High-resolution land cover mapping with GeoAI: instance segmentation for land cover analysis

Federica Gerla[1,2], Lindo Nepi[3], Mihnea Căţeanu[4,5], Francesca Ferroni[6,7], Khew Ee Hung[8], Krystian Mocny[9], Shahla Yadollahi[10], Caterina Balletti[1], Roberto Pierdicca[11], Mattia Balestra[11,12]

[1] Università Iuav di Venezia, CIRCE Lab, Venice, Italy - fgerla@iuav.it, balletti@iuav.it
[2] Sapienza Università di Roma, DICEA, Rome, Italy
[3] Università Politecnica delle Marche, DII, Ancona, Italy - l.nepi@staff.univpm.it
[4] Transilvania University of Brașov, Faculty of Silviculture and Forest Engineering, Brașov, Romania - cateanu.mihnea@unitbv.ro
[5] Technical University of Civil Engineering, Faculty of Geodesy, Bucharest, Romania
[6] Sapienza Università di Roma, Department of Environmental Biology, Rome, Italy
[7] FAO, Mountain Partnership Secretariat, Rome, Italy - francesca.ferroni@fao.org
[8] Hokkaido University, School of Agriculture, Japan - eehung.khew.f9@elms.hokudai.ac.jp
[9] Institute of Soil Science and Plant Cultivation, Department of Geomatics, Puławy, Poland - kmocny@iung.pulawy.pl
[10] Liège University, Environment Campus, Arlon, Belgium
[11] Università Politecnica delle Marche, DICEA, Ancona, Italy - r.pierdicca@staff.univpm.it, m.balestra@staff.univpm.it
[12] Università Politecnica delle Marche, D3A, Ancona, Italy

Keywords: GeoAI, Instance segmentation, Land-cover mapping, High-resolution Earth Observation

Abstract

Land cover classification has become a key method for understanding natural and ecological resources, as well as for sustainable land-use planning and management. Advances in deep learning (DL) and automated satellite image analysis have significantly improved the accuracy, speed and scalability of land cover map production. This paper investigates the potential of instance-based segmentation within a GeoAI workflow for high-resolution land cover classification in the area of San Vito di Cadore (Veneto), a UNESCO mountain region with high ecological heterogeneity. The dataset comprises 650 manually annotated orthophotos at a spatial resolution of 0.1-0.5 m, labelled using seven main land cover classes (Forest, Shrubland, Grassland, Cropland, Water Bodies, Artificial/Urban Areas, and Rocky/Bare Areas), subsequently harmonized with the CORINE Land Cover (CLC) categories for inter-comparison purposes. Snow and cloud were treated as auxiliary classes due to their frequent occurrence in alpine orthophotos. The YOLOv11 instance-segmentation model was trained on 1000 × 1000 px tiles. During inference, a sliced-inference approach was adopted exploiting the SAHI (Slicing Aided Hyper Inference) framework, enabling the processing of large-scale orthophotos without degrading spatial quality. The results show an overall precision of 0.847 and an overall recall of 0.575, with mAP@0.5 greater than 0.65. Quantitative comparison with the Regione Veneto land cover product (2023) shows good agreement for the dominant forest class (-1.4%), whereas the largest discrepancies occur for artificial surfaces (+20.8%) and agricultural areas (-36.3%), attributable to differences in spatial scale and training-sample imbalance.

1. Introduction

Land cover classification has become a cornerstone of environmental and territorial sciences. It is not merely a technical exercise in remote sensing data analysis, but rather an analytical tool supporting land use planning and natural resources management (Gaur and Singh, 2023; Klopfer and Gruehn, 2024). Information derived from these analyses enables the assessment of landscape dynamics, the rate of urbanization, and the loss of natural surfaces, providing key indicators for the definition of sustainable development strategies and territorial adaptation policies (Keshtkar et al., 2017; Bokhari et al., 2022; Mugiraneza et al., 2022). In the context of spatial planning, land cover data provide an essential basis for policy formulation and the development of urban planning instruments and zoning schemes, supporting the management of infrastructure, the mitigation of natural hazards, and the assessment of environmental (Chamling and Bera, 2020; Alvarez Gebelin et al., 2024). Land cover is defined as the observed physical cover on the Earth's surface, including vegetation (natural or planted) and human constructions (FAO, 2024). Within the United Nations (UN) framework, it is considered a fundamental data theme, as it influences sustainable development, climate change, biodiversity conservation, food security, and disaster risk reduction. The UN recognises land cover as one of the 14 fundamental data theme layers, highlighting its global relevance for environmental monitoring and sustainable land management. Land cover maps are further applied in the management of agricultural, forest, and water resources, facilitating crop monitoring, the assessment of hydrogeological vulnerability, and the conservation of biodiversity (Rana and Suryanarayana, 2020; Yang et al., 2025; Zhang et al., 2024).

The integration of Computer Vision (CV) into Geospatial Artificial Intelligence (GeoAI) has fundamentally shifted how we interpret Earth observation data. By applying deep learning architectures to satellite and aerial imagery, researchers can now automate the extraction of complex spatial patterns at scales previously unimaginable. Recent advances in deep learning (DL) and automatic satellite image analysis have been transforming the methodologies for the production and updating of land cover maps. In particular, Convolutional Neural Networks (CNNs) have established themselves as the standard for extracting high-resolution information, improving both the accuracy and the robustness compared to traditional machine learning methods (Carranza-Garcia et al., 2019; Fayaz et al., 2024; Eb-

This contribution has been peer-reviewed.
https://doi.org/10.5194/isprs-archives-XLIX-B4-2026-133-2026

rahimi and Kumar, 2025).

In recent years, CNNs have been joined by models based on the Vision Transformer (ViT) architecture, a technology that applies the Transformer approach, the de facto standard for natural language processing tasks, to computer vision. However, the use of pre-trained CNNs allows robust results to be obtained even with limited data and in complex territorial contexts. This makes them preferable in some cases to deep learning solutions based on ViT architectures, which in many cases deliver superior results compared to CNN models (Dosovitskiy et al., 2020; Pierdicca et al., 2023) but require a much larger amount of annotated data.

In the landscape of computer vision applications in the GeoAI field, image segmentation plays a very important role. This type of approach allows us not only to identify and localize an object, but also to identify its contours. By isolating the specific pixels belonging to an object, GeoAI segmentation models provide the high-fidelity measurements needed for rigorous scientific reporting. Within the GeoAI domain, the distinction between semantic and instance segmentation is pivotal for determining the depth of data analysis. Semantic segmentation assigns a thematic class label to every pixel in an image. In a GeoAI context, it treats all pixels of the same category (e.g., "forest" or "water") as a single, continuous mask without distinguishing between individual objects. The instance segmentation approach extends semantic segmentation by detecting and delineating every individual object of interest. It provides a unique identity to each instance, allowing the system to differentiate between "grassland A" and "grassland B", even if they are touching. The resulting object-level representations can be integrated into GIS workflows for distinct polygons, facilitating both expert validation and operational use for land-use and territorial planning (Xu et al., 2021; Li et al., 2022, 2024).

The present work, carried out within the Earth Sensing Summer School 2025[1], centred on remote sensing and GeoAI for environmental monitoring, investigates the potential of the YOLOv11 instance segmentation model, deployed within a SAHI-based inference framework, for high-resolution land cover mapping using manually annotated orthophotos. Compared to traditional pixel-wise semantic segmentation approaches, instance-based segmentation is expected to improve boundary delineation accuracy and spatial consistency of land cover units. A particular focus is placed on the role of annotation quality, class balance, and dataset composition in driving model performance, using a mountainous alpine environment (San Vito di Cadore, Veneto, Italy) as a case study.

2. Study area framework

The study was conducted in the municipality of San Vito di Cadore, a mountainous area located in the Province of Belluno, Veneto region, in the northeastern part of Italy (Figure 1). The area lies in the Dolomites, a UNESCO World Heritage Site characterized by heterogeneous land cover, predominantly steep topography, and a mixture of forest, grassland, rocky areas, and urbanized surfaces. The landscape is subject to ongoing ecological succession processes, with the expansion of forest cover at the expense of former pastures and agricultural land, particularly in lower areas with elevation under 1500 m (Giupponi et al., 2006; Dalla Valle et al., 2009). The resulting landscape mosaic combines strong spatial heterogeneity, ecotonal transitions between adjacent cover types, fragmented patches, and significant spectral overlap between spectrally similar classes such as shrubland and grassland or forest and shrubland. In mountainous environments of this type, rapid changes in slope, aspect, and terrain exposure create highly variable illumination conditions that amplify within-class spectral variability and reduce the separability of adjacent land cover classes, posing well-known challenges for automated classification models (Qichi et al., 2023). These characteristics make the area particularly suitable for testing land cover classification workflows based on high-resolution imagery.

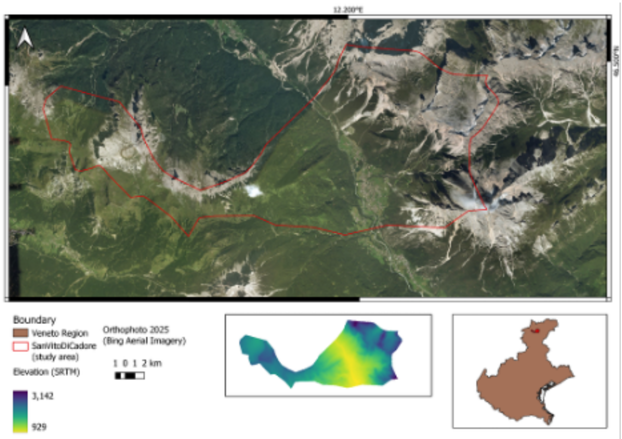

Figure 1. Study area of San Vito di Cadore (Dolomites, Italy) displayed on Bing Aerial Imagery (2025), with elevation data (SRTM) and study area boundary shown in red.

3. Materials and Methods

High-resolution orthophotos were obtained from multiple open-access sources spanning different countries, spatial resolutions, and acquisition conditions. This multi-source strategy was deliberately adopted to maximise the spectral and radiometric variability of the training dataset, thereby improving the model's ability to generalise across diverse imaging conditions. The imagery used was collected in various parts of Italy and Switzerland, and included data from the following sources:

- Italian National Geoportal (0.3–0.5 m resolution)
- Veneto Region Geoportal (0.3–0.5 m resolution)
- SWISS orthophotos (0.1 m resolution)
- Google Satellite imagery (0.1 m resolution)

Image selection aimed at maximizing the range of land cover conditions, capturing diverse spatial contexts while focusing mainly on the land cover categories identified as predominant in the study area of San Vito di Cadore. It was chosen not to resample the images to a common resolution, so the model worked with the native resolutions characteristic of each data source used.

All images collected from the data sources indicated above were divided into 1000×1000 pixel tiles before starting the annotation phase. This allowed the annotation experts to work more easily. At the same time, using images with smaller dimensions than those typically found in orthophotos made it possible to train the model without the need for enterprise-grade hardware.

[1] Link to the summer school official webpage

https://doi.org/10.5194/isprs-archives-XLIX-B4-2026-133-2026

By applying the tiling operation, 650 images were obtained: 81 percent (526 images) were used as training data, while 124 images (the remaining 19 percent) were reserved for model validation.

Based on the visual analysis of the San Vito di Cadore landscape and available reference datasets (Veneto regional land cover product, based on CLC updated to 2023), seven main land cover classes were defined for the thematic analysis:

- Forest
- Shrubland
- Grassland
- Cropland
- Water Bodies
- Artificial/Urban Areas
- Rocky/Bare Areas

Initially, a more detailed ten-class scheme was adopted, including grassland, forest, shrubland, cropland, rivers, lakes, road, urban area, rocky area, and snow. This scheme was subsequently harmonized in order to reduce class imbalance and improve the semantic consistency of the annotations. In particular, rivers and lakes were merged into the single class "Water Bodies", while road and urban area were aggregated into "Artificial/Urban Areas". In addition, snow was retained as an explicit class because of its frequent occurrence in high-resolution orthophotos of alpine environments, particularly at higher elevations. Cloud was also introduced as an additional class for the same reason, as both elements may interfere with the interpretation and segmentation of the target land cover classes. As a result, the final annotation and model-training workflow included nine foreground classes overall, consisting of seven main land cover classes for thematic analysis plus the additional "Snow" and "Cloud" classes. The process of instance segmentation involves two main steps: obtaining a dataset of labelled data for training and validating the deep learning model, and applying (testing) the model to a dataset from a different geographic area (in this case, an orthophoto of San Vito di Cadore) in order to obtain an estimate of land cover in the study area.

Labelling of training images was done manually using the Roboflow platform (`https://app.roboflow.com`). The main land cover classes present in each training image were delineated using either manually drawn polygons (Figure 2A) or polygons generated by the Roboflow platform using the Smart Polygon tool (Figure 2B). This automatic segmentation tool attempts to accurately delineate sections in the input image based on spectral similarity. In both cases, all polygon geometries were visually inspected and manually corrected by the annotators to ensure accurate delineation of the land cover boundaries. A total of 6624 polygons were manually labelled in the training dataset across the nine foreground classes included in the annotation and model-training workflow. Of these, 6382 polygons belong to the seven main land cover classes used for the thematic analysis, while the remaining polygons correspond to the additional "Snow" and "Cloud" classes. To ensure high quality and reliability of the annotations, a rigorous multi-stage quality control process was established. In the first stage, four annotators were randomly paired, and each annotator was tasked with independently re-annotating the images assigned to the other. In the second stage, two annotators conducted a systematic review of the annotations through visual inspection. This step focused on identifying and correcting mislabelled, incomplete, or improperly delineated objects, ensuring that annotations met predefined standards and guidelines. In the third stage, the refined annotations were further examined to ensure overall dataset consistency. This multi-stage quality assurance process is critical in achieving a high-quality dataset suitable for robust downstream applications (Balestra et al., 2025).

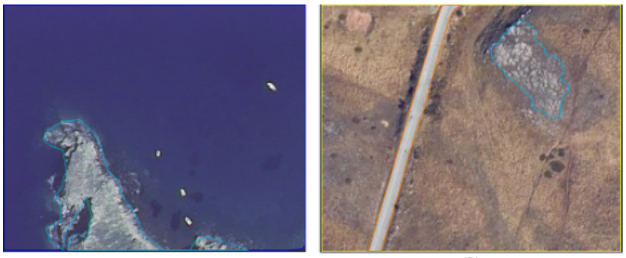

Figure 2. A: a training image showing a lake surface and a rocky area, with the latter delineated by hand; B: a training image showing mostly shrubland, with the road surface and rocky area delineated automatically using the Smart Polygon tool.

As images used for training and validation were chosen based on their similarity to the study area of San Vito di Cadore, classes specific to mountainous regions (such as forest, grassland, or shrubland) have more polygons associated with them compared to classes such as cropland and urban areas. Table 1 presents the distribution of all nine foreground classes considered in the annotation and model-training workflow, including the seven main land cover classes used for thematic analysis and the additional "Snow" and "Cloud" classes. Meanwhile, 124 validation images were annotated adopting the same methodology. As these images were randomly selected by the Roboflow platform, the distribution of polygons across land cover classes follows a pattern similar to that of the training images (Table 1). Annotators ensured that training polygons were evenly distributed across the spatial extent of each image and were representative of intra-class variability (e.g., dense vs. sparser forest, grassland with varying vegetation density and cover). Land cover classes were identified and delineated based on standard visual photo-interpretation criteria, primarily texture and spatial context, as commonly applied in remote sensing image analysis (Lillesand et al., 2015; Janga et al., 2023).

Table 1. Distribution of polygons before augmentation across the nine foreground classes included in the annotation and model-training workflow.

Classes	Train	Val.
Forest	1744	492
Shrubland	1217	384
Grassland	1474	325
Cropland	240	88
Water Bodies	238	73
Artificial/Urban	586	123
Rocky/Bare Areas	883	161
Snow	211	30
Cloud	31	8

After the initial labelling, a single annotator, to ensure consistency, undertook the quality control of labelling through visual verification and correction of mislabelled, incomplete, or improperly delineated objects. Once a set of labelled, controlled images is obtained, the next step is the augmentation of labelled images. The aim of augmentation is to generate additional training samples by applying automatic modifications to the original images using various techniques such as rotation, scaling, grayscaling, and changes in brightness and contrast (Casado-García et al., 2019; Hwang et al., 2025). Augmentation comes with multiple benefits: (1) it increases the diversity of training data, simulating real-world variations and thus helping the model learn to recognize land covers under different conditions (Hwang et al., 2025); (2) it improves generalization since it exposes the model to a wider range of scenarios, reducing overfitting and thus enhancing the model's ability to perform well on unseen data (Kupyn and Rupprecht, 2024); (3) it can address class imbalance, potentially improving performance on underrepresented classes (Liu and Xiao, 2023); and (4) it reduces annotation costs, as instance-level augmentations can expand datasets without the need for additional manual labeling (Ghiasi et al., 2021).

In this study, 9 additional images were generated in Roboflow for each manually labelled image, using the following augmentation techniques:

- Horizontal and vertical flipping
- Rotation of images: up to ±10 degrees
- Grayscale: applied randomly to 10 percent of images
- Saturation: up to ±7 percent
- Exposure: up to ±5 percent
- Blur: up to 0.2 px

Data augmentation was applied to both the training dataset and the validation dataset. At the end of the augmentation process, over 6,500 training images and approximately 1,100 validation images, with corresponding annotations, were obtained. It is important to note that data augmentation is applied not only to the images but also to the coordinates of the polygons included in the annotations for each image.

Finally, the dataset comprising the manually labelled and augmented images is exported from the Roboflow platform and imported into Ultralytics YOLOv11 for processing. As input to the model, each image is therefore associated with its corresponding annotation file, which constitutes the ground truth necessary for proper training and validation of the model.

The model is based on the Ultralytics YOLOv11 framework, an instance segmentation architecture optimized for object detection, instance segmentation, and other computer vision tasks. The model was trained by applying fine-tuning to the pretrained YOLOv11m-seg model, with GPU acceleration and the following hyperparameters (Table 2): LR = 0.007, WD = 0.0005, dropout = 0.3, optimizer = AdamW. We also set imgsz = 1000 and multi_scale = 0.25.

To preserve spatial detail, training was performed on 1000×1000 pixel tiles rather than on full-size images, avoiding the downsampling that would cause the loss of fine-scale objects. The choice of the Ultralytics YOLO framework in the present work was motivated by the need to process large-scale orthophotos at inference time.

Table 2. Final parameter values used for the instance segmentation model based on the YOLOv11 framework.

Parameter	Value
Epochs	350
Batch size	16
Optimizer	AdamW
Weight decay	0.0005
Dropout	0.3
LR	0.007
YOLO model size	Medium

YOLOv11 is part of a family of CNN-based models released by Ultralytics (`https://www.ultralytics.com`). Each new release introduces innovations that allow YOLO to achieve performance at the same level as ViT architectures in computer vision tasks, while requiring fewer training images. Each YOLO release is available in various configurations that provide different model sizes, capable of adapting to datasets of different sizes and avoiding the risk of overfitting or underfitting that can compromise the model's performance. However, the fine-tuning operation (optimal configuration of parameters, depending on the dataset used) requires several attempts to identify the right configuration, which, in our case too, required several training runs. Another feature of recent versions of YOLO is the ability to integrate easily with SAHI software, which allows inference on large-scale images using a sliced-inference approach. In the context of Geospatial Artificial Intelligence (GeoAI), sliced inference is a technique designed to overcome the hardware and algorithmic limitations of processing high-resolution imagery. It involves dividing a large-scale input image into smaller, overlapping patches (slices), performing inference on each patch independently, and then reassembling the results into a single global output. This approach, applied during the inference phase, allows models that natively support very small image sizes (640×640 or similar) to perform inference on large images without losing details and while avoiding system overload.

In this type of task, as well as more generally in GeoAI projects, sliced inference can be implemented through the Slicing Aided Hyper Inference (SAHI) functionality provided by the Ultralytics YOLO framework. The use of SAHI (Akyon et al., 2022) enabled inference on large-scale orthophotos (e.g., 14000×11000 pixels) in a straightforward manner by exploiting the built-in functions of the library, without the need to implement external scripts for image-slicing-based inference as is required by other models. This strategy reduces training time while maintaining detection and segmentation performance during inference by avoiding image resampling. The final model was used to create a land cover map of the study area using a 0.3-meter-resolution orthophoto of San Vito di Cadore as input (Figure 3). The predicted land cover classes were compared with the Veneto regional land cover product, based on CLC and updated to 2023. The comparison was carried out in two steps. First, the seven GeoAI land cover classes were harmonized with the CLC Level 1 categories through a cross-reference table, enabling a class-level alignment between the two datasets. Second, the mapped areas per class were extracted using zonal statistics and compared quantitatively, after clipping both datasets to the same spatial extent.

Figure 3. Orthophoto of the study area used for testing.

4. Results and discussion

4.1 The effect of dataset quality

The results obtained confirm that the YOLOv11 model is sensitive to class imbalance and annotation consistency. The first training run, carried out using 300 training images and 50 validation images, after 50 training epochs, achieved an average precision (mAP50) of less than 0.2. This may be attributed to the inconsistency of the initial labelling process and to class imbalance across several categories in the preliminary annotation scheme, particularly rivers, lakes, and snow, although a direct statistical validation of these factors was not carried out. To resolve these issues, a single annotator performed quality control on the labelled data, correcting misclassified or incomplete polygons prior to augmentation. Following this labelling cleanup, the dataset was expanded, also using data augmentation, to include 1,200 training images and 150 validation images. The second training session, after 50 training epochs, achieved an mAP50 of 0.34, whilst a third training session with 1,500 training images and 150 validation images resulted in a modest decrease in the mAP50 value. This reduction highlights the intrinsic complexity of class balance and the model’s sensitivity to high variability within classes, particularly in the grassland category. This suggests the importance of extending the training period beyond 300 epochs, according to YOLO documentation, to stabilise learning in cases of high intra-class variability, particularly in grasslands. This variability is evident in the samples used for training (Figure 4), where the grassland class encompasses a wide range of vegetation types and surface conditions, ranging from dense, well-structured grasslands with high tree cover to xeric or rocky formations characterised by sparse and patchy herbaceous cover. This ecological and structural heterogeneity results in significant spectral and structural variation within the class.

In the final training run, to which the metrics presented in this chapter refer, in addition to applying greater data augmentation and extending the number of training epochs, all annotated images were rechecked, excluding those that could pose difficulties in interpreting the object’s class or outlines. This resulted in significantly superior performance and confirms the need for thorough image quality control in the dataset before proceeding with model training. To further evaluate the model’s performance, a normalised confusion matrix was generated (Figure 5). The matrix reveals that forest (0.78), snow (0.69), urban

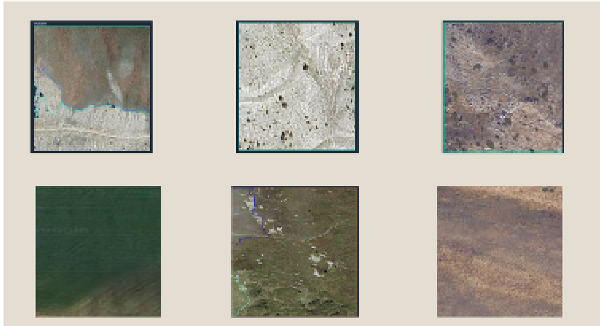

Figure 4. Examples of training samples illustrating the high within-class spectral variance and textural heterogeneity in the grassland category.

area (0.67), and cropland (0.66) are the best-classified categories, reflecting their relatively distinct spectral and structural signatures. The “cloud” class shows comparatively limited separability, with a low diagonal value and marked confusion with the background, reflecting its variable appearance in alpine orthophotos. Rocky area (0.46) and shrubland (0.54) show the lowest diagonal values, with both classes exhibiting substantial confusion with the background, suggesting a systematic tendency towards false negatives for spatially fragmented or spectrally ambiguous cover types. Grassland (0.62) also presents notable misclassification towards the background class (0.18), likely reflecting its high intra-class variability. The consistently high background row values across all classes confirm the model’s general tendency to under-detect instances, consistent with the relatively low overall recall value reported above.

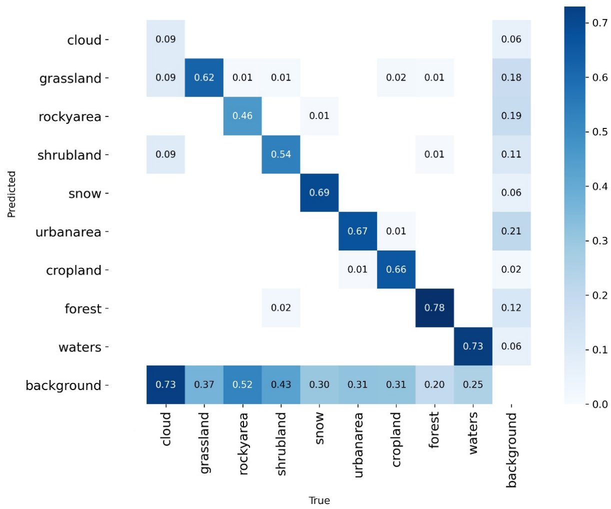

Figure 5. Normalised confusion matrix.

The test results and analysis of the graphs and the image relating to the model’s training and validation (Figures 6 and 7) show that the system achieved an overall precision of 0.847 and an overall recall of 0.575, with an mAP50 greater than 0.65. The relatively low recall value highlights the model’s tendency to produce false negatives, particularly for the “shrubland” and “rocky area” classes. Qualitative inspection of the model outputs on the test orthophoto confirmed this issue, which manifests predominantly in shadowed areas, at ecotonal transitions between adjacent classes, and in areas where spectrally similar textures belonging to different land cover types reduce the model’s ability to produce confident predictions. This problem is likely linked to the low number of samples in certain classes and to the difficulty, even for domain experts, in carrying out photointerpretation of certain images, where lighting conditions and the presence of shadows contribute to making the analysis uncertain. The high segmentation loss value is linked to the difficulty, even for experienced operators, in accurately delineating the boundaries between the different land-cover classes analysed in this study. This will cause the model, during inference, to inaccurately delineate the boundaries between areas with different land-cover classes.

4.2 Thematic consistency with the Veneto Region Cover Product (CLC 2023)

The final model was applied to the study area orthophoto to generate a land cover map, which was subsequently evaluated for spatial and thematic consistency with the Veneto regional land cover product based on CLC updated to 2023. It should be noted that the two datasets differ substantially in terms of spatial scale, minimum mapping unit, cartographic logic, and class aggregation level. The CLC-based product should therefore not be treated as high-resolution ground truth, but rather as a reference for coarse-scale thematic consistency assessment. The classification obtained using the YOLOv11 model (Figure 8) yielded a total mapped area of 6.56 km^2. The dominant class is “forest”, covering 4.71 km^2, equal to 71.8% of the total mapped area, followed by “urban area” with 1.22 km^2 (18.6%) and “grassland” with 0.59 km^2 (9.0%). The remaining classes occupy marginal areas: “water bodies” (0.017 km^2, 0.3%), “shrubland” (0.013 km^2, 0.2%), “rocky area” (0.004 km^2, 0.05%), and “cropland” (0.002 km^2, 0.02%).

The Land Cover layer for the Veneto region (Figure 9), based on CLC and updated to 2023, cropped to the study area, shows a total area of 6.80 km^2. The dominant class is Class 3 (Wooded areas and semi-natural environments) with 4.80 km^2 (70.6% of the total), followed by Class 1 (Artificial surfaces) with 1.01 km^2 (14.9%) and Class 2 (Agricultural surfaces) with 0.93 km^2 (13.6%). Class 5 (Water bodies) represents the smallest component with 0.064 km^2 (0.9%). To enable a quantitative comparison between the two datasets, the seven GeoAI classes have been aligned with the Level 1 classification of the Veneto Land Cover through a cross-reference table (Table 3). The “grassland” class was assigned to Class 2 (Agricultural land), consistent with its spatial correspondence to subclass 2.3.1 (Permanent grassland/pasture). “Forest”, “shrubland”, and “rocky area” were grouped into Class 3 (Wooded areas and semi-natural environments), whilst “Urban area” was mapped to Class 1 (Artificial surfaces) and “Water bodies” to Class 5 (Water bodies). Although the “cropland” class is assigned to Class 2, it accounts for a negligible proportion of the total area (0.002 km^2). The two analyses show good agreement for the dominant land cover class, wooded and semi-natural areas (-1.4%). The most marked discrepancies concern artificial surfaces, which are overestimated by the GeoAI model by 0.21 km^2 (+20.8%), and agricultural surfaces, which are underestimated by 0.34 km^2 (-36.3%). Water bodies show the greatest relative difference (-73.4%), which, however, has limited absolute significance given the very small extent of this class within the study area (0.064 km^2 in the CLC) (Table 4). The overestimation of artificial land cover is likely due to the GeoAI model’s ability to detect, at a fine scale, features such as roads, car parks, and isolated buildings which, falling below the CLC’s minimum mappable unit (25 ha), are not represented in the reference layer. The underestimation of agricultural areas, on the other hand, can be attributed to confusion between the “grassland” class and the “shrubland” class at the boundaries of ecotones. The substantial underestimation of water bodies (-73.4%) may reflect both the limited spatial extent of this class within the study area and the model’s difficulty in detecting small or narrow water features, such as streams and irrigation channels, which are below the detection threshold at the scale of the training dataset. Given the very small absolute area involved (0.064 km^2 in the CLC), however, this discrepancy should be interpreted with caution.

Table 3. Correspondence between GeoAI classes and CORINE Level 1 categories used for quantitative comparison.

CLC class	Description	GeoAI classes
1	Artificial surfaces	Urban Area
2	Agricultural surfaces	Cropland, Grassland
3	Forested and semi-natural areas	Forest, Shrubland, Rocky area
5	Water bodies	Water bodies

5. Conclusions and recommendations

This study has investigated the potential of instance-aware deep learning methods for detailed mapping of land cover in complex mountainous environments, using open-access high-resolution orthophotos and the YOLOv11 framework for instance segmentation.

The results show promise and suggest that instance-aware approaches may improve the capture of fine-scale landscape patterns and ecologically relevant transitions, overcoming the limitations of traditional semantic segmentation in terms of spatial consistency and boundary quality. The model achieved an overall precision of 0.847 and an overall recall of 0.575, with an mAP50 exceeding 0.65, confirming the system’s ability to correctly classify dominant classes, particularly forested areas, which also show the best agreement in the coarse-scale consistency check with the Veneto Land Cover data (-1.4%).

The recall achieved highlights the presence of false negatives (i.e., objects not detected by the model). This issue is primarily related to the high intra-class variability, particularly for classes such as shrubland, cropland, and grassland, which the model struggled to capture effectively due to the limited number of training samples. Additional factors affecting recall may include the presence of small-sized objects (e.g., low-order streams or low-density urban areas typical of mountainous regions) or partially occluded features (e.g., clouds, shadows, or shrubs along riverbanks), which are commonly observed in orthophotos. Improvements in recall performance could likely be achieved by incorporating very high-resolution imagery and increasing the size of the training dataset.

The main challenges identified relate to the imbalance between classes and high intra-class variability, particularly for ecologically heterogeneous categories. The iterative process of label cleaning and dataset rebalancing proved crucial for improving the model’s performance. The study confirms that achieving high levels of accuracy requires a substantial investment by domain experts in the collection, classification, and quality annotation of images. Several methodological limitations should nonetheless be acknowledged. The training dataset remains relatively small and unevenly distributed across classes, which

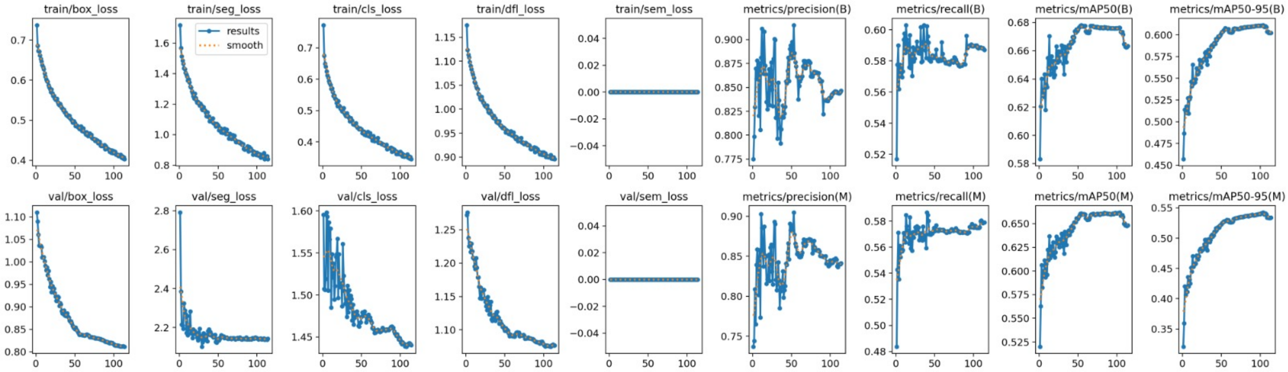

Figure 6. Model performance metrics during training: Precision, Recall, mAP50 and mAP50-95.

Table 4. Comparison of areas between test results and the 2023 Veneto Land Cover Map by Level 1 class.

CLC class	Description	CLC area (km^2)	GeoAI area (km^2)	Difference (%)
1	Artificial surfaces	1.01	1.22	+20.8%
2	Agricultural surfaces	0.93	0.59	-36.3%
3	Forested and semi-natural areas	4.80	4.73	-1.4%
5	Water bodies	0.064	0.017	-73.4%
Total		6.8	6.56	-3.6%

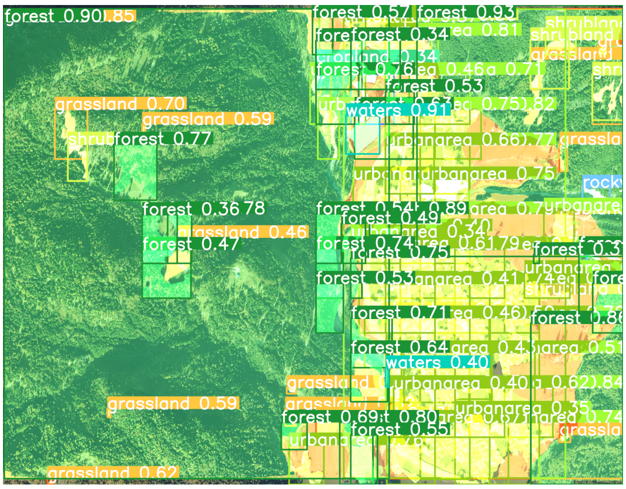

Figure 7. Sample output of the YOLOv11 instance segmentation model over the study area, with predicted land-cover classes and confidence scores.

Figure 8. Classification of the test area obtained using the YOLOv11 model.

Figure 9. Classification of the test area based on land cover in the Veneto Region using CLC 2023.

likely contributes to the low recall values observed for under-represented categories. The instance segmentation approach introduces overlapping polygons during inference, which complicates area estimation and limits direct comparability with polygon-based reference datasets. In this study, overlapping polygons were resolved through a dissolve-by-class geoprocessing operation prior to area calculation, enabling a more reliable quantitative comparison with the CLC reference product. The consistency assessment against the CLC-based Veneto Land Cover product was carried out only at an aggregated thematic scale, given the substantial differences in minimum mapping unit, spatial resolution, and cartographic generalisation between the two products. This comparison should therefore not be interpreted as a rigorous accuracy assessment. Despite these limitations, aligning the classification scheme with established standards such as CLC, combined with the integration of multi-source data, remains a priority for future work, which should also focus on increasing the training samples for under-represented classes such as cropland and rocky area, and resolving the issue of overlapping polygons typical of instance segmentation. Future work should explore the application of alternative instance segmentation models, in particular the Roboflow RF-DETR framework, which has shown promising results in the preliminary tests conducted in this study. Furthermore, the training dataset should be expanded with a larger volume of high-resolution images, in order to improve the model's discriminative power and significantly reduce false negatives during large-scale inference.

Acknowledgements

The authors wish to acknowledge the Earth Sensing Summer School 2025, centred on remote sensing and GeoAI for environmental monitoring, within which this work was carried out. The Summer School provided a valuable framework for interdisciplinary collaboration across geomatics, artificial intelligence, environmental sciences, and spatial planning.

References

Akyon, F. C., Altinuc, S. O., Temizel, A., 2022. Slicing aided hyper inference and fine-tuning for small object detection. *2022 IEEE international conference on image processing (ICIP)*, IEEE, 966–970.

Alvarez Gebelin, A., Borretti, M., Cohn, C., Minutti, G., 2024. Land Cover and Land Use in Uruguay Using Land Cover Classification System Methodology. *Land*, 13(12), 2168.

Balestra, M., Paolanti, M., Pierdicca, R., 2025. Whu-rs19 abzsl: An attribute-based dataset for remote sensing image understanding. *Remote Sensing*, 17(14), 2384.

Bokhari, S. A., Saqib, Z., Amir, S., Naseer, S., Shafiq, M., Ali, A., Zaman-ul Haq, M., Irshad, A., Hamam, H., 2022. Assessing land cover transformation for urban environmental sustainability through satellite sensing. *Sustainability*, 14(5), 2810.

Carranza-Garcia, M., García-Gutiérrez, J., Riquelme, J. C., 2019. A framework for evaluating land use and land cover classification using convolutional neural networks. *Remote Sensing*, 11(3), 274.

Casado-García, Á., Domínguez, C., García-Domínguez, M., Heras, J., Inés, A., Mata, E., Pascual, V., 2019. CLoDSA: a tool for augmentation in classification, localization, detection, semantic segmentation and instance segmentation tasks. *BMC bioinformatics*, 20(1), 323.

Chamling, M., Bera, B., 2020. Spatio-temporal Patterns of Land Use/Land Cover Change in the Bhutan–Bengal Foothill Region Between 1987 and 2019: Study Towards Geospatial Applications and Policy Making: M. Chamling, B. Bera. *Earth Systems and Environment*, 4(1), 117–130.

Dalla Valle, E., Lamedica, S., Pilli, R., Anfodillo, T., 2009. Land use change and forest carbon sink assessment in an alpine mountain area of the Veneto region (Northeast Italy). *Mountain Research and Development*, 29(2), 161–168.

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S. et al., 2020. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*.

Ebrahimi, S., Kumar, S., 2025. What helps to detect what? Explainable AI and multisensor fusion for semantic segmentation of simultaneous crop and land cover land use delineation. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 18, 5423–5444.

FAO, 2024. Land cover and crop monitoring. `https://www.fao.org/geospatial/our-focus/Land-Cover-Crop-Monitoring/en`. Food and Agriculture Organization of the United Nations. Accessed on 27 March 2026.

Fayaz, M., Nam, J., Dang, L. M., Song, H.-K., Moon, H., 2024. Land-cover classification using deep learning with high-resolution remote-sensing imagery. *Applied Sciences*, 14(5), 1844.

Gaur, S., Singh, R., 2023. A comprehensive review on land use/land cover (LULC) change modeling for urban development: current status and future prospects. *Sustainability*, 15(2), 903.

Ghiasi, G., Cui, Y., Srinivas, A., Qian, R., Lin, T.-Y., Cubuk, E. D., Le, Q. V., Zoph, B., 2021. Simple copy-paste is a strong data augmentation method for instance segmentation. *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2918–2928.

Giupponi, C., Ramanzin, M., Sturaro, E., Fuser, S., 2006. Climate and land use changes, biodiversity and agri-environmental measures in the Belluno province, Italy. *Environmental Science & Policy*, 9(2), 163–173.

Hwang, D., Kim, J.-J., Moon, S., Wang, S., 2025. Image augmentation approaches for building dimension estimation in street view images using object detection and instance segmentation based on deep learning. *Applied Sciences*, 15(5), 2525.

Janga, B., Asamani, G. P., Sun, Z., Cristea, N., 2023. A review of practical ai for remote sensing in earth sciences. *Remote Sensing*, 15(16), 4112.

Keshtkar, H., Voigt, W., Alizadeh, E., 2017. Land-cover classification and analysis of change using machine-learning classifiers and multi-temporal remote sensing imagery. *Arabian Journal of Geosciences*, 10(6), 154.

Klopfer, F., Gruehn, D., 2024. Leveraging open source land use and land cover data for urban and regional planning in transforming areas. *Regional Studies, Regional Science*, 11(1), 738–756.

Kupyn, O., Rupprecht, C., 2024. Dataset enhancement with instance-level augmentations. *European Conference on Computer Vision*, Springer, 384–402.

Li, W., Hsu, C.-Y., Wang, S., Witharana, C., Liljedahl, A., 2022. Real-time geoai for high-resolution mapping and segmentation of arctic permafrost features: the case of ice-wedge polygons. *Proceedings of the 5th ACM SIGSPATIAL international workshop on AI for geographic knowledge discovery*, 62–65.

Li, X., Zhao, H., Wu, D., Liu, Q., Tang, R., Li, L., Xu, Z., Lyu, X., 2024. SLMFNet: Enhancing land cover classification of remote sensing images through selective attentions and multi-level feature fusion. *Plos one*, 19(5), e0301134.

Lillesand, T., Kiefer, R. W., Chipman, J., 2015. *Remote sensing and image interpretation*. John Wiley & Sons.

Liu, A., Xiao, Z., 2023. Nighttime semantic segmentation with instance-level data augmentation: a case study of the dark zurich benchmark. *Proceedings of the 2023 7th international conference on machine learning and soft computing*, 175–180.

Mugiraneza, T., Hafner, S., Haas, J., Ban, Y., 2022. Monitoring urbanization and environmental impact in Kigali, Rwanda using Sentinel-2 MSI data and ecosystem service bundles. *International Journal of Applied Earth Observation and Geoinformation*, 109, 102775.

Pierdicca, R., Nepi, L., Mancini, A., Malinverni, E., Balestra, M., 2023. UAV4TREE: Deep learning-based system for automatic classification of tree species using RGB optical images obtained by an unmanned aerial vehicle. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 10, 1089–1096.

Qichi, Y., Lihui, W., Jinliang, H., Linzhi, L., Xiaodong, L., Fei, X., Yun, D., Xue, Y., Feng, L., 2023. A novel alpine land cover classification strategy based on a deep convolutional neural network and multi-source remote sensing data in Google Earth Engine. *GIScience & Remote Sensing*, 60(1), 2233756.

Rana, V. K., Suryanarayana, T. M. V., 2020. Performance evaluation of MLE, RF and SVM classification algorithms for watershed scale land use/land cover mapping using sentinel 2 bands. *Remote Sensing Applications: Society and Environment*, 19, 100351.

Xu, Z., Su, C., Zhang, X., 2021. A semantic segmentation method with category boundary for Land Use and Land Cover (LULC) mapping of Very-High Resolution (VHR) remote sensing image. *International journal of remote sensing*, 42(8), 3146–3165.

Yang, J., Jiang, Y., Song, Q., Wang, Z., Hu, Y., Li, K., Sun, Y., 2025. An approach for multi-source land use and land cover data fusion considering spatial correlations. *Remote Sensing*, 17(7), 1131.

Zhang, W., Yang, X., Yuan, Z., Chen, Z., Xu, Y., 2024. A Framework for Fine-Grained Land-Cover Classification Using 10 m Sentinel-2 Images. *Remote Sensing*, 16(2), 390.