OA)	0.9547	0.9866
Macro F1-score	0.9339	0.9742
Balanced Accuracy	0.9314	0.9684
Weighted F1-score	0.9551	0.9865
Cohen's Kappa	0.9468	0.9841
Matthews Correlation Coefficient (MCC)	0.9464	0.9837
Best boosting iteration	2499	1656

Table 7 - Comparison of classification performance between the 9-band baseline and the 27-band texture-enhanced model for the Brocan Lake dataset, showing consistent improvements across all evaluation metrics.

Class	F1 (9-band)	F1 (27-band)	ΔF1
Alpine Dock	0.96	0.99	+0.03
Lady Fern	0.91	0.95	+0.04
Green Alder	0.93	0.98	+0.05
Male Fern	0.92	0.96	+0.04
Dwarf Juniper	0.95	0.99	+0.04
Matgrass	0.96	0.98	+0.02
Red Raspberry	0.87	0.92	+0.05
Rusty Leaf Rhododendron	0.94	0.98	+0.04
Tussock Fescue	0.93	0.98	+0.05
Upright Matgrass	0.98	1.00	+0.02
Blueberry	0.87	0.97	+0.10

Table 8 - Comparison of classification performance between the 9-band baseline and the 27-band texture-enhanced model for the Brocan Lake dataset, showing consistent improvements across all evaluation metrics.

4.3 Transferability results

To evaluate the robustness and generalization capability of the proposed framework, an independent dataset from a second alpine lake environment was used. This dataset represents a larger and more complex scenario, characterized by different spatial structures, class distributions, and increased data volume.

Unlike the within-scene evaluation conducted in Case Study 1, this analysis examines the performance of the model under spatially distinct conditions. The independent dataset was used exclusively for validation and was not involved in model design, training, or parameter tuning.

The corresponding quantitative results and class-wise performance are presented in Section 4.3.1.

4.3.1 Classification performance on independent validation dataset (27-band Model)

The performance of the proposed framework was further evaluated on an independent dataset acquired over Lake Vej del Bouc, representing a larger and more complex alpine environment. The classification was performed using the 27-band feature configuration without modification of the model structure or training strategy, ensuring a consistent assessment of model transferability.

The dataset covers a very large spatial extent, consisting of approximately 849,598,608 total pixels, of which 237,882,024 pixels remain valid after masking. Only 378,748 pixels (~0.16% of the valid area) are labeled, highlighting the extreme label scarcity conditions under which the model is evaluated.

The model achieved an Overall Accuracy (OA) of 0.9881 and a Macro F1-score of 0.9813 on the independent test set. The Weighted F1-score was 0.9880, and the Balanced Accuracy reached 0.9798. In addition, Cohen's Kappa and the Matthews Correlation Coefficient (MCC) were both equal to 0.9863, indicating strong agreement between predicted and reference labels.

Class-wise evaluation shows consistently high performance across all vegetation classes, with all classes achieving F1-scores greater than or equal to 0.96. Dominant classes such as Paniculate grass, Matgrass, and Green Alder achieved near-perfect performance, while less represented classes also maintained strong classification results.

Full-scene inference was conducted in a block-wise manner over 3,304 processing windows, covering all valid pixels. This confirms the scalability of the proposed framework for large-scale UAV datasets.

The training process exhibited stable convergence behavior. The model was trained for 2,500 boosting iterations, with no early stopping triggered, and the optimal model was obtained at iteration 2,499. The training mlogloss reached 0.0025, while the validation mlogloss stabilized at 0.0464, resulting in a generalization gap of 0.0439.

These results indicate that the proposed framework maintains high classification performance under different spatial conditions and data distributions.

Metric / Attribute	Value
Overall Accuracy (OA)	0.9881
Macro F1-score	0.9813
Weighted F1-score	0.9880
Balanced Accuracy	0.9798
Cohen's Kappa	0.9863
Matthews Correlation Coefficient (MCC)	0.9860
Number of classes	12
Total pixels	849,598,608
Total valid pixels	237,882,024
Total labeled pixels	378,748
Label scarcity (valid pixels)	≈ 0.16%

Table 9 - Summary of classification performance and dataset characteristics for the Lake Vej del Bouc dataset, demonstrating high accuracy and robustness under extreme label scarcity.

Class	Precision	Recall	F1	Support
Adenostyles	0.97	0.97	0.97	1,093
Alpine Dock	0.97	0.95	0.96	873
Black Sedge	0.98	0.97	0.97	741
Blueberry	0.97	0.97	0.97	1,178
Cotton Grass	0.99	0.99	0.99	9,746
Dwarf Juniper	0.98	0.98	0.98	8,845
Goosefoot	0.98	0.98	0.98	1,667
Green Alder	0.99	0.99	0.99	10,067
Male Fern	0.99	0.98	0.98	3,272
Matgrass	0.99	0.99	0.99	12,662
Paniculate grass	1.00	1.00	1.00	15,350
Rhododendron	0.98	0.98	0.98	4,090

Table 10 - Per-class precision, recall, and F1-scores for the Lake Vej del Bouc dataset, indicating consistently high performance across all vegetation classes.

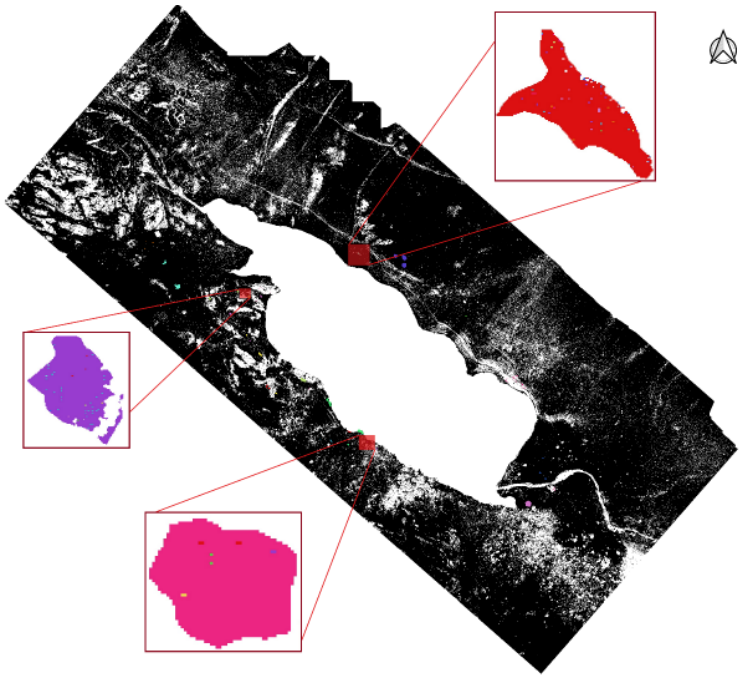


Figure 10 - Predicted vegetation classification map with representative sample patches extracted from different locations, illustrating spatial variability across the study area.



Figure 11 - Predicted Example of Class 1 (Adenostyles)



Figure 12 - Ground Truth Reference of Class 1 (Adenostyles)



Figure 13 Predicted Example of Class 5 (Cotton Grass)



Figure 14 Ground Truth Reference of Class 5 (Cotton Grass)



Figure 15 Predicted Example of Class 6 (Dwarf Juniper)



Figure 16 Ground Truth Reference of Class 6 (Dwarf Juniper)

4.4 Spatial analysis and salt-and-pepper noise

While the quantitative results presented in Sections 4.1 to 4.3 indicate high classification performance, the spatial characteristics of the generated maps reveal additional aspects that are not fully captured by standard evaluation metrics. In very high-resolution UAV imagery, pixel-based classification often produces spatial artifacts such as isolated misclassified pixels and fragmented class patterns, commonly referred to as salt-and-pepper noise (Gini et al., 2018).

These effects arise from the independent treatment of pixels, where spatial context is not explicitly incorporated during prediction. As a result, even models with high numerical accuracy may produce classification maps with limited spatial coherence, particularly in heterogeneous environments characterized by gradual transitions and strong spectral overlap.

To further investigate this behavior, a spatial analysis of the classification outputs was conducted. This analysis focuses on fragmentation patterns, boundary consistency, and the distribution of isolated errors across the mapped area.

The following subsections examine these aspects for both the primary case study and the independent validation dataset, and explore the relationship between numerical performance and spatial quality.

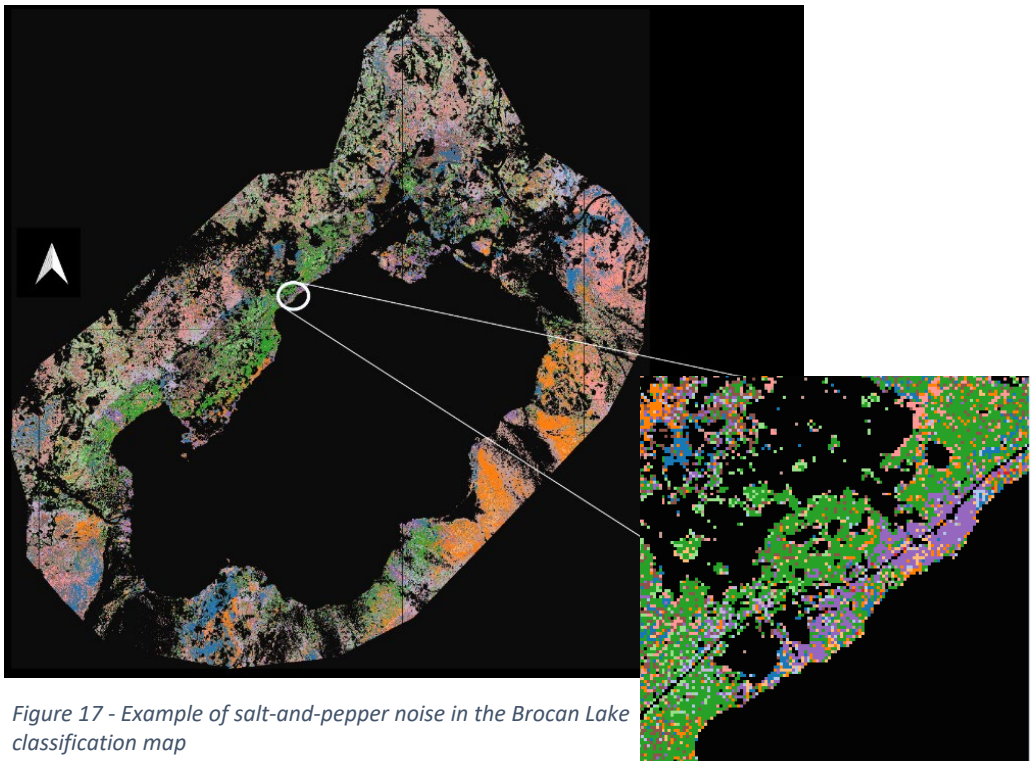


Figure 17 - Example of salt-and-pepper noise in the Brocan Lake classification map

4.4.1 Spatial behavior in case Study 1 (9-band vs 27-band)

The spatial behavior of the classification results in the primary study area shows a clear difference between the baseline (9-band) and texture-enhanced (27-band) configurations.

In the baseline model, the classification output exhibits a high level of salt-and-pepper noise, characterized by scattered misclassified pixels and fragmented class regions. These effects are particularly evident in transition zones and in areas with strong spectral similarity, resulting in irregular and discontinuous class boundaries.

In contrast, the inclusion of texture features leads to improved spatial consistency. The 27-band model produces more homogeneous class regions, with fewer isolated pixel errors and smoother transitions between neighboring classes. This improvement is especially noticeable in heterogeneous areas, where local spatial context supports the separation of structurally different but spectrally similar vegetation types.

The reduction in edge density from 24.49% to 15.79% quantitatively confirms the decrease in spatial fragmentation, indicating a more coherent spatial organization of the classified regions.

A summary of the main spatial differences between the two configurations is presented in Table 11.

Spatial Indicator	9-band (Baseline)	27-band (TextureEnhanced)	Interpretation
Salt-and-pepper noise	Pronounced	reduced	Fewer isolated misclassified pixels
Edge density (fragmentation)	24.49%	15.79%	Lower spatial fragmentation
Spatial coherence of regions	Low	High	More homogeneous class regions
Boundary stability	Irregular	More consistent	Smoother transitions between classes
Most improved class (structure-sensitive)	Blueberry	Blueberry	Highest improvement (+0.10 F1)
Additional improved minority class	Red Raspberry	Red Raspberry	Notable improvement (+0.05 F1)

Table 11 - Spatial comparison between baseline (9-band) and texture-enhanced (27-band) models in Brocan lake

4.4.2 Spatial behavior in independent validation dataset (27-band)

The spatial behavior of the classification results on the independent dataset reveals a more complex pattern compared to the primary case study. Despite the high quantitative performance reported in Section 4.3.1, the generated maps still exhibit noticeable salt-and-pepper noise and localized fragmentation.

These effects are particularly evident in heterogeneous regions and transition zones between vegetation classes, where small clusters of misclassified pixels persist. This behavior reflects the inherent limitation of pixel-based classification, in which spatial context is not explicitly incorporated during prediction.

At the full-scene scale, this effect becomes more pronounced. The classification was applied to over 237 million valid pixels under conditions of extreme label scarcity (~0.16%), where even minor spectral or radiometric variations can lead to inconsistent predictions between neighboring pixels.

This spatial fragmentation is quantitatively reflected in an edge density of 15.79%, indicating that a considerable proportion of class boundaries remain irregular despite the strong numerical performance.

Analysis of class adjacency patterns further reveals strong spatial intermixing between vegetation types. The most prominent boundary occurs between Matgrass and Paniculate grass, exceeding 12 million adjacency edges, indicating a highly dynamic transition zone.

To further characterize the spatial structure of the landscape, coverage and diversity indicators were analyzed. Coverage metrics describe how much of the area is occupied by dominant classes, while Shannon entropy measures class diversity and the Gini coefficient quantifies inequality in class distribution. Together, these indicators provide complementary insight into landscape composition and structural complexity.

A summary of the main spatial indicators and class boundary relationships is presented in Tables 12 and 13.

Indicator	Value	Interpretation
Total pixels	849,598,608	Full dataset size
Total valid pixels	237,882,024	Area used for classification
Labeled pixels	378,748	Reference data size
Label scarcity (valid area)	~0.16%	Extremely limited ground truth
Edge density (fragmentation)	15.79%	Degree of spatial fragmentation
Top-3 coverage share	52.75%	Dominance of few major classes
Top-5 coverage share	74.66%	Increasing concentration of dominant classes
Top-8 coverage share	90.39%	Majority of landscape covered by limited classes
Shannon entropy	3.203	High class diversity and landscape complexity
Gini coefficient	0.455	Moderate inequality in class distribution

Table 12 - Quantitative spatial indicators of classification results in the independent validation dataset (27-band model)

Class A	Class B	Boundary Edges
Matgrass	Paniculate grass	12,444,685
Dwarf Juniper	soil/rock	9,845,141
Dwarf Juniper	Matgrass	7,969,515
Matgrass	soil/rock	6,970,431
Cotton Grass	Dwarf Juniper	5,559,303
Paniculate grass	soil/rock	4,880,194
Cotton Grass	Green Alder	3,677,488
Goosefoot	Matgrass	3,601,013
Dwarf Juniper	Paniculate grass	3,383,374
Goosefoot	Paniculate grass	3,312,338
Alpine Dock	Goosefoot	2,898,962
Dwarf Juniper	Rhododendron	2,876,270
Alpine Dock	Matgrass	2,780,628
Matgrass	Rhododendron	2,610,753
Green Alder	Rhododendron	2,375,593

Table 13 - Strongest class boundaries based on 4-neighbor connectivity, indicating spatial adjacency and interaction between vegetation classes

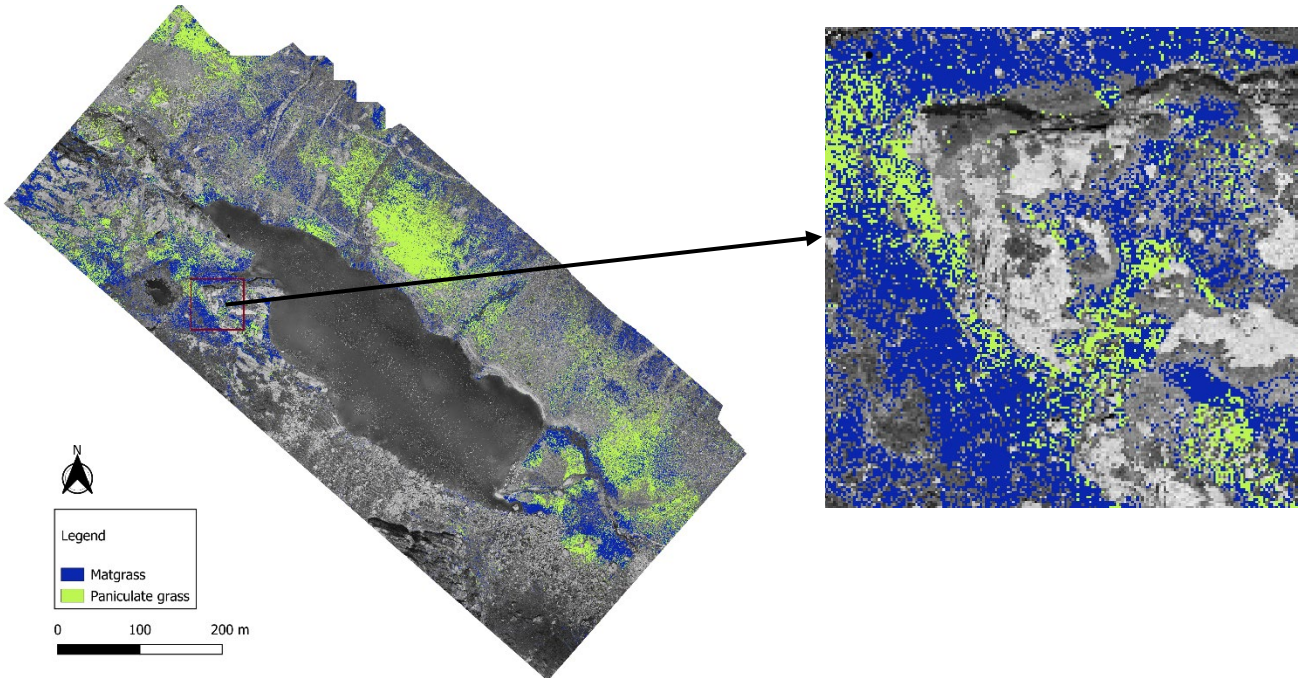


Figure 18 - Spatial intermixing between Matgrass and Paniculate grass, the strongest class boundary identified in the study area

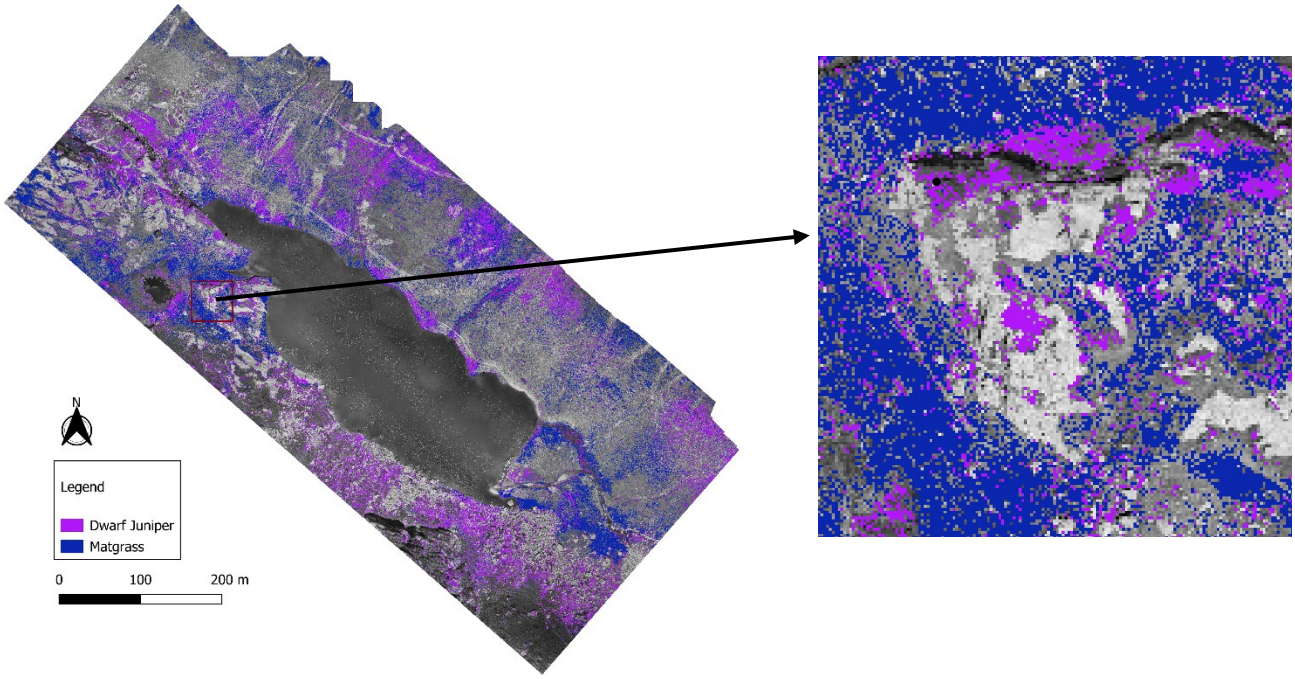


Figure 19 - Spatial intermixing between Matgrass and Dwarf Juniper, the strongest class boundary identified in the study area

4.4.3 Spatial leakage analysis

The reported classification performance is influenced by spatial autocorrelation within the dataset, which may lead to spatial leakage between training and test samples. In very high-resolution UAV imagery (3 cm), neighboring pixels often share highly similar spectral and structural properties, making randomly sampled pixels partially dependent (Roberts et al., 2017; Ploton et al., 2020).

In this study, data splitting was performed at the pixel level using a class-wise strategy, with approximately 80% of labeled pixels used for training and 20% for testing. Training and test samples were generated as disjoint subsets of pixel coordinates, ensuring that no identical pixel locations were shared between the two sets. While this approach maximizes the use of the extremely limited labeled data (~0.16% of the valid area), it may still include spatially adjacent pixels in both training and test subsets.

As a result, any potential optimistic bias does not arise from duplicated samples, but rather from spatial autocorrelation between neighboring pixels, particularly in homogeneous regions where spectral similarity is high. This distinction is important, as it indicates that the evaluation remains valid at the pixel level, while acknowledging the inherent spatial dependency of the data.

Consequently, although the test-set performance is high (Macro F1 ≈ 0.98), it does not fully represent the challenges observed in full-scene inference, where spatial heterogeneity, boundary effects, and unlabeled variability are more pronounced.

4.4.4 Discrepancy between test metrics and full-scene inference

Despite the high classification performance reported on the test set, a clear discrepancy is observed when the model is applied to full-scene inference. While quantitative metrics such as Overall Accuracy (0.9881) and Macro F1-score (0.9813) indicate strong performance, spatial analysis reveals that the generated maps still contain localized inconsistencies and fragmented patterns.

This discrepancy arises from the difference between evaluation conditions and real-world application. The test metrics are computed on a limited set of labeled pixels (378,748 samples, ~0.16% of the valid area), whereas full-scene inference is performed over 237,882,024 valid pixels, including regions that are not directly represented in the reference data. As a result, the model is exposed to a wider range of conditions, including out-of-distribution patterns that were not sufficiently captured during training.

In addition, the pixel-based classification framework predicts each pixel independently without explicit spatial constraints. Under these conditions, local variations in illumination, shadows, radiometric noise, and image blur can increase intra-class variability and lead to inconsistent predictions between neighboring pixels, particularly in heterogeneous areas. These effects manifest as localized noise and small clusters of misclassified pixels, especially along class boundaries and transition zones.

This behavior is quantitatively reflected in the spatial fragmentation of the classification outputs. The edge density of the full-scene prediction remains at 15.79%, indicating persistent irregular class boundaries despite high numerical performance. Post-processing analysis further highlights this discrepancy, showing that between 13.29% and 34.03% of valid pixels are modified during spatial refinement, which confirms the presence of substantial noise and structural instability in the initial predictions.

Overall, this discrepancy should not be interpreted as a simple weakness of the model. Instead, it reflects a structural characteristic of the problem, arising from the combination of limited labeled data, spatial autocorrelation between neighboring pixels, and the inherent limitations of independent pixel-based classification in very high-resolution imagery.

A summary of the discrepancy between test-set performance and full-scene behavior is presented in Table 14.

Indicator	Value	Interpretation
Test Macro F1-score	0.9813	High classification performance on labeled data
Overall Accuracy (test)	0.9881	Strong agreement with reference samples
Edge density (full scene)	15.79%	Remaining spatial fragmentation in predictions
Pixels modified after refinement	13.29% – 34.03%	Significant portion of predictions required correction
Total valid pixels	237,882,024	Large-scale inference domain
Labeled pixels	378,748 (~0.16%)	Extreme label scarcity

Table 14 - Indicators highlighting the discrepancy between test-set performance and full-scene classification behavior.

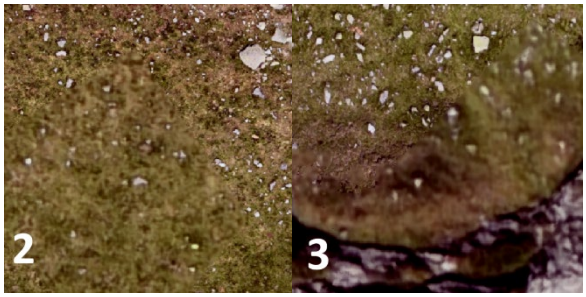


Figure 20 - Blur noise

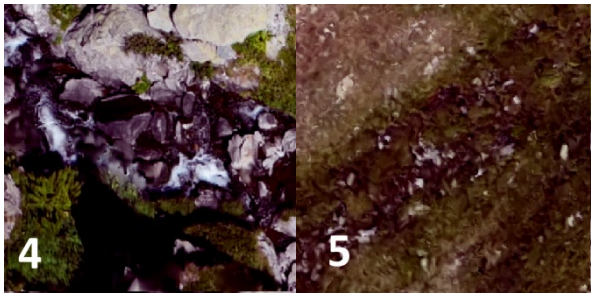


Figure 21 - Radiometric noise



Figure 22 - Shadow

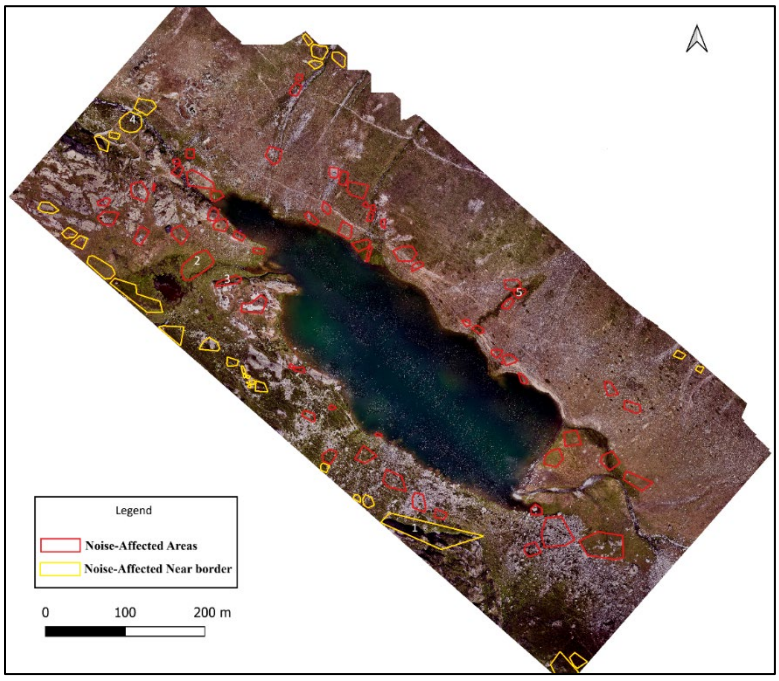


Figure 23 - Noise-affected areas in the UAV imagery highlighted in red

4.5 Post-processing for spatial refinement

To address the spatial inconsistencies observed in the classification outputs, post-processing techniques were applied to improve the structural quality of the resulting maps. These methods aim to reduce salt-and-pepper noise, eliminate small fragmented regions, and enhance spatial coherence.

Post-processing is particularly relevant in pixel-based classification, where predictions are generated independently and spatial context is not explicitly enforced. As a result, even models with high numerical accuracy may produce fragmented and locally inconsistent maps.

In this study, two complementary approaches were implemented: majority filtering and small patch removal. The first enforces local neighborhood consistency, while the second removes structurally isolated regions. Their impact is evaluated under different configurations in the following subsections.

4.5.1 Majority filtering (3×3 and 5×5)

To improve the spatial coherence of the classification outputs, a majority filtering approach was applied as a post-processing step. This method reassigns the class label of each pixel based on the most frequent class within a local neighborhood, thereby enforcing local spatial consistency.

Two neighborhood configurations were evaluated: a 3×3 window and a 5×5 window. The 3×3 configuration provides moderate smoothing while preserving fine spatial details, whereas the 5×5 configuration applies stronger smoothing by emphasizing dominant local patterns.

The filtering process was applied to the full-scene classification results using a window-based implementation. NoData pixels were preserved to maintain the integrity of masked areas.

The application of majority filtering resulted in substantial changes in the classification outputs. The 3×3 configuration modified 31,611,656 pixels (13.29% of valid pixels), while the 5×5 configuration affected 48,250,932 pixels (20.28%). These values indicate a significant reduction of isolated misclassified pixels and fragmented regions.

The larger 5×5 window produced a stronger smoothing effect, leading to more homogeneous regions, while the 3×3 configuration maintained a better balance between noise reduction and preservation of fine-scale spatial structures.

Configuration	Kernel Size	Pixels Modified	Percentage of Valid Pixels	Spatial Effect
Majority Filter	3×3	31,611,656	13.29%	Moderate smoothing, preserves fine details
Majority Filter	5×5	48,250,932	20.28%	Strong smoothing, higher spatial homogenization

Table 15 - Impact of majority filtering with different neighborhood sizes on full-scene classification outputs.

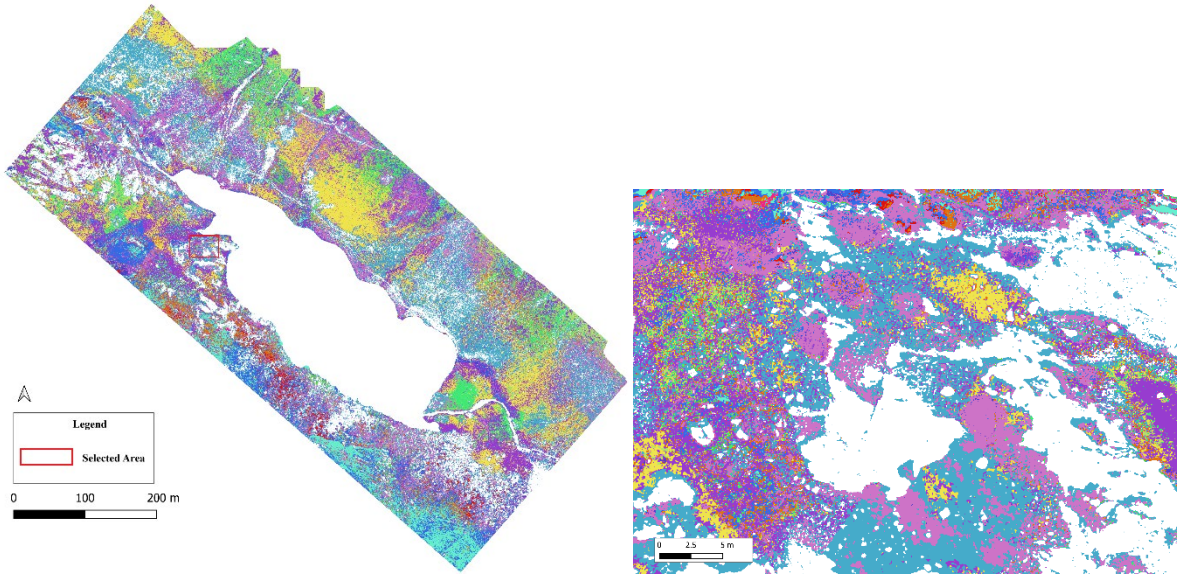


Figure 24 - Original vegetation classification produced by the XGBoost model before post-processing

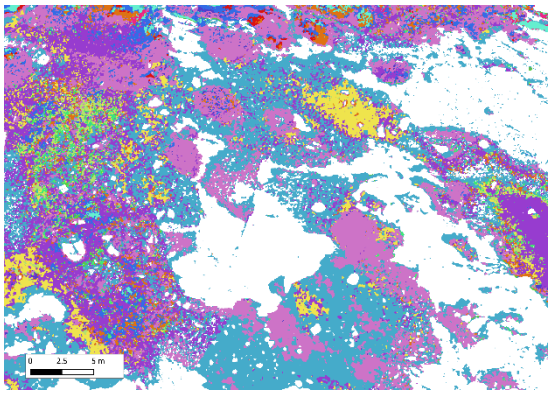


Figure 25 - Vegetation classification after spatial smoothing using the Majority Filter (3×3 window)

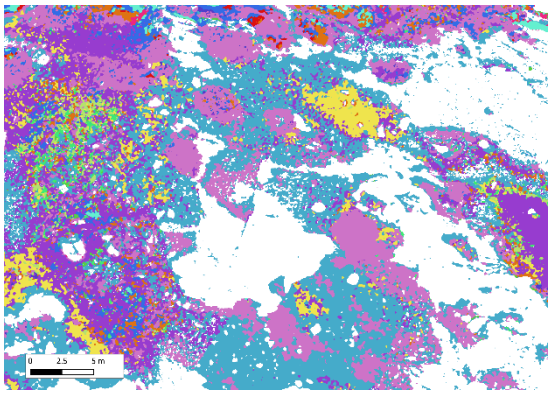


Figure 26 - Vegetation classification after spatial smoothing using the Majority Filter (5×5 window)

4.5.2 Small patch removal (<250 px and <500 px)

To further improve the structural consistency of the classification results, a small patch removal approach was applied as a post-processing step. This method is based on connected-component analysis, where spatially connected pixels belonging to the same class are identified as individual

regions. Patches below a predefined size threshold are considered unreliable and are reassigned to the dominant class in their surrounding neighborhood.

Two threshold levels were evaluated to assess the effect of this operation at different spatial scales: 250 pixels and 500 pixels. The smaller threshold preserves finer spatial details while removing very small and isolated regions, whereas the larger threshold results in stronger generalization by eliminating a greater number of fragmented patches.

The application of this method resulted in substantial modifications to the classification outputs. The threshold of 250 pixels led to the reassignment of 72,232,771 pixels (30.37% of valid pixels), while the 500-pixel threshold affected 80,957,225 pixels (34.03%).

These results indicate that a significant portion of the initial classification consisted of small and fragmented regions, particularly in heterogeneous areas. Increasing the threshold produced more homogeneous spatial patterns, at the cost of reducing fine-scale detail.

Configuration	Patch Size Threshold	Pixels Modified	Percentage of Valid Pixels	Spatial Effect
Small Patch Removal	<250 px	72,232,771	30.37%	Removes small noisy regions, preserves medium-scale structure
Small Patch Removal	<500 px	80,957,225	34.03%	Strong generalization, reduces fragmentation further

Table 16 - Impact of small patch removal using different size thresholds on full-scene classification outputs.

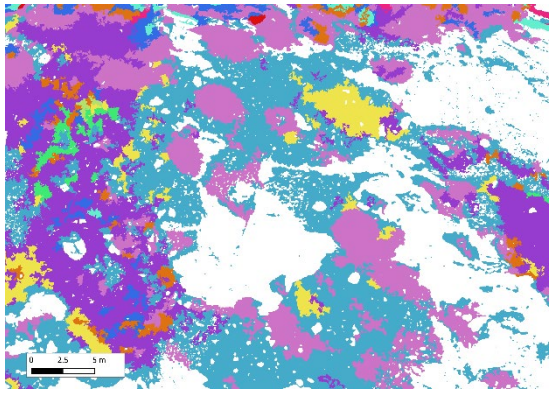


Figure 27 - Vegetation classification after small-patch removal with a minimum patch size threshold of 250 pixels

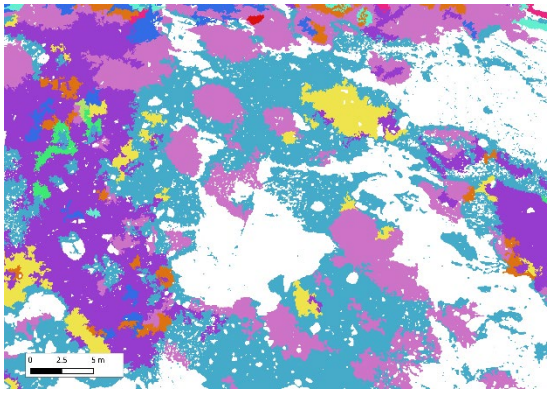


Figure 28 - Vegetation classification after small-patch removal with a minimum patch size threshold of 500 pixels

4.5.3 Comparative impact of post-processing methods

The impact of majority filtering and small patch removal was compared across different configurations.

Majority filtering operates at the local neighborhood level and is effective in reducing isolated pixel noise. The 3×3 configuration preserves fine spatial details, while the 5×5 configuration produces stronger smoothing and more homogeneous regions.

In contrast, small patch removal operates at a structural level by eliminating connected regions below a defined size threshold. This method resulted in a substantially higher number of modified pixels, indicating a stronger effect on the overall spatial organization of the classification.

The <250 px threshold preserves medium-scale structures while removing small fragmented regions, whereas the <500 px configuration produces stronger generalization by further reducing fragmentation.

Overall, majority filtering is more effective for local noise reduction, while small patch removal provides stronger structural refinement.

Method	Configuration	Pixels Modified	% of Valid Pixels
Majority Filtering	3×3	31,611,656	13.29%
Majority Filtering	5×5	48,250,932	20.28%
Small Patch Removal	<250 px	72,232,771	30.37%
Small Patch Removal	<500 px	80,957,225	34.03%

Table 17 - Comparative impact of post-processing methods on spatial refinement of classification outputs.

ID	Class	Δ 3×3 (%)	Δ 5×5 (%)	Δ <250 px (%)	Δ <500 px (%)
1	Adenostyles	+0.011	-0.344	-1.685	-1.806
2	Alpine Dock	+0.043	-0.167	-1.315	-1.560
3	Black Sedge	-0.097	-0.236	-0.668	-0.789
4	Blueberry	-0.003	-0.058	-0.368	-0.472
5	Cotton Grass	+0.199	+0.083	-0.540	-0.758
6	Dwarf Juniper	+0.593	+0.599	-0.373	-0.627
7	Goosefoot	+0.271	+0.414	+1.193	+1.239
8	Green Alder	+0.004	+0.086	+0.124	+0.168
9	Male Fern	-0.241	-0.358	-0.550	-0.610
10	Matgrass	-0.050	-0.037	+0.604	+0.697
11	Paniculate grass	-0.056	+0.238	+1.453	+1.825
12	Rhododendron	-0.702	-0.952	-0.867	-0.809

Table 18 - Percentage Change (Δ%) After Post-Processing Methods

4.6 Summary of results

This section summarizes the main findings of the study across the primary case study and the independent validation dataset, focusing on both quantitative performance and spatial behavior.

4.6.1 Case study 1 (Baseline vs Texture)

In the primary study area, the baseline configuration based on spectral bands and vegetation indices achieved satisfactory classification performance. However, the integration of spatial texture features led to a clear and consistent improvement across all evaluation metrics, particularly for spectrally overlapping and minority classes.

In addition to numerical gains, the texture-enhanced configuration significantly improved the spatial quality of the classification outputs. A substantial reduction in salt-and-pepper noise and spatial fragmentation was observed, resulting in more coherent class regions and more stable boundaries. These results highlight the importance of incorporating spatial information for accurate vegetation mapping in very high-resolution UAV imagery.

4.6.2 Independent validation dataset (27-band + spatial refinement)

The evaluation on the independent dataset confirmed the robustness of the proposed framework under different spatial conditions and increased data complexity. The model maintained high classification performance, demonstrating its ability to generalize beyond the primary study area.

However, spatial analysis revealed the presence of residual noise and localized inconsistencies in full-scene predictions, despite strong numerical metrics. This discrepancy reflects the limitations of pixel-based classification and the influence of spatial autocorrelation.

The application of post-processing techniques further improved the structural quality of the classification outputs. Majority filtering effectively reduced local noise, while small patch removal significantly enhanced spatial consistency by eliminating fragmented regions. The combination of feature-space enrichment and spatial refinement resulted in more reliable and interpretable classification maps at the full-scene scale.

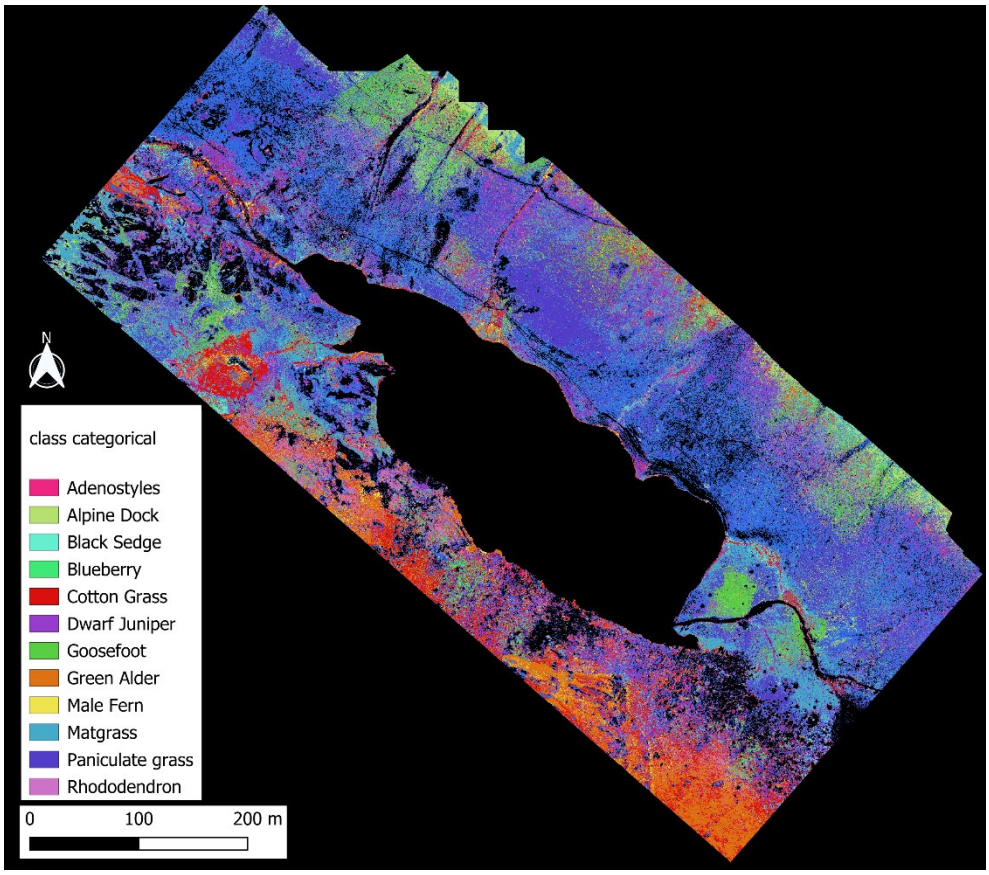


Figure 29 - Full-scene vegetation classification map showing the spatial distribution of predicted vegetation Vej Del Bouc)

5. Discussion

5.1 Interpretation of main findings

The results of this research indicate that, in pixel-based vegetation classification using very high-resolution UAV imagery under severe reference-data scarcity, model performance is more strongly influenced by feature representation than by the complexity of the classifier.

In the primary case study, only a very small fraction of the available data was labeled (about 0.12% of valid pixels), which significantly constrains the learning process. Under such conditions, the observed behavior aligns with previous studies emphasizing the importance of data representation and feature design over model complexity (Persello and Bruzzone, 2016; Zhu et al., 2018; Fan et al., 2025).

The baseline configuration, based on spectral bands and vegetation indices, achieved satisfactory overall performance. However, classes with strong spectral similarity, such as Blueberry and Red Raspberry, showed higher confusion, reflecting the limitations of purely spectral information in very high-resolution environments (Gini et al., 2018; Zou et al., 2021).

At the same time, the classification maps revealed clear spatial limitations, including scattered errors and fragmented patterns, particularly in transition zones. This indicates that performance measured on labeled samples does not fully represent model behavior at the full-scene scale, especially in large and heterogeneous datasets (Persello and Bruzzone, 2016; Ploton et al., 2020). This gap between test-set performance and full-scene behavior, discussed in Section 4.4.4, highlights an important aspect of model interpretation in large and label-sparse datasets.

The inclusion of texture features introduced local spatial context into the feature space, allowing the model to better capture structural differences between vegetation types. This resulted in improved classification performance, particularly for difficult classes, along with more coherent spatial patterns and smoother class boundaries.

A similar behavior is observed in the independent validation dataset, where the model maintains high numerical performance despite the much larger scale and similar label scarcity. However, spatial analysis still reveals residual fragmentation and localized inconsistencies, especially in areas with complex class interactions, confirming that these effects are not limited to a single dataset but are linked to the nature of the problem itself.

5.2 Role of texture features

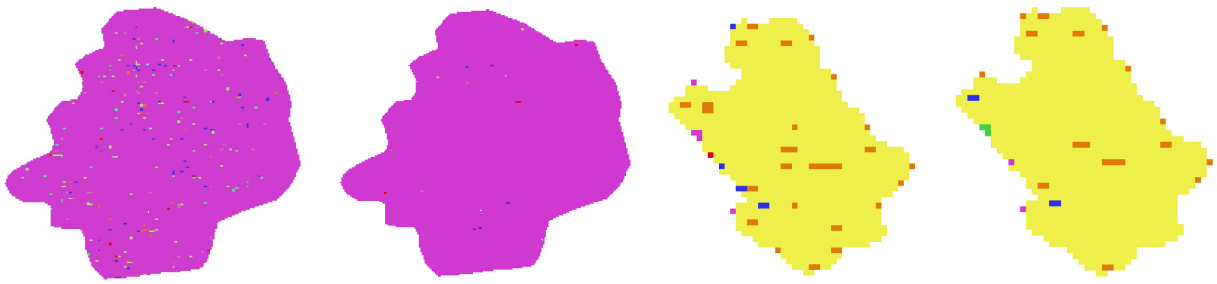
The results presented in Chapter 4 highlight the importance of incorporating spatial texture information in pixel-based vegetation classification, particularly under conditions of limited reference data. In the baseline configuration, spectral bands and vegetation indices provided a reasonable level of discrimination, but spatial instability and class confusion remained evident, especially for spectrally similar and under-represented classes.

The introduction of first-order texture features allowed the model to integrate local spatial context into the classification process. This led to a clear improvement in performance (e.g., Macro F1 increasing from 0.9339 to 0.9742), indicating that part of the limitation in the baseline approach was related to missing spatial information rather than model capacity.

The feature design also played a key role. Texture extraction was limited to a small set of informative layers, maintaining a balance between feature richness and model stability. This is particularly important in label-scarce scenarios, where excessive feature expansion may negatively affect generalization.

The impact of texture features is also visible in the spatial structure of the classification outputs. Compared to the spectral-only approach, the texture-enhanced model produces more coherent vegetation patches, with fewer isolated errors and reduced fragmentation. These improvements are especially noticeable in heterogeneous areas and transition zones, where local spatial context helps distinguish between similar vegetation types.

The reduction of isolated misclassified pixels suggests that texture features effectively transfer neighborhood-level information into the pixel-based decision process, allowing the model to better capture structural differences between classes while preserving a simple and scalable classification framework (Gini et al., 2018; Feng et al., 2015).



(a) (b) (c) (d)

Figure 30 - Visual comparison of pixel-level classification noise in Brocan Lake: (a) Dwarf Juniper (9-band model), (b) Dwarf Juniper (27-band model), (c) Red Bilberry (9-band model), and (d) Red Bilberry (27-band model), showing reduced noise and improvement

5.3 Comparison with similar studies

The results of this research are consistent with findings reported in vegetation mapping studies conducted in alpine, wetland, and transition-zone environments, where spectral information alone is often insufficient for reliable species discrimination in very high-resolution imagery. In such contexts, several studies highlight the importance of incorporating spatial or texture information to improve both classification accuracy and the interpretability of the resulting maps (Feng et al., 2015; Zou et al., 2021).

In comparison with previous work under similar ecological conditions, the proposed framework shows that a controlled enrichment of the feature space can improve both statistical performance and spatial quality without increasing model complexity. This behavior is in line with studies that demonstrate the effectiveness of combining multispectral information with carefully designed feature representations for vegetation classification in heterogeneous environments, particularly when dealing with ecologically complex and spatially variable habitats (Belcore et al., 2021). Such findings further support the idea that feature engineering plays a more critical role than algorithmic complexity in improving class separability under high-resolution conditions.

At the same time, the observations from this study are consistent with the broader literature in showing that, under conditions of extreme label scarcity, achieving fully stable and consistent predictions across the entire scene remains difficult. Even when classification metrics are high, spatial inconsistencies may still appear in large and heterogeneous datasets, especially in areas that are weakly represented in the reference data (Ploton et al., 2020; Zhu et al., 2018).

5.4 Robustness, class imbalance and generalization

A key challenge in this research is the presence of strong class imbalance together with extremely limited reference data. Under these conditions, the combination of inverse class weighting and feature-space enrichment helps improve the representation of minority classes and leads to more balanced classification performance across classes, as also reported in previous vegetation classification studies (He and Garcia, 2009; Fan et al., 2025).

At the same time, the results indicate that improvements in classification metrics do not fully resolve the issue of generalization at the full-scene scale. While performance on labeled data remains high, spatial inconsistencies are still observed when the model is applied to large and heterogeneous areas with limited ground truth (Persello and Bruzzone, 2016; Zhu et al., 2018).

This behavior is not related to data leakage between training and test sets, as these were constructed as disjoint pixel-coordinate subsets. Instead, it reflects the imbalance between the size of the

dataset and the availability of reference data, as well as the inherent spatial structure of high-resolution imagery. Under such conditions, neighboring pixels may share similar properties, while being only partially represented in the training data.

These limitations highlight the importance of combining data-centric strategies with approaches that account for spatial context and structural consistency, including feature design and post-processing methods, as discussed in recent remote sensing studies (Ploton et al., 2020).

5.5 Strengths and limitations

A key strength of this research lies in the design of a transparent and controlled data-centric framework, where the contribution of different feature components can be evaluated in a structured manner (Persello and Bruzzone, 2016; Gini et al., 2018). The results show that strong classification performance can be achieved even under conditions of extremely limited reference data, without increasing model complexity, which highlights the role of feature-space design in such scenarios (Fan et al., 2025).

In addition to numerical performance, improvements are also reflected in the spatial characteristics of the classification outputs. The integration of texture features contributes to more coherent class regions and reduced fragmentation, supporting more interpretable vegetation maps, particularly in heterogeneous environments (Feng et al., 2015; Zou et al., 2021)

At the same time, several limitations should be considered. The scarcity of reference data remains a major constraint, affecting both model training and the reliability of performance evaluation. In addition, although training and test samples were constructed as disjoint pixel-coordinate subsets, spatial dependence may still arise due to autocorrelation between neighboring pixels. This may slightly influence the reported accuracy and limit the representativeness of the test results, particularly in homogeneous regions (Persello and Bruzzone, 2016; Ploton et al., 2020).

Another limitation is related to the temporal dimension of the data. The analysis is based on a single acquisition period, which does not allow direct assessment of seasonal or phenological variations in vegetation patterns (Matrone et al., 2025).

5.6 Data-centric vs. Model-centric perspective

The results of this study highlight the importance of adopting a data-centric perspective in vegetation classification tasks characterized by limited reference data. Under such conditions, improvements in feature representation have a stronger influence on model performance than increasing model complexity or extensive parameter tuning (Zhu et al., 2018; Fan et al., 2025).

In the experiments conducted, the introduction of spatial texture features led to more stable classification outputs and reduced pixel-level noise, while changes in model configuration played a comparatively smaller role. This suggests that the main limitations of the baseline approach were related to insufficient feature representation rather than to the capacity of the classifier itself (Feng et al., 2015; Gini et al., 2018).

This observation is particularly relevant in very high-resolution UAV imagery, where data volumes are large but labeled samples remain extremely limited. In such scenarios, increasing model capacity without improving the informational content of the input data may lead to unstable predictions or overfitting, especially in heterogeneous environments (Persello and Bruzzone, 2016; Safonova et al., 2023).

At the same time, the results show that high numerical performance does not necessarily translate into spatially consistent outputs at the full-scene scale. As discussed in Section 4.4.4, the gap between test-set performance and full-scene behavior reflects the combined effect of limited labeled data, spatial autocorrelation between neighboring pixels, and the independent nature of pixel-based classification.

In this context, the observed limitations should not be interpreted as a weakness of the model, but as a structural characteristic of the problem. Addressing these challenges requires improving data representation and integrating spatial context, rather than relying solely on more complex models.

5.7 Implications for UAV-based vegetation mapping

The findings of this research highlight several practical implications for UAV-based vegetation mapping, particularly in complex and data-limited environments.

First, feature-space design plays a key role in very high-resolution UAV imagery. While spectral information provides a strong baseline, the integration of simple texture features allows the model to capture local spatial patterns more effectively, improving class separability without increasing model complexity.

Second, traditional accuracy metrics should be complemented with spatial evaluation. Metrics such as Overall Accuracy and Macro F1-score do not fully reflect map quality, as spatial fragmentation and local inconsistencies may still persist. Indicators such as edge density, class adjacency, and landscape diversity provide a more complete assessment of classification outputs.

Third, pixel-based classification remains a scalable and practical solution for large UAV datasets. The ability to process more than 237 million pixels demonstrates its applicability, although post-processing is often required to improve spatial coherence and reduce noise.

Finally, the results support a data-centric perspective, where improving feature representation is more effective than increasing model complexity under limited labeled data conditions.

These observations indicate that effective UAV-based vegetation mapping relies on a balanced integration of spectral features, spatial context, appropriate evaluation, and scalable processing strategies.

6. Conclusions and outlook

6.1 Conclusions and final methodological insight

This research presented a pixel-based machine learning framework for vegetation classification using very high-resolution multispectral UAV imagery under conditions of severe reference-data scarcity. In the primary dataset, only about 0.12% of valid pixels were labeled out of more than 131 million, creating a strong imbalance between data volume and available supervision.

The baseline configuration, based on spectral bands and vegetation indices, provided a solid starting point and achieved satisfactory numerical performance. However, the resulting maps showed visible fragmentation and local inconsistencies, indicating that performance on labeled samples does not fully reflect full-scene behavior.

By adding simple texture features derived from local neighborhoods, the feature space was enriched without increasing model complexity. This led to improved classification results and more spatially coherent outputs, with reduced noise and more stable class boundaries, especially in heterogeneous areas.

The evaluation on an independent dataset confirmed that the framework maintains strong performance under different spatial conditions and larger data volumes. At the same time, similar spatial patterns of fragmentation were observed, showing that some limitations are inherent to pixel-based classification rather than specific to a single dataset.

Overall, the results show that under limited data conditions, improving feature representation and processing design is more effective than increasing model complexity. A data-centric workflow that combines spectral features, texture information, and class imbalance handling provides a practical solution for high-resolution UAV-based vegetation mapping.

6.2 Scientific and practical contributions

Scientific Contributions

- A data-centric, pixel-based framework for vegetation classification in alpine and aquatic environments under severe reference-data scarcity

- Empirical evidence for the role of simple texture features in improving class separability under high spectral similarity

- A controlled experimental design enabling direct comparison between spectral (9-band) and texture-enhanced (27-band) feature spaces

- Support for data-centric approaches over model-centric strategies in data-limited remote sensing problems

- Demonstration of the effectiveness of inverse class weighting in multi-class pixel-based classification

Practical contributions

- A lightweight, transparent, and scalable XGBoost-based classification pipeline

- Demonstration of applicability to large UAV datasets at full-scene scale

- Improved spatial coherence of classification outputs through reduced noise and fragmentation

- A fully reproducible, script-based workflow covering preprocessing, feature engineering, training, and evaluation

6.3 Limitations and future research directions

Despite the strong performance achieved, several limitations should be considered, mainly related to data characteristics and practical constraints.

The most important limitation is the extreme scarcity of labeled data relative to image size. Although the dataset contains more than 131 million valid pixels, only a very small fraction is labeled. Under these conditions, the model can perform well on labeled samples while still showing spatial inconsistencies across the full scene.

A related limitation concerns spatial autocorrelation in the pixel-based data splitting. Even though training and test samples were built as disjoint pixel-coordinate subsets, neighboring

pixels can still end up in both subsets. This may lead to slightly optimistic performance estimates, particularly in homogeneous areas.

The use of single-date imagery is another constraint. Alpine vegetation changes throughout the season, and additional acquisition dates would likely improve class separability by capturing this seasonal variability. A revisit of the study area is already planned for an upcoming season, which would allow for temporal consistency analysis and a re-evaluation of the model under different growth conditions.

The feature space used in this work is also limited in a few ways. Texture description relies only on first-order statistics (mean, standard deviation, range), and no three-dimensional structural information, such as canopy height from DSM, DTM, or nDSM, is currently included. Adding this kind of structural information could help separate species that look similar in the spectral domain but differ in height or canopy structure.

Future work could also focus on improving the reference data through more targeted sampling, especially in transition zones and underrepresented classes. Multi-scale texture features could further strengthen the spatial representation, and combining pixel-based classification with region- or object-based refinement could improve spatial coherence and reduce the reliance on post-processing steps.

Finally, semi-supervised and self-supervised learning approaches represent a promising direction for making better use of the large amount of unlabeled UAV data available, reducing the dependency on manual labeling while improving the model's generalization without requiring more labeled samples.

6.4 Generalization across case studies

The evaluation of the framework across two independent case studies provides insight into its generalization capability.

In the primary study area, improvements were mainly driven by feature-space enrichment, leading to better numerical performance and more stable spatial outputs. The second case study introduced a larger and more complex dataset, with different class distributions and landscape structure. Despite these differences, the framework maintained strong classification performance.

At the same time, similar spatial patterns were observed in both datasets, including fragmentation and local inconsistencies at the full-scene scale. This indicates that these effects are not specific to a single dataset, but are related to the nature of pixel-based classification in very high-resolution imagery.

The comparison shows that the framework is transferable across similar environments in terms of classification performance, while also producing consistent spatial behavior. This consistency suggests that both strengths and limitations of the approach are stable and predictable.

Overall, the results indicate that the proposed method can be applied to comparable alpine environments, while also highlighting the importance of combining feature-space design with spatial refinement to achieve reliable mapping results.

References

Bazzo, C.O.G., Kamali, B., Vianna, M.S., Behrend, D., Hueging, H., Schleip, I., Mosebach, P., Haub, A., Behrendt, A. and Gaiser, T. (2024)
'Integration of UAV-sensed features using machine learning methods to assess species richness in wet grassland ecosystems', *Ecological Informatics*, 83, 102813.

Belcore, E., Pittarello, M., Lingua, A.M. and 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.

Belgiu, M. and Drăguţ, L. (2016)
'Random forest in remote sensing: A review of applications and future directions', *ISPRS Journal of Photogrammetry and Remote Sensing*, 114, pp. 24–31.

Chen, J., Chen, Z., Huang, R., You, H., Han, X., Yue, T. and Zhou, G. (2022)
'The effects of spatial resolution and resampling on the classification accuracy of wetland vegetation species and ground objects based on high spatial resolution UAV images', *Remote Sensing*, 14, 1775.

Fan, J., Yu, X., Ke, L., Li, Q., Li, Z., Zhao, R., Du, X. and Yang, J. (2025)
'Classification of vegetation types in saline–alkali grassland of the Songnen Plain using multispectral UAV imagery and ensemble learning', *Ecological Informatics*, 92, 103509. https://doi.org/10.1016/j.ecoinf.2025.103509

Feng, Q., Liu, J. and Gong, J. (2015)
'UAV remote sensing for urban vegetation mapping using random forest and texture analysis', *Remote Sensing*, 7(1), pp. 1074–1094. https://doi.org/10.3390/rs70101074

Foody, G.M. (2002)
'Status of land cover classification accuracy assessment', *Remote Sensing of Environment*, 80(1), pp. 185–201.

Foody, G.M. (2023)
'Challenges in the real world use of classification accuracy metrics: From recall and precision to the Matthews correlation coefficient', *PLOS ONE*, 18(1), e0291908.

Gini, R., Sona, G., Ronchetti, G., Passoni, D. and Pinto, L. (2018)
'Improving tree species classification using UAS multispectral images and texture measures', *ISPRS International Journal of Geo-Information*, 7(8), 315. https://doi.org/10.3390/ijgi7080315

Gitelson, A.A., Kaufman, Y.J. and Merzlyak, M.N. (1996)
'Use of a green channel in remote sensing of global vegetation from EOS-MODIS', *Remote Sensing of Environment*, 58(3), pp. 289–298.

Gonzalez, R.C. and Woods, R.E. (2008)
Digital Image Processing. 3rd edn. Prentice Hall.

He, H. and Garcia, E.A. (2009)
'Learning from imbalanced data', *IEEE Transactions on Knowledge and Data Engineering*, 21(9), pp. 1263–1284. https://doi.org/10.1109/TKDE.2008.239

Huete, A.R. (1988)
'Soil-adjusted vegetation index (SAVI)', *Remote Sensing of Environment*, 25(3), pp. 295–309.

Huete, A., Didan, K., Miura, T., Rodriguez, E.P., Gao, X. and Ferreira, L.G. (2002)
'Overview of the radiometric and biophysical performance of the MODIS vegetation indices', *Remote Sensing of Environment*, 83(1–2), pp. 195–213.

Jintasuttisak, T., Chabplan, P., Issaro, S., Saeung, O. and Suwanroj, T. (2022)
'Accurate segmentation of vegetation in UAV desert imagery using HSV–GLCM features and SVM classification', *Remote Sensing*, 14(24), 6248.

King, G. and Zeng, L. (2001)
'Logistic regression in rare events data', *Political Analysis*, 9(2), pp. 137–163.

Kramarczyk, P. and Hejmanowska, B. (2025)
'AccuClass: A comprehensive tool for accuracy metrics evaluation in machine learning and remote sensing classification', *SoftwareX*, 21, 102332.

Li, Q., Yuan, H., Fu, T., Yu, Z., Zheng, B. and Chen, S. (2025)
'Multispectral semantic segmentation for UAVs: A benchmark dataset and baseline', *Remote Sensing*, 17(3), 512.

Liu, Y., Zhang, Y., Wang, C. and Chen, X. (2021)
'UAV-based vegetation classification using gradient boosting decision trees', *Remote Sensing*, 13(18), 3654.

Łoś, H., Sousa Mendes, G., Cordeiro, D., Grosso, N., Costa, H., Benevides, P. and Caetano, M. (2021)
'Evaluation of XGBoost and LightGBM performance in tree species classification with Sentinel-2 data', in *Proceedings of the 2021 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*. Brussels, Belgium: IEEE, pp. 4726–4729.

Ma, L., Liu, Y., Zhang, X., Ye, Y., Yin, G. and Johnson, B.A. (2023)
'Deep learning in remote sensing applications: A meta-analysis and review', *ISPRS Journal of Photogrammetry and Remote Sensing*, 196, pp. 277–294.

Matrone, F., Graziani, C., Spadaro, A., Hemmatianzadeh, S. and 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', *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLVIII-2/W11-2025, pp. 203–210. https://doi.org/10.5194/isprs-archives-XLVIII-2-W11-2025-203-2025

Maxwell, A.E., Warner, T.A. and Fang, F. (2018)
'Implementation of machine-learning classification in remote sensing: An applied review', *International Journal of Remote Sensing*, 39(9), pp. 2784–2817.

Mountrakis, G., Im, J. and Ogole, C. (2011)
'Support vector machines in remote sensing: A review', *ISPRS Journal of Photogrammetry and Remote Sensing*, 66(3), pp. 247–259.

Pan, W., Wang, X., Sun, Y., Wang, J., Li, Y. and Li, S. (2022)
'Karst vegetation coverage detection using UAV multispectral vegetation indices and machine learning algorithms', *Remote Sensing*, 14, 331.

Persello, C. and Bruzzone, L. (2016)
'Very high resolution remote sensing image classification with scarce training data', *IEEE Transactions on Geoscience and Remote Sensing*, 54(2), pp. 871–885.

Ploton, P., Mortier, F., Réjou-Méchain, M., Barbier, N., Picard, N., Rossi, V. and Pélissier, R. (2020)
'Spatial validation reveals poor predictive performance of large-scale ecological mapping models', *Nature Communications*, 11, 4540.

Rodriguez-Galiano, V.F., Ghimire, B., Rogan, J., Chica-Olmo, M. and Rigol-Sanchez, J.P. (2012)
'An assessment of the effectiveness of a random forest classifier for land-cover classification', *ISPRS Journal of Photogrammetry and Remote Sensing*, 67, pp. 93–104.

Roberts, D.R., Bahn, V., Ciuti, S., Boyce, M.S., Elith, J., Guillera-Arroita, G., Hauenstein, S., Lahoz-Monfort, J.J., Schröder, B., Thuiller, W., Warton, D.I., Wintle, B.A., Hartig, F. and Dormann, C.F. (2017)
'Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure', *Ecography*, 40(8), pp. 913–929.

Safonova, A., Ghazaryan, G., Stiller, S., Main-Knorn, M., Nendel, C. and Ryo, M. (2023)
'Ten deep learning techniques to address small data problems with remote sensing', *International Journal of Applied Earth Observation and Geoinformation*, 125, 103606. https://doi.org/10.1016/j.jag.2023.103606

Samat, A., Li, E., Wang, W., Liu, S., Lin, C. and Abuduwaili, J. (2020)
'Meta-XGBoost for hyperspectral image classification using extended MSER-guided morphological profiles', *Remote Sensing*, 12(12), 1973.

Soille, P. (2003)
Morphological Image Analysis: Principles and Applications. 2nd edn. Springer.

Walz, J.H. and Weber, K.T. (2009)
Identifying actively growing vegetation using NDVI threshold values. Pocatello, ID: GIS Training and Research Center, Idaho State University.

Wang, X., Pan, W., Sun, Y., Li, Y. and Li, S. (2022)
'Handling class imbalance in UAV-based vegetation classification using ensemble learning', *Remote Sensing*, 14, 331.

Zhang, L., Zhang, J., Liu, X. and Zhang, Y. (2022)
'Feature-driven versus model-driven learning strategies for high-resolution UAV vegetation mapping', *Remote Sensing*, 14(21), 5478.

Zhou, Z., Zheng, C., Liu, X., Tian, Y., Chen, X., Chen, X. and Dong, Z. (2023)
'A dynamic effective class-balanced approach for remote sensing imagery semantic segmentation of imbalanced data', *Remote Sensing*, 15(7), 1768.

Zhu, X.X., Tuia, D., Mou, L., Xia, G.-S., Zhang, L., Xu, F. and Fraundorfer, F. (2018)
'Deep learning in remote sensing: A comprehensive review and list of resources', *IEEE Geoscience and Remote Sensing Magazine*, 5(4), pp. 8–36.

Zou, M., Liu, Y., Fu, M., Li, C., Zhou, Z., Meng, H., Xing, E. and Ren, Y. (2021)
'Combining spectral and texture features of UAV imagery with plant height to improve leaf area index estimation of winter wheat at jointing stage', *Remote Sensing*, 13(15), 3001.