



Comparison between different approaches for Land Cover classification

TESI DI LAUREA MAGISTRALE IN
GEOINFORMATICS ENGINEERING
INGEGNERIA GEOINFORMATICA

Author: **Yuxin Chen**

Student ID: 10875909

Advisor: Ludovico Giorgio Aldo Biagi

Co-advisor: Nikolina Zallemi

Academic Year: 2024-25

Abstract

With the rapid development of remote sensing and geographic information systems, accurate land cover classification has become essential for ecological monitoring and land use management. This study evaluates three classification algorithms—KMeans, Random Forest, and Multi-Layer Perceptron—using Sentinel-2 multispectral imagery across three regions: urbanized Shanghai, the agricultural hills of Tuscany, and the alpine area of Val d'Aosta. A standardized workflow was implemented, including image preprocessing, manual annotation, feature extraction, model training, result visualization, and accuracy assessment.

The results show that supervised models clearly outperform the unsupervised KMeans algorithm. Among them, MLP achieved the highest accuracy and spatial consistency, producing coherent and meaningful land cover maps. Random Forest also performed well in distinguishing spectrally distinct classes but was less effective in refining boundaries and representing minority categories. KMeans, by contrast, only captured coarse patterns and showed high levels of misclassification. The comparison of sampling strategies further revealed that uniform sampling improved recognition of minority classes and reduced imbalance, strengthening model robustness.

Overall, the findings indicate that MLP with uniform sampling provides the most effective framework for diverse land cover classification, offering guidance for large-scale remote sensing applications and ecological assessment.

The thesis is organized into seven chapters. Chapter 1 outlines the Earth Observation background, research objectives, and the Sentinel-2 dataset. Chapter 2 introduces the Google Earth Engine platform, describing data catalogs, classifier implementations, and evaluation metrics. Chapter 3 presents the study areas and training samples, while Chapter 4 details the preprocessing workflow, including cloud masking, band selection, and spectral index computation. Chapter 5 reports the classification results for KMeans, Random Forest, and Multi-Layer Perceptron across the study regions. Chapter 6 compares proportional and uniform sampling strategies, highlighting their impact on classification robustness. Finally, Chapter 7 synthesizes the overall results, emphasizing model performance and the role of sampling strategies, and provides recommendations for remote sensing applications and ecological assessments.

Key-words: GEE, Land Classification, Random Forest, K-means, MLP.

Abstract in italiano

Con il rapido sviluppo del telerilevamento e dei sistemi informativi geografici, una classificazione accurata della copertura del suolo è diventata fondamentale per il monitoraggio ecologico e la gestione del territorio. Questo studio valuta tre algoritmi di classificazione—KMeans, Random Forest e Multi-Layer Perceptron—utilizzando immagini multispettrali Sentinel-2 in tre regioni: l'area urbanizzata di Shanghai, le colline agricole della Toscana e la zona alpina della Valle d'Aosta. È adottato un flusso di lavoro standardizzato che comprende la pre-elaborazione delle immagini, l'annotazione manuale, l'estrazione delle caratteristiche, l'addestramento dei modelli, la visualizzazione dei risultati e la valutazione dell'accuratezza.

I risultati mostrano che i modelli supervisionati superano l'algoritmo non supervisionato KMeans. L'MLP ha raggiunto la massima accuratezza e coerenza spaziale, producendo mappe coerenti e significative. Random Forest ha dato buoni risultati nella distinzione di classi spettralmente diverse, ma è stato meno efficace nel definire i confini e rappresentare categorie minoritarie. KMeans, invece, ha colto solo schemi grossolani con elevata misclassificazione. Il confronto tra strategie di campionamento ha evidenziato che quello uniforme migliora il riconoscimento delle classi minoritarie e riduce lo squilibrio, rafforzando la robustezza del modello.

Nel complesso, i risultati indicano che l'MLP con campionamento uniforme rappresenta l'approccio più efficace per la classificazione della copertura del suolo, fornendo orientamenti per applicazioni di telerilevamento su larga scala e valutazioni ecologiche.

La tesi è articolata in sette capitoli: il Capitolo 1 presenta il contesto dell'Osservazione della Terra, gli obiettivi di ricerca e il dataset Sentinel-2; il Capitolo 2 introduce la piattaforma Google Earth Engine, descrivendo cataloghi di dati, implementazioni dei classificatori e metriche di valutazione; il Capitolo 3 illustra le aree di studio e i campioni di addestramento; il Capitolo 4 descrive il flusso di pre-elaborazione, comprendendo mascheramento delle nuvole, selezione delle bande e calcolo di indici spettrali; il Capitolo 5 riporta i risultati ottenuti con KMeans, Random Forest e Multi-Layer Perceptron; il Capitolo 6 confronta il campionamento proporzionale e uniforme, sottolineandone l'impatto sulla robustezza; infine, il Capitolo 7 sintetizza i risultati complessivi, evidenziando prestazioni dei modelli e implicazioni di strategie di campionamento, e propone raccomandazioni per applicazioni di telerilevamento e valutazioni ecologiche.

Parole chiave: GEE, Classificazione del territorio, Random Forest, K-means, MLP.

Contents

Abstract	i
Abstract in italiano	iii
Contents	v
1 Introduction	2
1.1. Earth Observation Background.....	2
1.2. Research Objective	4
1.3. Sentinel-2 dataset	5
2 Google Earth Engine	7
2.1. Overview	7
2.2. Data Catalogs in GEE	8
2.3. Classifiers in GEE	9
2.3.1. K-Means	10
2.3.2. Random Forest.....	12
2.3.3. Multi Layer Perceptron.....	15
2.4. Algorithm Performance Evaluation	22
2.4.1. Confusion Matrix.....	22
2.4.2. Accuracy Metrics	23
3 Study Areas	25
3.1. Shanghai	25
3.1.1. Description	25
3.1.2. Original Image	26
3.1.3. Training Samples.....	27
3.2. Tuscany	28
3.2.1. Description	28
3.2.2. Original Image	30
3.2.3. Training Samples	30
3.3. Aosta	31
3.3.1. Description	31
3.3.2. Original Image	33
3.3.3. Training Samples	33

4 Data Processing.....	34
4.1.1. Grid Number Determination	34
4.1.2. Cloud and Shadow Mask	35
4.1.3. Band Selection	36
4.1.4. Spectral Index Computation	37
4.1.5. Experimental Workflow	39
5 Results Analysis	40
5.1. KMeans	40
5.1.1. Shanghai.....	40
5.1.2. Tuscany	43
5.1.3. Aosta	46
5.2. Random Forest.....	50
5.2.1. Shanghai.....	50
5.2.2. Tuscany	52
5.2.3. Aosta	55
5.3. Multi Layer Perceptron.....	58
5.3.1. Hyperparameter Configurations.....	58
5.3.2. Shanghai.....	58
5.3.3. Tuscany	61
5.3.4. Aosta	65
6 Comparison of Uniform and Proportional Sampling Strategies	69
6.1. Shanghai: comparison on RF	69
6.2. Tuscany: comparison on RF.....	72
6.3. Aosta: comparison on RF	75
7 Conclusion	79
7.1. Comparative Analysis of Classification Models.....	79
7.2. Impact of Sampling Strategies on Classification Performance	80
Bibliography	81
List of Figures	84
List of Tables	86
Acknowledgments	87

1 Introduction

1.1. Earth Observation Background

In recent years, with the increasing environmental challenges facing the world, such as climate change, forest degradation, urban expansion, and land use change, the demand for obtaining dynamic information on the earth's surface has increased significantly. In this context, Earth Observation has been widely used in geographic information science, environmental science, agriculture, urban planning, and resource management as a non-contact, efficient, and sustainable monitoring method.

EO systems usually rely on remote sensing technology to regularly observe the earth's surface through sensors carried on spacecraft (such as satellites) and aircraft (such as drones and manned aircraft) to collect multi-phase, multi-scale, and multi-type surface information. Remote sensing systems can be divided into visible light remote sensing, multispectral remote sensing, hyperspectral remote sensing, and radar remote sensing (such as SAR) according to the type of electromagnetic wave bands collected. Among them, multispectral remote sensing is widely used in tasks such as land cover classification, vegetation monitoring, water body identification, and urban expansion analysis due to its moderate cost, wide spatial coverage, and flexible band combination.

Multispectral sensors usually cover several discrete bands from visible light (blue, green, and red) to near infrared and short-wave infrared. Different types of surface cover have different reflectance characteristics in these bands. For example, green vegetation has low reflectance in the red-light band and high reflectance in the near infrared band. This characteristic can be highlighted by calculating the NDVI (see Section 4.1.4. Spectral Index Computation); urban buildings, bare land, water bodies, etc. also present their own unique spectral responses. Therefore, the classification model based on multispectral remote sensing images can accurately identify different types of land on the surface without field sampling.

With the continuous advancement of Earth Observation (EO) technology, the accessibility and timeliness of data have significantly improved. For instance, the Sentinel series under the European Union's Copernicus programme, the US Landsat series, Japan's ALOS, and China's Gaofen series all provide medium-to-high resolution remote sensing imagery accessible to the public. Among these, the Sentinel-2 satellite boasts a spatial resolution of 10–60 metres, offering 13 spectral bands and a global revisit capability of five days. It stands as one of the most widely utilised data sources

in multispectral remote sensing today, proving particularly suitable for large-scale land cover monitoring. Specifically, Sentinel-2 bands can be categorised into three groups: 10-metre resolution bands include blue (B2), green (B3), red (B4), and near-infrared (B8); 20-metre resolution bands encompass the red edge bands (B5, B6, B7), narrow near-infrared band (B8A), and shortwave infrared bands (B11, B12); while 60-metre resolution bands are primarily used for atmospheric correction, including coastal aerosol (B1), water vapour (B9), and cirrus clouds (B10). This multi-resolution, multi-spectral design enables Sentinel-2 data to balance the extraction of fine surface features with atmospheric correction, thereby playing a vital role in studies such as land cover classification, ecological monitoring, and environmental assessment.

Satellite	Agency	Spatial Resolution	Spectral Bands	Band Details	Revisit Cycle	Open Access
Sentinel-2	ESA	10–60m	13	4 visible/NIR (10 m: B2–B4, B8); 6 red edge & SWIR (20 m: B5–B7, B8A, B11, B12); 3 atmospheric (60 m: B1, B9, B10)	5 days	Yes
Landsat-8/9	NASA	15–100m	11	OLI: 9 bands (coastal, blue, green, red, NIR, SWIR1, SWIR2, cirrus, pan 15 m); TIRS: 2 thermal bands (100 m)	16 days	Yes
Gaofen	CNSA	2–16m	4–8	Typically 4 multispectral (Blue, Green, Red, NIR) + Panchromatic (2 m); some satellites carry 8 bands (adding SWIR/Red Edge)	4–10 days	No
MODIS	NASA	250–1000m	36	36 spectral bands from 0.4–14.4 μm : 2 bands (250 m), 5 bands (500 m), 29 bands (1000 m)	Daily	Yes

Table 1. Comparison of Major Earth Observation Satellites.

In addition, the acquisition and processing of remote sensing data are increasingly dependent on the support of cloud computing platforms. Traditional local computing methods face limitations such as heavy computing burden and insufficient memory when processing large-area, multi-temporal remote sensing data. Platforms such as Google Earth Engine (GEE) provide instant access to global remote sensing data and cloud processing capabilities, greatly improving the analysis efficiency of EO data and promoting the application of machine learning and other methods in remote sensing classification.

In summary, EO data, especially multispectral remote sensing images, play an irreplaceable role in land cover classification. Its high spatiotemporal resolution, multispectral structure and open access characteristics provide data foundation and technical support for this study to carry out land feature classification analysis in three typical areas: urban (Shanghai), agricultural (Tuscany) and mountainous (Aosta).

1.2. Research Objective

Land-cover classification is a fundamental tool for environmental monitoring, resource management, and sustainable-development planning. Recent advances in remote-sensing technology—particularly the widespread use of high-resolution multispectral satellites such as Sentinel-2—have enabled large-scale, high-frequency monitoring of the Earth's surface. Nevertheless, because urban, agricultural, and mountainous areas differ markedly in land-cover composition and spatial structure, a classification method that performs well in one setting may be less effective in another. Urban areas exhibit complex and rapidly changing mosaics of buildings, roads, and green space; agricultural landscapes are dominated by the cyclic phenology of crops and intensive human management; mountainous regions are characterized by pronounced terrain, natural vegetation, exposed rock, and seasonal snow. These contrasts motivate a detailed examination of algorithm suitability across representative environments.

To this end, we analyse three study areas—Shanghai, China (February), Tuscany, Italy (April), and Aosta Valley, Italy (October)—chosen to capture typical seasonal conditions: late-winter urban stability in Shanghai, vigorous spring crop growth in Tuscany, and the autumn–winter transition in the Alpine Aosta region. Sentinel-2 multispectral imagery is classified with three widely used algorithms: K-means clustering, Random Forest, and a Multi-Layer Perceptron.

The specific research objectives include:

1. Establish a unified workflow—including data acquisition, preprocessing, feature extraction, model training, and evaluation—to ensure consistent inputs across algorithms;
2. Integrate Google Earth Engine with local deep-learning frameworks to implement K-means, RF, and MLP models, visualizing results and conducting accuracy assessments;
3. Evaluate algorithm performance in urban (Shanghai), agricultural (Tuscany), and mountainous (Aosta) contexts using confusion matrices and accuracy metrics, thereby exploring the algorithms' applicability and stability.

By adopting a regional comparative perspective, this study systematically assesses mainstream machine-learning classifiers under diverse geographical and seasonal conditions, enriches the body of knowledge on land-cover classification adaptability, and provides a solid scientific and practical foundation for precise land-resource monitoring and sustainable-development decision-making.

1.3. Sentinel-2 dataset

Sentinel-2 is a core component of the Copernicus Earth Observation Program led by the European Space Agency (ESA). The constellation consists of two polar-orbiting satellites: Sentinel-2A (launched in 2015) and Sentinel-2B (launched in 2017). Both satellites operate in a near-polar sun-synchronous orbit at an altitude of approximately 786 km, with an orbital period of about 100 minutes, and are equipped with the MultiSpectral Instrument (MSI). The system is characterized by high spatial resolution, wide spectral coverage, and high temporal frequency, enabling systematic acquisition of surface information. Sentinel-2 data have been widely used in applications such as land-use classification, crop monitoring, water resource assessment, forest inventory, and environmental change detection.

This study employs the Sentinel-2 Level-2A product, which provides surface reflectance data that have been atmospherically corrected by the Sen2Cor algorithm, thus more accurately reflecting the true reflectance characteristics of ground objects. The Sentinel-2 data products are organized into different processing levels:

- Level-0: Raw, unprocessed sensor data (original signal).
- Level-1B: Radiometrically calibrated data, still in detector geometry.
- Level-1C: Orthorectified Top-of-Atmosphere (TOA) reflectance, with spatial resolutions of 10 m, 20 m, and 60 m.
- Level-2A: Derived from Level-1C by atmospheric correction with Sen2Cor, providing Surface Reflectance (SR) along with scene classification and cloud masks. This level is more suitable for land-cover classification and environmental analysis tasks.

On the Google Earth Engine (GEE) platform, Level-2A data are available under the dataset COPERNICUS/S2_SR, which contains 12 primary spectral bands with spatial resolutions of 10 m, 20 m, and 60 m. More details are provided in Table 2. S2_SR Bands. Specifically, B1–B12 represent the primary spectral bands, each with a well-defined central wavelength (nm), whereas AOT, WVP, SCL, TCI, MSK, and QA are auxiliary information layers that do not correspond to specific spectral wavelengths; therefore, the wavelength column for these bands is left blank.

Sentinel-2 has a revisit period of about 5 days, making it particularly suitable for multitemporal change detection and classification studies in areas with high surface heterogeneity, such as urban regions, agricultural landscapes, and mountainous areas. Accordingly, this study focuses on three representative geographical regions—Shanghai, Tuscany, and Aosta—to conduct classification experiments and comparative analyses using Sentinel-2 data.

Name	Central Wavelength (nm)	Pixel Size	Description
B1	443	60 meters	Aerosols
B2	390	10 meters	Blue
B3	560	10 meters	Green
B4	665	10 meters	Red
B5	705	20 meters	Red Edge 1
B6	740	20 meters	Red Edge 2
B7	783	20 meters	Red Edge 3
B8	842	10 meters	NIR
B8A	865	20 meters	Red Edge 4
B9	945	60 meters	Water vapor
B11	1610	20 meters	SWIR 1
B12	2190	20 meters	SWIR 2
AOT	—	10 meters	Aerosol Optical Thickness
WVP	—	10 meters	Water Vapor Pressure. The height the water would occupy if the vapor were condensed into liquid and spread evenly across the column.
SCL	—	20 meters	Scene Classification Map (The "No Data" value of 0 is masked out)
TCI_R	—	10 meters	True Color Image, Red channel
TCI_G	—	10 meters	True Color Image, Green channel

TCI_B	—	10 meters	True Color Image, Blue channel
MSK_CLDPRB	—	20 meters	Cloud Probability Map (missing in some products)
MSK_SNWPRB	—	10 meters	Snow Probability Map (missing in some products)
QA10	—	10 meters	Always empty
QA20	—	20 meters	Always empty
QA60	—	60 meters	Cloud mask. Masked out between 2022-01-25 to 2024-02-28 inclusive.

Table 2. S2_SR Bands.

2 Google Earth Engine

2.1. Overview

Google Earth Engine (GEE) is a cloud-based platform developed by Google for large-scale geospatial data storage, processing, and analysis. It provides access to an extensive archive of open-source remote sensing and geographic datasets—including Landsat, MODIS, Sentinel missions, global digital elevation models (DEMs), and various land cover and climate products. GEE enables users to write compact and efficient scripts in JavaScript or Python to perform rapid geospatial analyses.

Compared with traditional workflows that require downloading and preprocessing imagery locally, GEE significantly reduces dependence on local storage and computational power. This makes it particularly suitable for regional-to-global land cover classification. Its distributed cloud architecture supports parallel computing across Google’s data centers, allowing users to run complex operations without hosting large datasets locally.

In this study, GEE was accessed via the Python API within a Jupyter Notebook environment (using VS Code), supporting a seamless end-to-end pipeline, including data ingestion, feature extraction, training sample upload, classification model deployment, map visualization, and accuracy assessment.

GEE was selected for the following reasons:

1. Data accessibility and consistency:

Sentinel-2 Surface Reflectance imagery, along with auxiliary quality assurance layers (e.g., QA60, SCL, AOT, WVP), is available in a harmonized, atmospherically corrected format. This ensures a consistent and reliable data source across all processing stages without the need for manual downloads.

2. Operational efficiency:

Although this study utilizes only a single Sentinel-2 image per region, GEE simplifies the process with built-in functions for cloud masking, temporal filtering, and region-based clipping. Moreover, it facilitates cross-regional comparisons (e.g., Shanghai, Tuscany, Aosta) through a consistent data structure and processing environment.

3. Reproducible and integrable workflow:

By combining GEE's server-side computation with the Python ecosystem of machine learning libraries, this study trains models locally (e.g., MLP), pushes predictions back to GEE for map generation, and shares the entire codebase for reproducibility. This hybrid cloud-local approach balances computational efficiency with development flexibility.

Together, these capabilities make GEE a scalable, transparent, and efficient framework for land cover classification tasks in this research.

2.2. Data Catalogs in GEE

One of GEE's core strengths is its structured Data Catalog, which provides programmatic access to a broad array of global datasets. These datasets are curated and organized by source agencies (e.g., ESA, NASA, USGS, NOAA) and are grouped by product type—such as satellite imagery, climate data, land cover maps, and elevation models. Each dataset is assigned a unique identifier, which allows users to retrieve and analyze it directly through the GEE API, without any local storage or download requirement.

In this study, the primary data source is the Sentinel-2 Level-2A Surface Reflectance dataset, accessed via the GEE data catalog identifier: **COPERNICUS/S2_SR**

This dataset, provided by the European Space Agency (ESA) under the Copernicus program, has the following characteristics:

- Spatial resolution: 10–60 meters (depending on band)
- Temporal resolution: ~5 days revisit frequency
- Processing level: Level-2A (atmospherically corrected surface reflectance)
- Spectral bands: 13 bands in total; this study used bands B2–B12

For a full description of the Sentinel-2 satellite platform and its imaging specifications, refer to Section 1.3: Sentinel-2 Dataset.

To enhance feature separability during classification, several spectral indices were derived from the Sentinel-2 bands, including NDVI, NDWI, NDBI, and BSI. The specific formulations and interpretations of these indices are described in Section 4.1.4: Spectral Index Computation.

By leveraging the COPERNICUS/S2_SR dataset and GEE's built-in tools for band selection and index computation, the study efficiently transformed raw satellite imagery into structured, information-rich feature sets that form the basis for classification modeling.

2.3. Classifiers in GEE

A key feature of Google Earth Engine is its built-in `ee.Classifier` module, which supports a wide range of machine learning algorithms for both supervised and unsupervised classification. This cloud-based approach eliminates the need for local data storage or high-performance computing, enabling users to perform complex classification tasks directly on the cloud.

The available classifiers in GEE include, but are not limited to, Random Forest, Naive Bayes, Support Vector Machine (SVM), K-means, and Cobweb. These classifiers enable flexible modeling of diverse land cover scenarios across various spatial and temporal scales.

In this study, three representative classification approaches were selected to represent different machine learning paradigms:

- **Unsupervised classification:** Performed using the K-means clustering algorithm (`ee.Clusterer.wekaKMeans`) in GEE, allowing for pattern discovery without labeled data.
- **Supervised classification:** Implemented using the Random Forest classifier (`ee.Classifier.smileRandomForest`) within GEE, trained on manually labeled sample points.
- **Neural network classification:** Conducted outside of GEE using the Multi-Layer Perceptron (`MLPClassifier`) from the scikit-learn library in Python, enabling the modeling of non-linear class boundaries based on deep learning principles.

This multi-strategy design facilitates a comparative analysis of classification performance across different algorithm families. By combining cloud-based and local approaches, the study ensures both scalability and flexibility in the land cover mapping workflow.

2.3.1. K-Means

2.3.1.1. Algorithm Principle

K-means is a partition-based unsupervised clustering algorithm that aims to divide data into K non-overlapping clusters so that the similarity between samples within a cluster is as high as possible, while the similarity between samples between clusters is as low as possible. The algorithm takes minimizing the sum of squared errors (SSE) between samples and the center of their cluster as its objective function. Each cluster is represented by a "centroid", which is the mean of the features of all samples in the cluster.

The core steps of the algorithm include:

1. Initialization: randomly select K initial centroids;
2. Assignment step: assign each sample to the cluster with the nearest centroid;
3. Update step: recalculate the mean of all samples in each cluster as the new centroid;
4. Iterative optimization: repeat the "assignment-update" process until the centroid position converges or the set maximum number of iterations is reached.

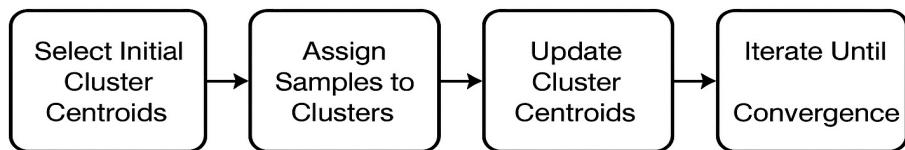


Figure 1. K-means steps.

K-means usually uses Euclidean distance to measure the similarity between samples and centroids, which is defined as follows:

$$D = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \dots + (x_n - y_n)^2} \quad (1)$$

where x and y are the feature vectors of two data points.

The results of K-means clustering are highly dependent on the selection of the initial centroid and may fall into a local optimal solution. Therefore, in practical applications, multiple random initialization runs are often used, and the result with the smallest SSE is selected as the final clustering output to improve the stability and reliability of the algorithm.

2.3.1.2. GEE Platform Implementation

In the Google Earth Engine platform, the K-means clustering algorithm is implemented via the `ee.Clusterer.wekaK-means()` interface, which is based on the Weka machine learning library. In practical applications, the input for K-means is

typically a FeatureCollection used for training, which contains attributes for clustering. Input features can come from selected remote sensing image bands or derived features calculated from bands (such as NDVI, NDWI, etc.).

In this study, the number of clusters was defined according to the land cover classes specified for each study area. Specifically, five clusters were used for Shanghai and Tuscany, while six clusters were applied for Aosta. The sampling scale was set to 10 m, and the geometries parameter was specified as False in order to optimize computational efficiency.

Usage	Returns
<code>ee.Clusterer.wekaK-means(nClusters,init,canopies,maxCandidates, periodicPruning,minDensity,t1,t2,distanceFunction,maxIterations, preserveOrder, fast, seed)</code>	Clusters

Table 3. K-means Usage.

Argument	Type	Details
nClusters	Integer	Number of clusters.
init	Integer, default: 0	Initialization method to use. 0 = random, 1 = K-means++, 2 = canopy, 3 = farthest first.
canopies	Boolean, default: false	Use canopies to reduce the number of distance calculations.
maxCandidates	Integer, default: 100	Maximum number of candidate canopies to retain in memory at any one time when using canopy clustering. T2 distance plus, data characteristics, will determine how many candidate canopies are formed before periodic and final pruning are performed, which might result in excess memory consumption. This setting avoids large numbers of candidate canopies consuming memory.
periodicPruning	Integer, default: 1000	How often to prune low density canopies when using canopy clustering.
minDensity	Integer, default: 2	Minimum canopy density, when using canopy clustering, below which a canopy will be pruned during periodic pruning.
t1	Float, default: -1.5	The T1 distance to use when using canopy clustering. A value < 0 is taken as a positive multiplier for T2.

t2	Float, default: -1	The T2 distance to use when using canopy clustering. Values < 0 cause a heuristic based on attribute std. deviation to be used.
distanceFunction	String, default: "Euclidean"	Distance function to use. Options are: Euclidean and Manhattan.
maxIterations	Integer, default: null	Maximum number of iterations.
preserveOrder	Boolean, default: false	Preserve order of instances.
fast	Boolean, default: false	Enables faster distance calculations, using cut-off values. Disables the calculation/output of squared errors/distances.
seed	Integer, default: 10	The randomization seed.

Table 4. K-means Parameter.

2.3.1.3. Algorithm Construction Process

The construction process of the K-means classifier includes the following key steps: first, a number of initial cluster centroids are selected randomly from the input data according to the predefined number of clusters K; then, each sample is assigned to the cluster whose centroid has the minimum Euclidean distance from the sample; after all samples are assigned, the centroid of each cluster is recalculated as the mean of all samples within that cluster; this assignment–update process is iteratively repeated—each time reassigning samples and updating centroids—until the centroids converge or the maximum number of iterations is reached. Since K-means is an unsupervised method, the resulting cluster labels are numerical and not semantically meaningful. Therefore, a post-processing step is required to remap the cluster labels to actual land cover types, based on prior knowledge, sample distribution, or visual interpretation.

2.3.2. Random Forest

Random Forest (RF) is a supervised learning algorithm based on the idea of integration, which consists of multiple mutually independent decision trees. By introducing data perturbation and feature perturbation mechanisms, the algorithm effectively suppresses the overfitting phenomenon while improving the generalization ability of the model, and thus it is widely used in the land cover classification task of remote sensing images.

2.3.2.1. Algorithm Principle

The basic idea of random forest is to build a set of independent decision trees during the training phase, each of which is generated by different training samples and feature subsets. During classification, each tree independently determines the category to which the input sample belongs, and finally determines the classification result by "majority voting". Compared with the traditional single decision tree, random forest

effectively improves the stability and accuracy of the overall model by integrating multiple high-variance weak learners while maintaining a low bias.

Its core mechanisms include:

- Bootstrap Sampling: Each tree is trained using a subsample set randomly extracted from the original training set with replacement.
- Random feature selection: Each time a node is split, only a portion of all features is randomly selected to participate in the division, which enhances the differences between trees.
- No pruning strategy: Trees in random forests usually grow completely to leaf nodes without pruning, and the risk of overfitting is controlled by the overall integration effect.
- Majority voting decision: The prediction results of all trees determine the final classification by voting, which improves the robustness of the prediction.

2.3.2.2. Training Process

Training a random forest classifier typically involves the following steps:

1. Prepare training data: Construct feature matrix $X \in R^{n \times p}$, label vector $y \in R^n$, where n is the number of samples and p is the feature dimension.
1. Build forest (train T trees in total):
 - For each tree: generate a sample subset from the original training set using bootstrap sampling;
 - At each node split: randomly select m features from p features (generally set $m = \sqrt{p}$); calculate the optimal partition among these m features (for example, based on Gini impurity); perform node partitioning and recursively grow the entire tree until the stopping condition is met (such as the minimum number of leaf node samples or the maximum tree depth).
1. Prediction and classification: For a new input sample x , each tree gives a predicted category, and finally the classification output is determined by the majority vote of all trees.

2.3.2.3. GEE Platform Implementation

In Google Earth Engine, random forests are easily available through the `ee.Classifier.smileRandomForest()` interface based on the SMILE (Statistical Machine Intelligence and Learning Engine) library.

In this study, the Random Forest classifier was configured with the following parameters: the number of trees was set to 100 to ensure sufficient decision-making capacity; the random seed was fixed at 42 to guarantee reproducibility of the experiments; the minimum leaf population was defined as 10 in order to prevent

overfitting and improve classification stability; and the bag fraction was set to 0.7, introducing randomness during the bootstrap sampling process to enhance the model's generalization ability.

Usage	Returns
<code>ee.Classifier.smileRandomForest(numberOfTrees,variablesPerSplit, minLeafPopulation, bagFraction, maxNodes, seed)</code>	Classifier

Table 5. Random Forest Usage.

Argument	Type	Details
number Of Trees	Integer	The number of decision trees to create.
variables Per Split	Integer, default: null	The number of variables per split. If unspecified, uses the square root of the number of variables.
min Leaf Population	Integer, default: 1	Only create nodes whose training set contains at least this many points.
bag Fraction	Float, default: 0.5	The fraction of input to bag per tree.
max Nodes	Integer, default: null	The maximum number of leaf nodes in each tree. If unspecified, defaults to no limit.
seed	Integer, default: 0	The randomization seed.

Table 6. Random Forest Parameter.

2.3.2.4. Algorithm Construction Process

The core idea of Random Forest is to construct a "forest" of uncorrelated decision trees during training. When a new input sample is classified, each tree independently provides a classification result, and the final output is determined by majority voting across all trees.

The construction process of the random forest classifier includes the following key steps: first, bootstrap sampling with replacement is performed from the original training set to generate the training subset required for each decision tree; then, when each node is split, a subset is randomly selected from all the features, and the optimal partitioning feature is determined among them for node partitioning; then, the entire tree is recursively grown according to the above strategy until the preset termination condition (such as the minimum number of leaf node samples or the maximum tree depth) is met, and no pruning operation is performed during the entire process; finally, a random forest is formed by constructing a large number of decision trees, and a majority vote is performed on the new samples in the classification stage to output the final classification result.

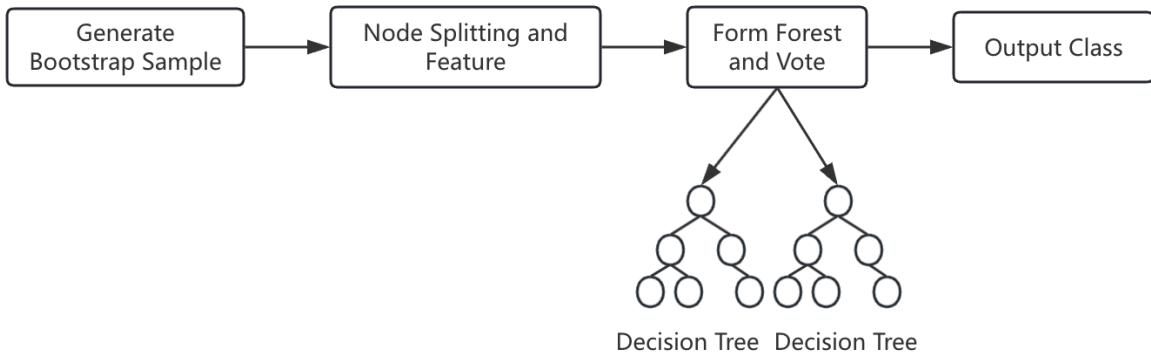


Figure 2. Random Forest Construction.

2.3.3. Multi Layer Perceptron

2.3.3.1. Neural Network

Neural network is an artificial neural network model designed based on the structure and functional characteristics of biological nervous system, with strong adaptability and nonlinear mapping ability. Neural network consists of multiple neurons (or nodes), which are connected to each other through connection weights to form a multi-layer network structure. Each neuron receives signals from other neurons, and performs nonlinear transformation through activation function after weighted linear combination of these signals, and finally outputs them to the next layer of neurons or output layer.

Neural networks can be divided into many different types. Here are some common types of neural networks:

1. Feedforward Neural Network: Feedforward neural network is the most basic type of neural network and the most common type of neural network in deep learning. It consists of several neurons in a certain hierarchical structure. Each neuron receives the output of the previous layer and generates the output of this layer, thereby realizing the transmission and processing of information.
2. Convolutional Neural Network: Convolutional Neural Network is a type of neural network specifically used for image processing and computer vision tasks. It can extract features from images through operations such as convolution and pooling, thereby realizing tasks such as image classification, target detection, and image segmentation.
3. Recurrent Neural Network: Recurrent Neural Network is a type of neural network that can process sequence data. It can process sequence data of any length through memory units and gating mechanisms, thereby realizing tasks such as natural language processing and speech recognition.

4. Autoencoder: Autoencoder is a type of neural network for unsupervised learning. Its goal is to compress and decompress input data to achieve tasks such as feature extraction and dimensionality reduction.
5. Deep Belief Network: A deep belief network is a type of neural network composed of multiple restricted Boltzmann machines. It can achieve efficient feature learning and classification tasks through layer-by-layer greedy pre-training and fine-tuning.

In addition to the types of neural networks listed above, there are many other types of neural networks, such as back propagation neural networks, Hopfield networks, Boltzmann machines, etc. Different neural network types are suitable for different tasks and data types.

2.3.3.2. Algorithm Principle

Multi-Layer Perceptron is a typical feedforward artificial neural network, which is widely used to deal with complex nonlinear classification problems, especially for the recognition and discrimination of land object types in remote sensing images. Its basic structure consists of three parts: input layer, one or more hidden layers and output layer. Each layer contains several neurons, and information is transmitted between layers through weighted connections.

In the network, the input layer is responsible for receiving the multidimensional features of image pixels, the output layer is used to output the final classification results, and the middle hidden layer is responsible for feature extraction and nonlinear mapping. Each neuron performs weighted summation on the output from the previous layer and applies a nonlinear activation function (such as ReLU, Sigmoid or Tanh) to enhance the network's expression ability.

The core algorithm of MLP is backpropagation, which continuously adjusts the weights and bias parameters in the network by minimizing the loss function (such as cross entropy or mean square error). The training process is optimized using gradient descent or its improved algorithms (such as Adam, RMSprop).

Benefit from its multilayer structure and nonlinear activation mechanism, MLP has a strong fitting ability to automatically learn hidden features from complex remote sensing data, capture the higher-order relationship between input features and target categories, and realize accurate classification tasks.

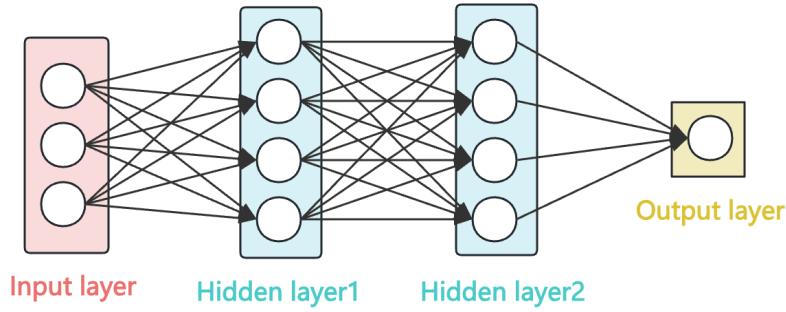


Figure 3. MLP Structure.

2.3.3.3. Training Process

The training process of MLP aims to learn the nonlinear mapping relationship between output categories through input features. The training goal is to minimize the error between the model prediction results and the true labels, so as to obtain a classifier that can generalize. The whole training process is as follows:

1. Data preparation

Extract training samples from remote sensing images and construct training data sets. Assume that the input feature matrix is:

$$\mathbf{X} = \begin{bmatrix} x^{(1)} \\ x^{(2)} \\ \vdots \\ x^{(n)} \end{bmatrix} \in \mathbb{R}^{n \times p}, \quad x^{(i)} = [x_1^{(i)}, x_2^{(i)}, \dots, x_p^{(i)}]^T \quad (2)$$

Where n is the number of samples and p is the feature dimension of each sample. The corresponding labels are:

$$\mathbf{y} = \begin{bmatrix} y^{(1)} \\ y^{(2)} \\ \vdots \\ y^{(n)} \end{bmatrix} \in \mathbb{R}^n, \quad y^{(i)} \in \{1, 2, \dots, C\} \quad (3)$$

Where C is the total number of categories for classification.

2. Network structure design

The MLP network contains L layers (including hidden layers and output layers), and the lth layer has $h^{(l)}$ neurons. The network structure includes:

- Input layer: size is p;
- Hidden layer (one or more): the number of neurons in each layer is $h^{(1)}, h^{(2)}, \dots$;
- Output layer: size is C, corresponding to the number of categories.

The calculation of neurons in each layer is as follows:

$$z^{(l)} = W^{(l)}a^{(1-l)} + b^{(l)}, a^{(l)} = f^{(l)}z^{(l)} \quad (4)$$

- $W^{(l)} \in \mathbb{R}^{h^{(l)} \times h^{(l-1)}}$ is the weight matrix;
- $b^{(l)} \in \mathbb{R}^{h^{(l)}}$ is the bias vector;
- $f^{(l)}$ is the activation function, such as ReLU: $f(x) = \max(0, x)$, or Sigmoid: $f(x) = \frac{1}{1+e^{-x}}$;
- $a^{(0)} = x$ is the input sample.

3. Forward Propagation

For each input sample $x^{(i)}$, we start from the input layer and calculate layer by layer, and finally generate the predicted probability $\hat{y}^{(i)}$ at the output layer:

$$\hat{y}^{(i)} = \text{softmax}(z^{(L)}) = \frac{e^{z_j^{(L)}}}{\sum_{k=1}^C e^{z_k^{(L)}}} \quad (5)$$

Output the predicted probability for each category.

4. Loss Computation

Use the cross-entropy loss function to measure the error between the predicted value and the true label:

$$\mathcal{L}(\hat{y}, y) = - \sum_{j=1}^C y_j \log(\hat{y}_j) \quad (6)$$

- $y_j \in \{0, 1\}$ is the One-Hot encoding of the true label;
- \hat{y}_j is the predicted probability.

5. Backpropagation and Optimization

Calculate the partial derivatives of the loss for each layer of parameters using the chain rule:

$$\frac{\partial \mathcal{L}}{\partial \mathbf{W}^{(l)}} = \delta^{(l)}(a^{(l-1)})^T, \quad \frac{\partial \mathcal{L}}{\partial \mathbf{b}^{(l)}} = \delta^{(l)} \quad (7)$$

The error term recursive formula is:

$$\delta^{(L)} = \hat{y} - y, \quad \delta^{(l)} = \left((\mathbf{W}^{(l+1)})^T \delta^{(l+1)} \right) \circ f'^{(l)}(z^{(l)}) \quad (8)$$

where \circ represents element-wise multiplication (Hadamard product).

Update parameters using optimization algorithms (such as SGD, Adam):

$$\mathbf{W}^{(l)} \leftarrow \mathbf{W}^{(l)} - \eta \cdot \frac{\partial \mathcal{L}}{\partial \mathbf{W}^{(l)}} \quad (9)$$

6. Training Loop

Repeat the process of forward propagation, loss calculation, backward propagation, and weight updating until the loss function converges or the maximum number of training epochs is reached.

2.3.3.4. GEE Platform Implementation

Unlike RF and K-means, Google Earth Engine does not natively support neural network models. Therefore, in this study, the training of the MLP model was conducted in a local Python environment (e.g., VS Code) using common deep learning libraries like scikit-learn (**from sklearn.neural_network import MLPClassifier**).

Usage	Returns
<code>MLPClassifier(hidden_layer_sizes, activation, solver, alpha, batch_size, learning_rate, learning_rate_init, power_t, max_iter, shuffle, random_state, tol, verbose, warm_start, momentum, nesterovs_momentum, early_stopping, validation_fraction, beta_1, beta_2, epsilon, n_iter_no_change, max_fun)</code>	Predict

Table 7. MLP Usage.

Argument	Type	Details
hidden_layer_sizes	array-like of shape(n_layers - 2,), default=(100,)	The ith element represents the number of neurons in the ith hidden layer.
activation	{'identity', 'logistic', 'tanh', 'relu'}, default='relu'	Activation function for the hidden layer.
solver	{'lbfgs', 'sgd', 'adam'}, default='adam'	The solver for weight optimization.
alpha	float, default=0.0001	Strength of the L2 regularization term.
batch_size	int, default='auto'	Size of minibatches for stochastic optimizers.

learning_rate	{'constant', 'invscaling', 'adaptive'}, default='constant'	Learning rate schedule for weight updates.
learning_rate_	initfloat, default=0.001	The initial learning rate used.
power_t	float, default=0.5	The exponent for inverse scaling learning rate.
max_iter	int, default=200	Maximum number of iterations.
shuffle	bool, default=True	Whether to shuffle samples in each iteration.
random_state	int, RandomState instance, default=None	Random seed for reproducibility.
tol	float, default=1e-4	Tolerance for the optimization.
verbose	bool, default=False	Whether to print progress messages to stdout.
warm_start	bool, default=False	When set to True, reuse the solution of the previous call to fit as initialization, otherwise, just erase the previous solution.
momentum	float, default=0.9	Momentum for gradient descent update. Should be between 0 and 1.
nesterovs_momentum	bool, default=True	Whether to use Nesterov's momentum. Only used when solver='sgd' and momentum > 0.
early_stopping	bool, default=False	Whether to use early stopping to terminate training when validation score is not improving.
validation_fraction	float, default=0.1	The proportion of training data to set aside as validation set for early stopping.
beta_1	float, default=0.9	Exponential decay rate for estimates of first moment vector in adam, should be in [0, 1).
beta_2	float, default=0.999	Exponential decay rate for estimates of second moment vector in adam, should be in [0, 1).
epsilon	float, default=1e-8	Value for numerical stability in adam.
n_iter_no_change	int, default=10	Maximum number of epochs to not meet tol improvement.
max_fun	int, default=15000	Maximum number of loss function calls.

Table 8. MLP Parameter.

After training, the model's prediction results or classification weights were imported into the GEE platform for classification inference and map generation, thereby achieving an efficient integration of local training with cloud-based visualization.

2.3.3.5. Parameter Tuning

In the training of the Multi-Layer Perceptron classifier, the choice of hyperparameters plays a crucial role in determining both convergence speed and classification accuracy. To avoid the uncertainty of manually chosen parameters, this study employed **Grid Search combined with Cross-Validation** to systematically identify the optimal hyperparameter configuration within a predefined candidate space.

1. Principle of GridSearchCV

The fundamental idea of grid search is to exhaustively evaluate all possible combinations within a given hyperparameter space, and assess their performance via cross-validation. Specifically, the training data are first standardized, and then each parameter configuration is evaluated using three-fold cross-validation ($cv=3$) with classification accuracy as the performance metric. Finally, GridSearchCV returns the parameter set that achieves the highest average performance across folds, which is then used as the optimal model for subsequent inference.

2. Design of the Hyperparameter Search Space

In designing the search space, this study considered four key hyperparameters, each with multiple candidate values. Their significance and design rationale are as follows:

a) activation: tanh / relu

The activation function introduces non-linearity into the hidden layers. The tanh function squashes outputs into the range $-1,1$ and is zero-centered, which can create clearer decision boundaries in datasets with complex distributions. The relu function, on the other hand, is unbounded for positive values and computationally efficient, favoring sparse activation. Comparing the two allows us to assess which non-linear mapping better suits the spectral characteristics of remote sensing data.

b) alpha: 0.0001 / 0.001 / 0.05

This parameter controls the strength of L2 regularization, penalizing large weights to prevent overfitting. The default value is 0.0001, while larger values (e.g., 0.05) indicate stronger regularization, forcing the model to be simpler and more generalizable. If smaller alpha values lead to overfitting on the training set, a larger alpha can significantly improve performance on unseen validation data.

c) hidden_layer_sizes: (50,50,50) / (100,50) / (128,64,32)

This parameter defines the architecture of the neural network, i.e., the number of layers and neurons per layer. The three candidate architectures represent a

shallow three-layer network, a moderate two-layer network, and a deeper three-layer network, respectively. Simpler networks are less prone to memorizing noise, while deeper networks provide stronger capacity to capture complex spectral patterns in land cover data.

d) **learning_rate: constant / adaptive**

The learning rate determines how weights are updated during training. The constant option applies a fixed learning rate, while adaptive decreases the learning rate after several stagnant iterations. Since the adam optimizer already includes adaptive mechanisms, a constant base learning rate often provides greater stability, allowing adam to operate more effectively without drastic external adjustments.

3. Training and Validation Procedure

During training, input features were first standardized using StandardScaler: the training set was transformed with `fit_transform`, while the validation and inference datasets were standardized using the same scaler via `transform` to maintain consistency. Subsequently, GridSearchCV exhaustively searched through all parameter combinations within the defined space, training models and computing the average accuracy via three-fold cross-validation. The parameter set with the best average performance was selected as the optimal configuration for each AOI and further validated on independent test data.

2.4. Algorithm Performance Evaluation

In remote sensing land cover classification and machine learning applications, evaluating the performance of classification algorithms is a critical step in ensuring the reliability and scientific validity of the results. In particular, for land cover classification, the model's effectiveness depends not only on its ability to assign labels but also on how accurately it distinguishes between different land cover types.

This section introduces the performance evaluation framework adopted in this study, including the principles and structure of the confusion matrix and a set of derived accuracy metrics used to quantify classification performance.

2.4.1. Confusion Matrix

The confusion matrix is a fundamental and widely used tool for evaluating the performance of classification models. In the context of land cover classification using remote sensing imagery, it provides an intuitive and comprehensive way to assess how well the predicted results agree with the actual ground truth data.

The confusion matrix is essentially a two-dimensional contingency table where:

- Rows correspond to the true labels (reference land cover classes),

- Columns correspond to the predicted labels (model outputs),
- Diagonal elements represent the number of correctly classified samples (True Positives),
- Off-diagonal elements reflect misclassified samples, including both False Positives and False Negatives.

By comparing model predictions with actual categories, the confusion matrix reveals not only the overall classification accuracy but also class-specific strengths and weaknesses—i.e., which land cover types the model identifies reliably and where it tends to confuse categories. This diagnostic capability makes the confusion matrix a critical basis for further model tuning and performance improvement.

In its typical form, the confusion matrix includes four fundamental components:

- True Positive (TP): The number of instances correctly predicted as belonging to a given class.
- False Positive (FP): The number of instances incorrectly predicted as belonging to a class (actually belong to another class).
- True Negative (TN): The number of instances correctly predicted as not belonging to a class.
- False Negative (FN): The number of instances incorrectly predicted as not belonging to a class (actually do belong).

An example of a binary confusion matrix is presented in Table 10, with TP, FP, TN, and FN values clearly organized.

		Predict	
		TP	FN
True	TP		
	FP		TN

Table 9. Confusion Matrix.

2.4.2. Accuracy Metrics

Based on the confusion matrix, several performance indicators can be computed to quantify how well the model performs across different classes. This study uses four common and informative metrics: User's Accuracy (UA), Producer's Accuracy (PA), Overall Accuracy (OA), and the Kappa coefficient (k).

2.4.2.1. User's Accuracy (UA)

User's Accuracy measures the proportion of correctly classified samples among all those predicted as a certain class. It reflects the reliability of the map from a user's perspective:

$$UA = \frac{TP}{TP+FP} \quad (10)$$

A high UA indicates fewer commission errors and higher prediction reliability.

2.4.2.2. Producer's Accuracy (PA)

Producer's Accuracy measures the proportion of correctly classified samples among all reference samples that actually belong to that class. It reflects the model's sensitivity or ability to correctly detect a class:

$$PA = \frac{TP}{TP+FN} \quad (11)$$

A high PA implies fewer omission errors and better model recall for the given class.

2.4.2.3. Overall Accuracy (OA)

Overall Accuracy is the most intuitive metric and indicates the percentage of correctly classified samples over the total number of samples:

$$OA = \frac{TP+TN}{TP+FN+FP+TN} \quad (12)$$

2.4.2.4. Kappa Coefficient (k)

The Kappa coefficient measures the agreement between the classification result and the reference data while accounting for chance agreement. It complements OA by adjusting for randomly correct predictions:

$$k = \frac{OA-E}{1-E} \quad (13)$$

Where:

- E: Expected accuracy by chance, based on marginal totals in the confusion matrix

Interpretation of k values is as follows:

Kappa Value	Interpretation
< 0.00	Poor agreement
0.00 – 0.20	Slight agreement
0.21 – 0.40	Fair agreement
0.41 – 0.60	Moderate agreement
0.61 – 0.80	Substantial agreement
0.81 – 1.00	Almost perfect agreement

Table 10. Kappa values.

3 Study Areas

3.1. Shanghai

3.1.1. Description

Shanghai is located on the eastern coast of China, at the mouth of the Changjiang River. Its geographical coordinates are approximately 120°52' to 122°12' east longitude and 30°40' to 31°53' north latitude. It is one of China's four municipalities and a national economic, financial, trade, and shipping center. The city covers a total area of approximately 6,340 square kilometers, with a flat terrain and an average elevation of less than 5 meters. It is characterized by a typical subtropical monsoon climate, featuring distinct seasons, abundant annual precipitation, and an average annual temperature ranging from 15–18°C.

In terms of land use structure, Shanghai is primarily characterized by high-density artificial construction land, including residential, commercial, and industrial zones, as well as a widespread network of roads, infrastructure, and urban open spaces. The main land categories can be divided into: built, water, trees, grass, bare and crops. The spatial mixing of urban land surface types is high, geometric structures are complex, and spectral differences are subtle, placing high demands on remote sensing classification accuracy.

This study selected February Sentinel-2 imagery as the classification time point for the Shanghai region, primarily based on the following considerations: first, February is at the end of winter, when vegetation is in a dormant or deciduous state, NDVI and other indices are relatively low, making artificial features (such as buildings and bare) more prominent in the imagery, which enhances the distinguishability of non-vegetation land categories. Second, atmospheric conditions are relatively stable during this period, making it easier to obtain high-quality imagery with low cloud cover. Additionally, fixed urban features such as buildings and water bodies are least affected by seasonal changes, facilitating the construction of stable single-period classification results and providing reliable urban reference samples.

Shanghai, as a typical urban area with complex artificial landforms and a highly urbanized background, serves as an important experimental region for testing and validating remote sensing classification methods, particularly in terms of the accuracy, stability, and adaptability of landform classification algorithms.



Figure 4. Map of China with Shanghai position.

3.1.2. Original Image

The study area is located in the eastern part of Shanghai, China, with grid number 51RUQ, covering an area of about 50km x 40km. The data collection time is 27th February, 2022. The area is dominated by a large number of built-up areas with typical urbanization features.

Image ID: 20220227T023639_20220227T024821_T51RUQ.

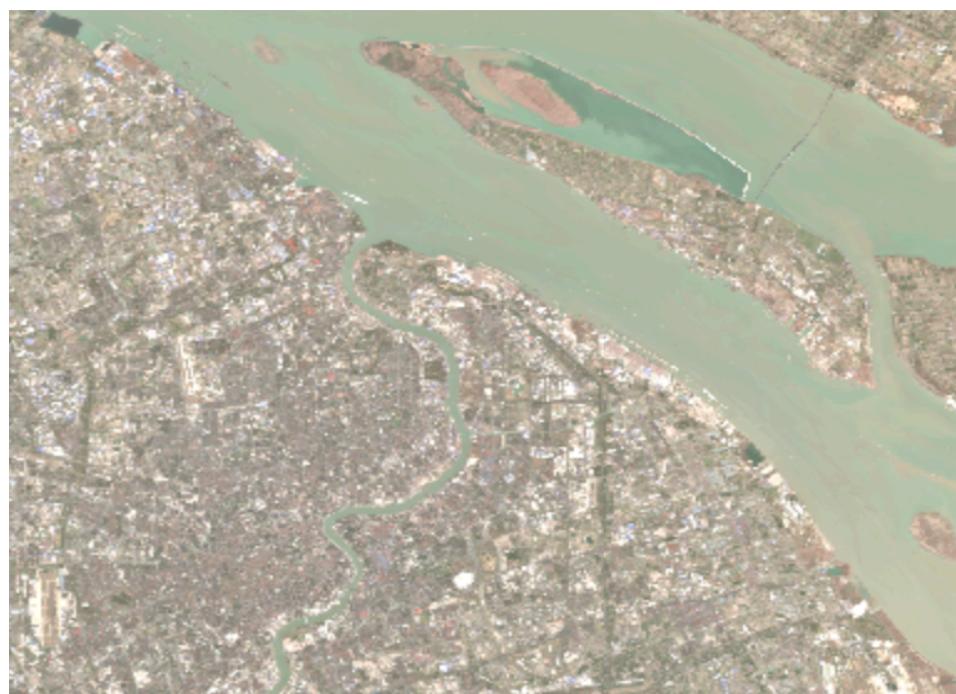


Figure 5. Shanghai Original Image.

3.1.3. Training Samples

In supervised classification, the selection and distribution of training samples is one of the key factors affecting classification accuracy. An important principle in training sample construction is to avoid class imbalance that may introduce bias during model learning. If one category is overrepresented in the training set, the model is likely to be biased toward the majority class, leading to misclassification or omission of minority classes. Conversely, artificially oversampling minority categories may result in overfitting, where the model performs well on the training data but shows reduced accuracy in practical applications.

To mitigate these issues, this study adopts a **Uniform Sampling** strategy, wherein 1,000 training pixels are evenly selected for each land cover class. This approach ensures that the model learns the spectral characteristics of all land cover types with equal attention. It effectively reduces class bias during training and improves the model's ability to distinguish among various land cover types, particularly enhancing recognition accuracy for minority classes such as trees and bare land.

In the land cover classification for the Shanghai area, training samples were selected based on multispectral image features and manual interpretation. Five representative land cover types were included: **Water, Trees, Crops, Built, and Bare**. A total of **5,000 training pixels** were collected—**1,000 pixels per class**.

Spatially, the training samples were distributed to cover representative areas across the study region, including the urban core, suburban agricultural zones, and areas with dense water networks. Samples of each class were extracted using multiple small polygons to avoid spatial autocorrelation issues caused by oversampling within a single plot. This is especially important in highly urbanized areas, where the spectral differences between built-up land and bare soil are often subtle. Sampling across diverse locations helps improve the model's ability to distinguish between spectrally similar classes.

By adopting the uniform sampling strategy, this study not only enhances the generalization ability of the classification model but also ensures robust recognition of marginal classes, resulting in a more balanced and equitable training dataset.

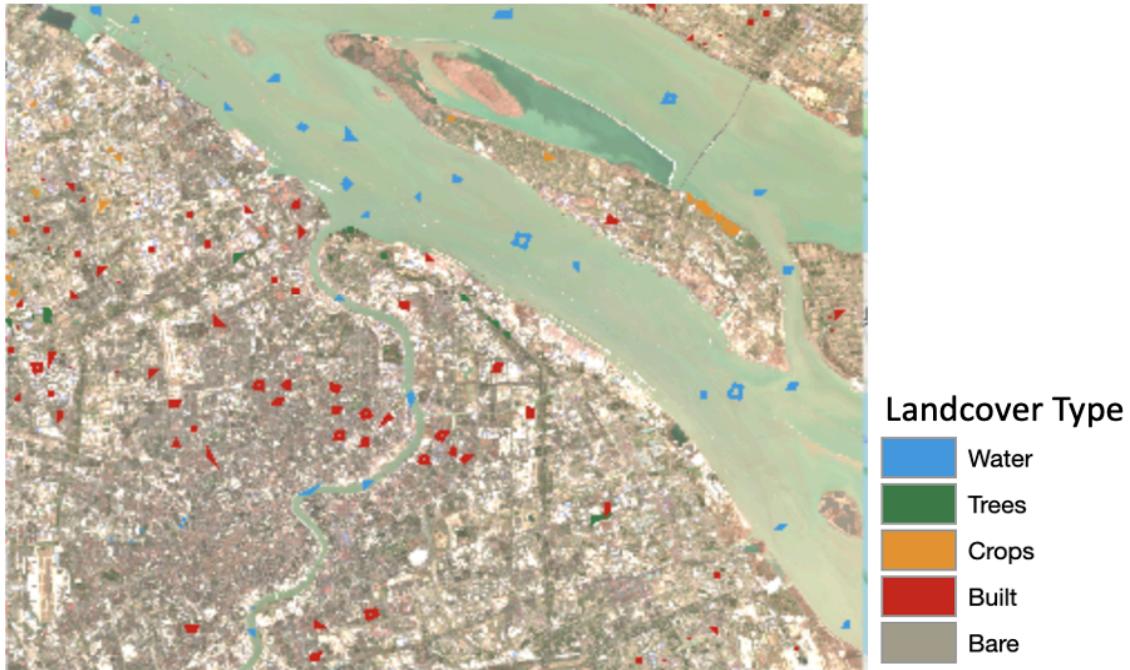


Figure 6. Shanghai Training Samples.

Type	Water	Trees	Crops	Built	Bare
Pixels	1,000	1,000	1,000	1,000	1,000

Table 11. Shanghai Training Samples.

3.2. Tuscany

3.2.1. Description

Tuscany is located in central Italy, on the west side of the Apennine Mountains, facing the Tyrrhenian Sea. Its landforms are dominated by hills and mountains, accounting for about 66% of the total area of the region. The plains are relatively small, mainly distributed along the coast and in some inland river valleys. This topographic pattern has created a diverse land use structure in Tuscany, forming a typical "mountain-hill-valley" transition zone landscape.

The region has a long history of agriculture and is one of Italy's important agricultural product production areas, especially famous for wine, olive oil, wheat, sunflowers and vegetable cultivation. Agricultural land occupies a significant proportion in Tuscany, widely distributed in the hills, low slopes and plains, forming a highly fragmented agricultural landscape with large fields, terraces, farmhouses and woodlands alternating.

This study selected agricultural land as the main observation object. On the one hand, agriculture is one of the most representative land use types in the region, with obvious seasonal changes and spatial structural characteristics; on the other hand, the

spatiotemporal changes of agricultural activities directly reflect the dynamic characteristics of surface coverage, which is suitable as an important basis for the evaluation and comparison of remote sensing classification models. At the same time, due to the large number of agricultural land types and high spectral similarity, the recognition ability of the classification algorithm is more challenging.

In terms of time, this study selected Sentinel-2 satellite images in April 2022 for analysis. This period is in the middle of the local spring, when crops enter the greening and rapid growth period. Most fields have been sown, and vegetation begins to show obvious spectral characteristics (such as higher NDVI values), but has not yet reached the high-density coverage state of summer. At this time, the farmland has good identifiability, which helps to distinguish different types of crops from non-arable land. At the same time, April is also a season with less cloud cover and moderate precipitation in central Italy, which is conducive to obtaining high-quality, cloud-free optical remote sensing images and improving the stability and accuracy of the classification model.

Therefore, as a typical region with complex terrain and prominent agriculture, conducting agricultural land classification research in spring is not only representative but also helpful to evaluate the adaptability and accuracy of different classification algorithms in complex terrain environments.



Figure 7. Map of Italy with Tuscany position.

Figure 8 illustrates the Digital Terrain Model (DTM) of the Tuscany study area, highlighting its complex elevation variations and topographic gradients. The DTM provides important contextual information for interpreting land-cover patterns in this agricultural environment.

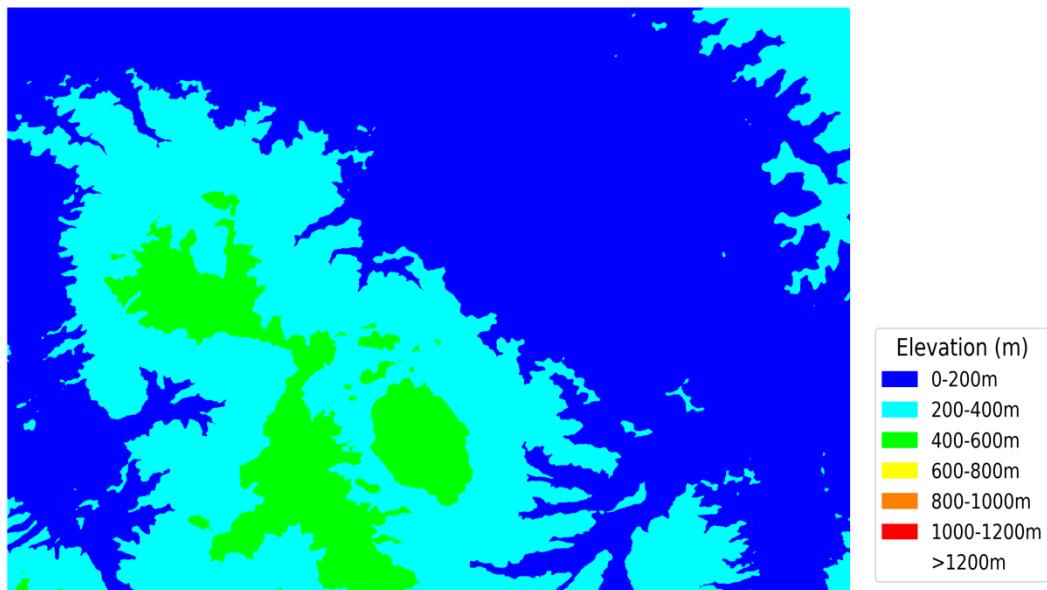


Figure 8. Tuscany DTM Image.

3.2.2. Original Image

The study area is located in the northwestern part of Tuscany, Italy, with grid number 32TPP, covering an area of approximately 25 km x 20 km, and the data collection time is 18th April, 2022. The area consists mainly of agricultural land, dominated by trees and crops, showing a typical agricultural landscape of the Tuscany region.

Image ID: 20220418T100601_20220418T101300_T32TPP.

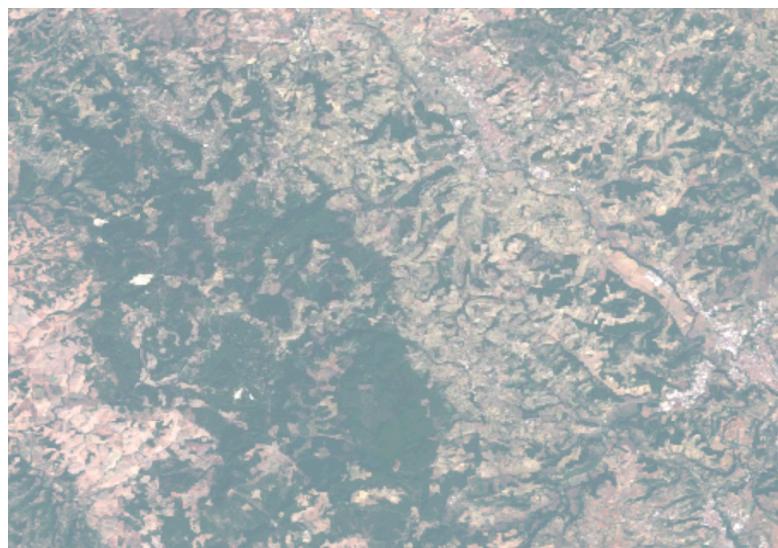


Figure 9. Tuscany Original Image.

3.2.3. Training Samples

In the land cover classification process in Tuscany, a uniform sampling strategy was adopted to ensure balanced model learning across different land-cover types. A total

of 5,000 training pixels were collected—1,000 pixels for each class, covering the five main land-cover types in the region: **Trees**, **Grass**, **Crops**, **Shrubs**, and **Built**.

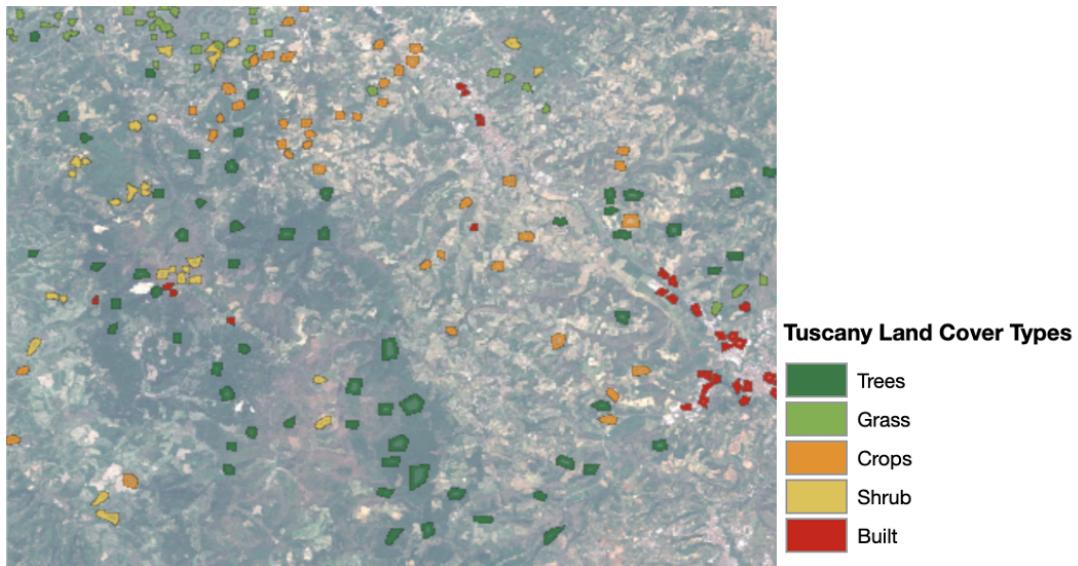


Figure 10. Tuscany Training Samples.

Type	Trees	Grass	Crops	Shrub	Built
Pixels	1,000	1,000	1,000	1,000	1,000

Table 12. Tuscany Training Samples.

3.3. Aosta

3.3.1. Description

The Aosta region is located in the northwest of Italy, in the core area of the Alps, near the borders of France and Switzerland. It is a typical alpine landform area, with elevations ranging from less than 600 meters in the valley to over 4,000 meters on snow-capped peaks. The valley is narrow and flanked by snow-covered mountains. The region features diverse landforms, including rivers, glaciers, bare rocks, coniferous forests, and meadows.

In terms of topographic distribution, Aosta exhibits a clear altitudinal zonation: low-altitude areas are dominated by woodland, grassland, and shrubs; mid-altitudes by bare land and shrubs; and high-altitudes by perennial snow and seasonal glaciers. Due to elevation, the vegetation growing season is short, limited to June–August, followed by an early snowfall.

1. This study selected Sentinel-2 imagery from early October, based on two considerations:

Early October marks the transition from late summer to autumn, when snow

begins to accumulate at high elevations and NDVI values decline, enhancing the spectral contrast between snow and other surfaces;

2. Compared to winter images, October scenes have better solar illumination and visibility, making it easier to identify the boundaries of snow and glaciers.

However, classification in alpine environments like Aosta faces multiple challenges: the mountain shadow effect and slope-induced illumination variation affect spectral response; confusion between bare land, shrubs, and snow-ice is common in visible and SWIR bands; and snow cover is spatially and temporally heterogeneous, leading to blurred and unstable class boundaries.

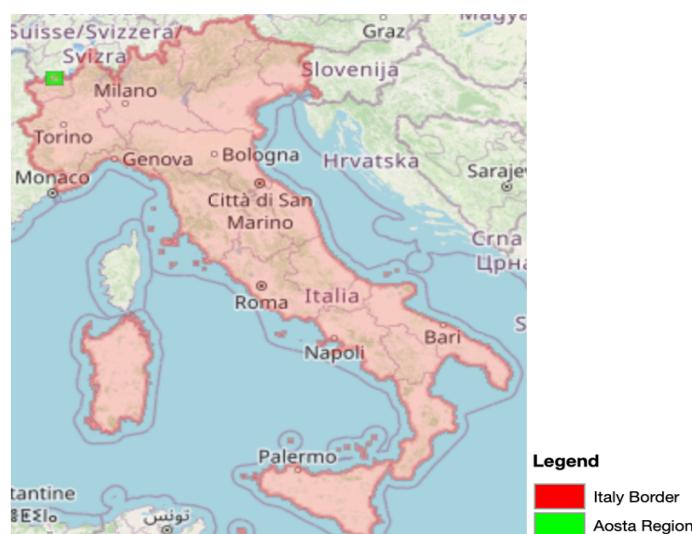


Figure 11. Map of Italy with Aosta position.

Figure 12 illustrates the Digital Terrain Model (DTM) of the Aosta study area, highlighting its complex elevation variations and topographic gradients.

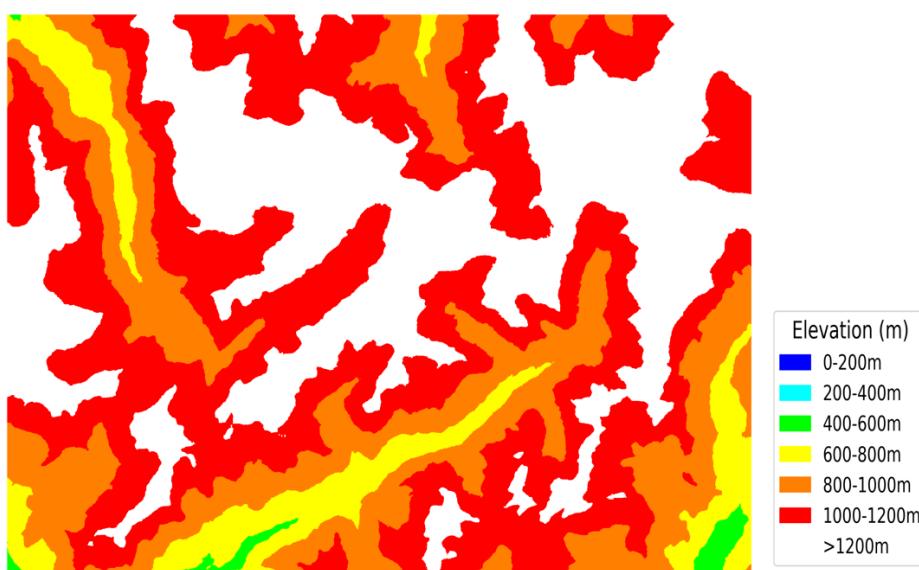


Figure 12. Aosta DTM Image.

3.3.2. Original Image

The study area is located in the northern part of the Aosta Mountains, Italy, with the grid number 32TLR, covering an area of approximately 25 km × 20 km. The data collection time is 6th October, 2022. Due to the high altitude and mountainous climatic characteristics, a lot of snow cover can be seen, and a small amount of bare ground can be observed more clearly.

Image ID: 20221006T102949_20221006T103655_T32TLR



Figure 13. Aosta Original Image.

3.3.3. Training Samples

In the land cover classification process in Aosta, a uniform sampling strategy was adopted to ensure balanced model learning across different land-cover types. A total of **6,000 training pixels** were collected—**1,000 pixels for each class**, covering the five main land-cover types in the region: **Water, Trees, Shrub, Grass, Bare, and Snow-ice**.

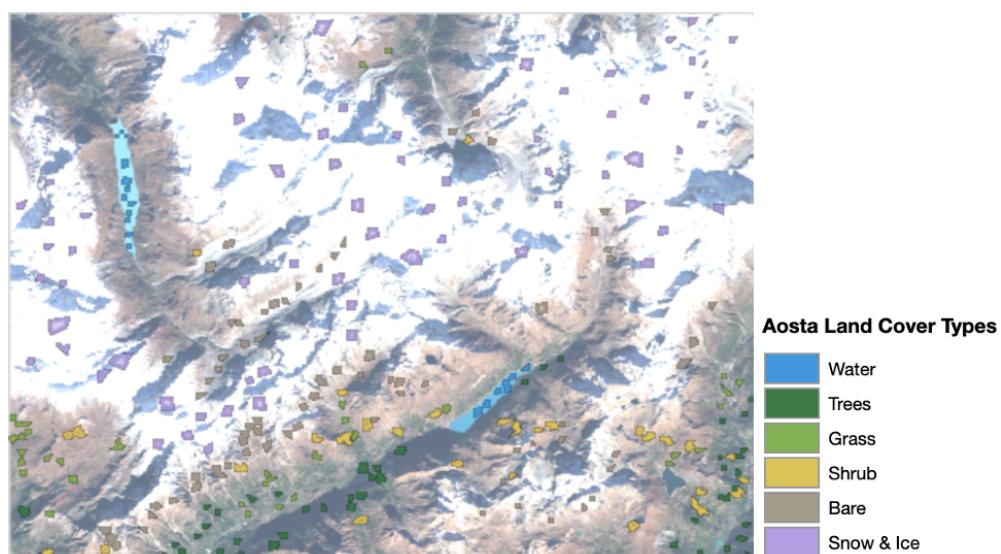


Figure 14. Aosta Training Samples.

Type	Water	Trees	Shrub	Grass	Bare	Snow-ice
Pixels	1,000	1,000	1,000	1,000	1,000	1,000

Table 13. Aosta Training Samples.

4 Data Processing

In remote sensing land cover classification, raw satellite images often have problems such as cloud interference, inconsistent band resolution, and redundant information. If they are not effectively processed, they will easily affect the subsequent feature extraction and classification accuracy. Therefore, a scientific and standardized data preprocessing process is the basic link of remote sensing classification analysis.

Based on the Google Earth Engine platform, this study carried out the following preprocessing work on the Sentinel-2 images of the study area: grid number determination, cloud and shadow mask, band selection, spectral index calculation, etc., to improve data quality and lay the foundation for subsequent feature extraction and modeling.

4.1.1. Grid Number Determination

When conducting remote sensing classification based on Sentinel-2 data, the spatial organization structure of the imagery directly impacts subsequent data processing and model training. The Sentinel-2 satellite system uses the Military Grid Reference System (MGRS) for imagery zoning, with each grid (Tile) covering an area of approximately 100 km × 100 km. Imagery data on the GEE is also stored and accessed according to this numbering system.

However, if the study area spans multiple tiles, there may be differences between images in terms of projection systems, imaging times, and image quality. Directly merging such images may result in pixel misalignment, spectral inconsistencies, or classification errors. Additionally, some image processing functions in GEE (such as image compositing, resampling, and classifier application) require input data to have spatial continuity by default. When input images span multiple tiles, this can lead to spatial alignment issues or processing failures.

Therefore, in the data preparation stage of this study, the study area was first matched and identified with the Sentinel-2 grid system. By loading the MGRS grid boundary layer and combining it with the GEE platform image metadata, the unique tile numbers corresponding to the Shanghai, Tuscany, and Aosta areas were clearly

identified. During subsequent image selection and retrieval, only areas fully contained within a single grid were selected as the analysis scope, and the data used was restricted to Sentinel-2 imagery corresponding to that tile, thereby eliminating processing interference caused by cross-grid stitching from the source.

By prioritizing the confirmation of grid numbers and the definition of spatial boundaries, this approach not only ensures that the study area has a consistent spatial reference system and complete image coverage on the GEE platform but also significantly simplifies processes such as image reprojection, spatial cropping, and sample annotation, thereby enhancing the stability and reproducibility of the classification system. This strategy provides a standardized and reliable data foundation for subsequent band selection, feature extraction, and model training.

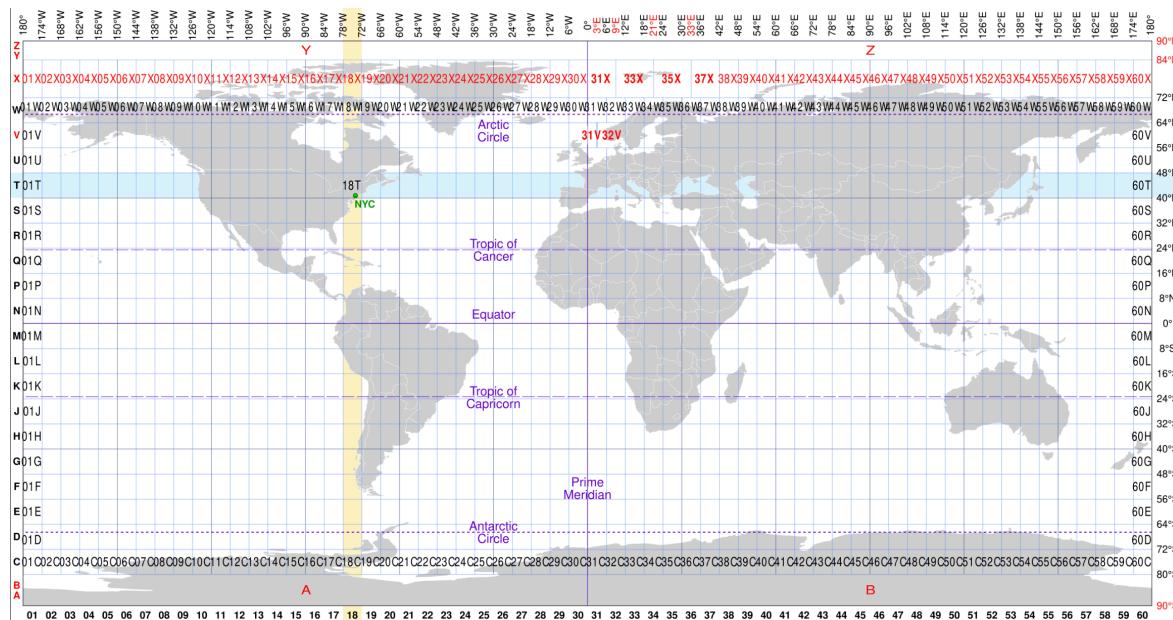


Figure 15. Military Grid Reference System.

4.1.2. Cloud and Shadow Mask

In remote sensing image classification tasks, clouds and their shadows often lead to abnormal pixel reflectivity, which seriously interferes with the identification of ground feature features and the training of classification models. Especially in winter and urban areas, high reflectivity of buildings or cloud shadows will increase classification uncertainty. Although the Sentinel-2 Level-2A (S2_SR) product has been atmospherically corrected and contains auxiliary mask information such as QA60 (cloud quality mask) and SCL (scene classification map) for automatic identification of clouds, shadows and invalid pixels, the existing automatic masking algorithm may still produce misjudgments or omissions at the edge of buildings, snowy surfaces or local thin cloud areas, thus affecting the purity of model training data.

In order to minimize the noise interference caused by cloud mask errors, this study did not directly use QA60 or SCL bands for masking operations, but adopted a

manually-led image screening strategy to strictly control cloud and shadow interference at the image selection source stage to ensure the spectral consistency of the input data and the reliability of ground feature expression. The strategy mainly includes the following steps:

1. Grid number determination:

First determine the Sentinel-2 grid number (Tile ID) covered by the study area.

2. Load candidate images:

Load the S2_SR image collection on the Google Earth Engine platform, and limit the time range to the target observation period to obtain all available image sequences of the Tile during the period.

3. Cloud coverage information extraction:

Each Sentinel-2 image is accompanied by the metadata field CLOUDY_PIXEL_PERCENTAGE, which can be directly used to quantify its overall cloud coverage. This study calls this parameter in GEE, outputs the cloud percentage of the candidate images and sorts them in ascending order, and preliminarily screens out data with obviously high cloud cover.

4. Manual visual inspection:

Based on the cloud cover sorting, the GEE visualization tool is further used to browse and compare the candidate images one by one, focusing on evaluating the cloud distribution position, shadow occlusion degree and ground object clarity in the study area, to ensure that the key features in the area are not blocked and the reflectivity is normal.

5. Final image determination:

Taking into account the cloud coverage, object clarity and shadow effects, a single high-quality cloud-free image is selected from multiple scene images as the classification analysis data source.

This image selection method based on cloud coverage information extraction and manual visual screening not only effectively avoids the classification uncertainty caused by automatic mask errors, but also provides a cleaner and unified input source for subsequent band resampling, spectral index calculation and feature extraction. This strategy is particularly suitable for small-scale, single-scene, multi-class urban object recognition tasks such as this study, and can lay a solid foundation for improving the performance of the classification model and the spatial consistency of the output results.

4.1.3. Band Selection

Considering that the land-cover classification scheme in this study encompasses Water, Built, Shrub, Trees, Grass, Crops, Bare, and Snow-ice, the selection of input spectral features was guided by both the spectral sensitivity of these classes and established findings from remote sensing research. Accordingly, ten representative bands from Sentinel-2 were adopted as classification features. This subset is designed to maximize

the separability of heterogeneous land-cover types while minimizing spectral redundancy, thereby ensuring computational efficiency without compromising classification accuracy.

- 10 m resolution bands: B2 (Blue), B3 (Green), B4 (Red), and B8 (NIR). These bands span the visible to near-infrared spectrum and serve as the fundamental basis for vegetation indices (e.g., NDVI), water indices (e.g., NDWI), and general land-cover discrimination.
- 20 m resolution bands: B5, B6, B7 (Red Edge), B8A (Narrow NIR), B11, and B12 (SWIR). The red-edge bands are particularly sensitive to chlorophyll content, canopy structure, and vegetation stress, enabling discrimination among diverse vegetation forms such as trees, shrubs, grasses, and crops. The SWIR bands, by contrast, are crucial for detecting soil and built-up features, as well as for distinguishing snow and ice from clouds or bare surfaces.

The choice of these bands is further substantiated by the distinct spectral responses of different land-cover classes:

- Water exhibits high reflectance in the blue (B2) and green (B3) bands, but very low reflectance in the NIR (B8) and SWIR (B11, B12), making it amenable to NDWI-based extraction.
- Built-up areas and bare soils reveal marked spectral contrasts in the red (B4), NIR (B8), and especially SWIR bands (B11, B12), where SWIR is highly effective in separating impervious surfaces from dry soils.
- Vegetation (trees, shrubs, grasslands, crops) is characterized by strong absorption in the red band (B4) and high reflectance in the NIR (B8), forming the basis of NDVI, while the red-edge bands (B5, B6, B7, B8A) provide enhanced sensitivity to canopy biochemical and structural properties, allowing for more refined discrimination among vegetation subtypes.
- Snow and ice display high reflectance in the visible range (B2, B3, B4) but very low reflectance in the SWIR (B11, B12), which provides a robust means of separating snow/ice from bare land or clouds.

In summary, the selected 10 bands integrate complementary information across the visible, red-edge, near-infrared, and shortwave-infrared regions, ensuring comprehensive coverage of the spectral signatures relevant to the eight land-cover categories under study. This balanced feature set not only enhances classification robustness but also prevents unnecessary redundancy, thereby achieving an optimal trade-off between accuracy and efficiency in multispectral land-cover mapping.

4.1.4. Spectral Index Computation

Spectral indices represent a commonly employed feature construction method in remote sensing image analysis. By combining multiple spectral bands, these indices effectively enhance the spectral characteristics of target objects, offering significant

advantages in improving the distinguishability between different land cover types. They are particularly well-suited for high-precision tasks such as land cover classification.

In this study, the initial step in feature construction involved cropping the study area from Sentinel-2 Level-2A data and selecting ten representative raw bands (selected_bands: B2, B3, B4, B5, B6, B7, B8, B8A, B11, B12) as foundational input features. These bands cover key spectral regions including visible light (RGB), red edge, near-infrared (NIR), and shortwave infrared (SWIR), enabling comprehensive representation of spectral response characteristics for primary land types within the study area (e.g., water, grass, trees, crops, bare, built, snow-ice).

Building upon this foundation, four representative spectral indices were further calculated to enhance recognition capabilities for distinct land features. Specifically, NDVI was employed to highlight vegetation features, NDWI enhanced water body information, NDBI was used to identify built-up areas, and BSI aided in extracting bare soil characteristics. These spectral indices were overlaid onto the original imagery as derived features, forming a composite image comprising 10 original bands and 4 spectral indices. The resulting feature set, all_bands, contained 14 dimensions in total.

This feature set not only ensures spectral diversity and stability but also provides a more discriminative input foundation for subsequent supervised classification and unsupervised clustering algorithms by introducing biophysically meaningful feature variables.

Their calculation formulas are as follows:

- NDVI (Normalized Difference Vegetation Index)

$$NDVI = \frac{B8 - B4}{B8 + B4} \quad (44)$$

Used to measure vegetation coverage and health, areas with high near-infrared reflectance and strong red light absorption generally indicate lush vegetation.

- NDWI (Normalized Difference Water Index)

$$NDWI = \frac{B3 - B8}{B3 + B8} \quad (45)$$

Enhance the ability to identify water bodies. The green band has strong reflection on water bodies, while the near infrared is absorbed by water bodies.

- NDBI (Normalized Difference Built-up Index)

$$NDBI = \frac{B11 - B8}{B11 + B8} \quad (46)$$

Enhance the spectral distinction between urban building areas and other landforms. The SWIR band has strong reflection on the building surface.

- BSI (Bare Soil Index)

$$BSI = \frac{(B11 + B4) - (B8 + B2)}{(B11 + B4) + (B8 + B2)} \quad (47)$$

It is used to highlight the exposed soil areas and comprehensively reflect the differences among SWIR, red light, near infrared and blue light.

4.1.5. Experimental Workflow

The experimental workflow of this study comprises five principal stages: data acquisition and preprocessing, sample collection and partitioning, model training and classification, result visualization, and accuracy evaluation.

Initially, three study areas characterized by distinct land cover patterns were selected: the highly urbanized metropolitan area of Shanghai (China), the agriculturally intensive Tuscany region (Italy), and the alpine environment of Aosta (Italy), characterized by snow cover and rugged topography. For each region, a single-date Sentinel-2 Level-2A surface reflectance product was acquired. Preprocessing procedures included spectral band selection, cloud masking, and the derivation of commonly used spectral indices such as NDVI, NDWI, and NDBI to enhance land cover separability.

During the sample preparation phase, training samples were manually delineated using polygonal annotations on the Google Earth Engine (GEE) platform and subsequently exported in CSV format. A uniform sampling strategy was employed, whereby 1,000 pixels were randomly selected for each land cover class to mitigate class imbalance. The annotated dataset was then stratified into training and validation subsets at a ratio of 7:3, ensuring an independent evaluation of model performance.

The classification phase involved the implementation of three representative algorithms: K-means clustering (unsupervised), Random Forest (supervised), and Multi-Layer Perceptron (deep learning). The former two were operationalized within the GEE Python API, while the MLP model was trained locally using the scikit-learn library in a Python environment. The resulting classification outputs were subsequently imported into GEE for spatial visualization.

Finally, the accuracy assessment was conducted using the independent validation set to construct confusion matrices and compute standard evaluation metrics, including overall accuracy, user's and producer's accuracy, and the Kappa coefficient. The classification maps generated from each model facilitated a spatially explicit comparison of land cover distributions, offering intuitive and quantitative insights into model performance across heterogeneous landscapes.

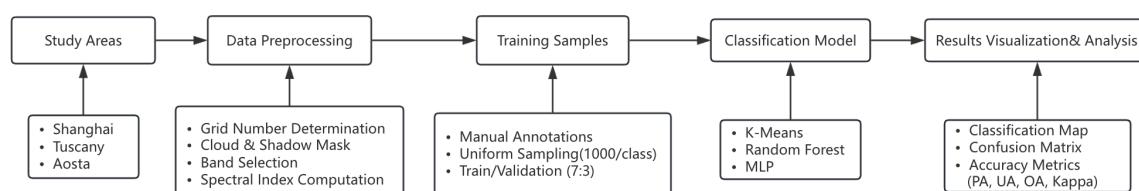


Figure 16. Experimental Workflow

5 Results Analysis

5.1. KMeans

5.1.1. Shanghai

5.1.1.1. Visual Interpretation of the Classification Map

Figure 16 shows the classification results of the KMeans clustering algorithm for the Shanghai region. Overall, these feature types exhibit a highly structured spatial distribution. The Built category is the most prominent and dominant category, widely distributed in the lower left region of the image. It forms a large, continuous patch on the west bank of the Yangtze River, corresponding to the densely populated buildings (including residential and commercial buildings) in downtown Shanghai, as well as hard surfaces such as urban roads. This demonstrates good clustering performance at the urban scale. The crops category is primarily concentrated in the eastern and upper left corners of the image, where large expanses of farmland form. The water category accurately reconstructs the main channel of the Yangtze River and its tributaries, with smooth boundaries, indicating that the spectral characteristics of water bodies are highly consistent across the cluster dimension. In contrast, the trees and bare land categories exhibit a more fragmented distribution pattern on the map: tree patches are limited and discrete. Bare land is primarily distributed in the buffer zone between built-up areas and crop areas, with a scattered distribution and blurred boundaries.

The water and building categories perform best in terms of clustering. The water category, due to its low near-infrared reflectance and stable spectral morphology, has clear boundaries and is easily identifiable in the classification map. The building category exhibits high-density clustering with strong spatial consistency, tending to cluster around central hubs. However, the classification results for the trees category are less than ideal, likely due to the low proportion of urban green space, the small trees canopy area, and interference from other categories such as buildings, which results in a fragmented and disjointed spatial distribution. Furthermore, in transitional areas with complex urban terrain, some category confusion is evident, particularly with blurred boundaries between built and bare land.



Figure 17. Shanghai KMeans.

5.1.1.2. Quantitative Accuracy Assessment and Discussion

As shown in the classification results, the Water class achieved the highest classification performance, with both the Producer's Accuracy (PA) and User's Accuracy (UA) reaching 1.00. This indicates that all validation pixels labeled as water were correctly classified, with no commission or omission errors. This exceptional result can be attributed to the distinct spectral characteristics of water bodies, particularly their strong absorption in the visible and near-infrared bands, which allows them to form well-separated and compact clusters in the spectral feature space. The Crops and Trees classes also demonstrated high classification accuracy, with PA and UA values of 0.93 / 0.90 and 0.86 / 0.95, respectively. These results suggest that the model is capable of reliably identifying these land cover types. According to the confusion matrix, a small number of crop pixels were misclassified as bare (10 pixels), trees (8 pixels), and built (3 pixels). Similarly, misclassification of tree pixels was more prominent, with 36 pixels confused as built, and an additional 3 pixels each confused as crops and bare, likely due to spectral mixing in densely built urban environments where green spaces are often interwoven with buildings.

By contrast, the Built class exhibited the lowest classification performance, with a PA of 0.79 and a UA of only 0.72, reflecting considerable confusion near the boundaries between built-up and non-built areas. Specifically, built pixels were frequently misclassified as crops (22 pixels), bare (30 pixels), and trees (7 pixels). This issue is particularly evident in urban fringe zones and construction sites, where heterogeneous surface materials and subtle spectral differences increase clustering uncertainty. The

Bare class showed moderate classification performance, with a PA of 0.83 and a UA of 0.86. Most misclassifications occurred at the interface with built areas, where 50 bare pixels were confused with built, and an additional 5 pixels were misclassified as crops. These findings indicate that spectral overlap between bare and other classes—particularly in transitional zones such as demolition sites or unpaved land—is a persistent challenge in unsupervised classification.

Overall, the KMeans clustering achieved an Overall Accuracy of 0.88 and a Kappa coefficient of 0.85, indicating that the results have good reliability and consistency. The model performed particularly well in distinguishing spectrally distinct classes such as water and crops, validating the effectiveness of the selected input features and band combinations. Nonetheless, the classification map still contains scattered noise and fuzzy class boundaries, especially in mixed or transitional zones between built, bare, and trees surfaces.

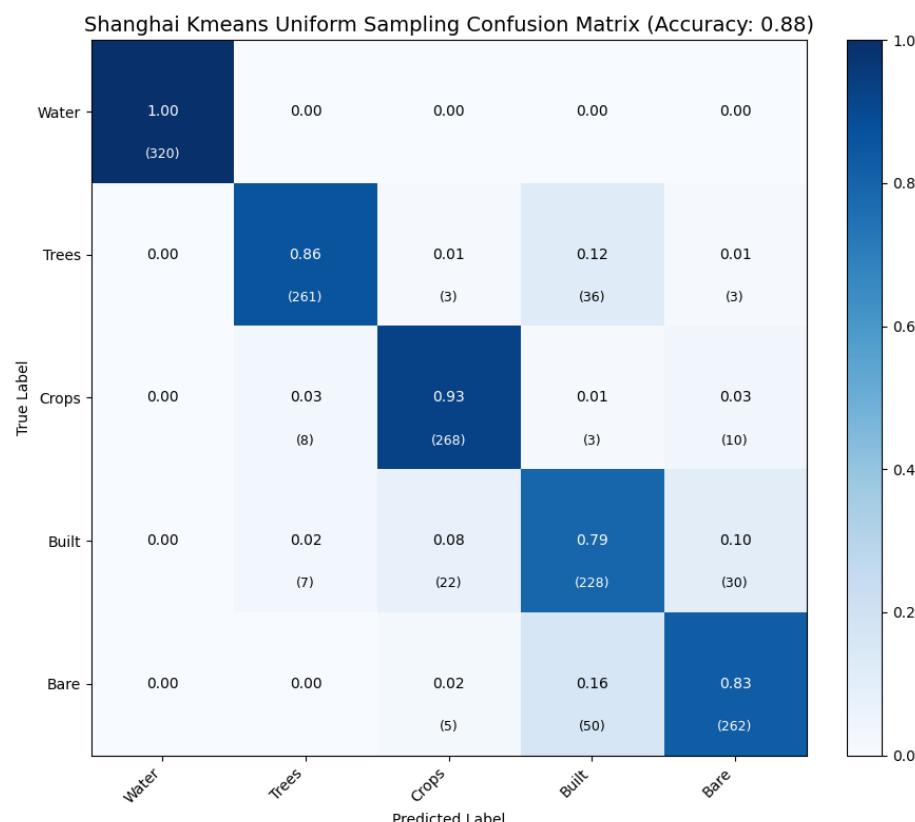


Figure 18. Shanghai KMeans Confusion Matrix.

Class	PA	UA
Water	1.00	1.00
Trees	0.86	0.95
Crops	0.93	0.90
Built	0.79	0.72
Bare	0.83	0.86

Overall Accuracy	0.88
Kappa Coefficient	0.85

Table 14. Shanghai Accuracy Report.

5.1.2. Tuscany

5.1.2.1. Visual Interpretation of the Classification Map

Figure 18 shows the results of land cover classification for Tuscany, Italy, using the KMeans unsupervised clustering algorithm. In this experiment, the number of clusters was set to five, representing five feature types: trees, grass, crops, shrubs, and built. The visualization depicts an overall hilly agricultural landscape, with trees, grass, crops, shrubs, and built interwoven. The classification map reflects the dominant spatial distribution of each feature. Trees are widely distributed in the central-western hills of the image. Grass and trees exhibit a clear spatial continuity. Crops features are primarily concentrated in the central and eastern parts of the image, with an irregular patch structure typical of Mediterranean crop cultivation areas (such as olives and grapes). Shrub features are primarily distributed at the junction of trees, grass, and crops, forming scattered, irregular patches. Building features are relatively sparsely distributed, appearing sporadically in the eastern and northern parts of the image. They are often nested between crops and grass, suggesting rural settlements, roads, or small-scale agricultural infrastructure.

Although some spatial structural features can be visually discerned, the overall classification performance of KMeans clustering in Tuscany is limited. First, there is significant spectral overlap between trees and grass, leading to confusion in some areas, particularly at tree edges and intersecting grass, where class boundaries are unclear. Second, the boundaries between crops, grass, and shrubs are ambiguous, leading to frequent cross-classification errors on hilly slopes or in mixed shrub-grass areas. This reflects the KMeans algorithm's limited ability to distinguish intermediate transition features. Furthermore, due to the high degree of discreteness and spatial patchiness of architectural features, they are easily misclassified due to interference from surrounding grass, shrub, and crop pixels. This results in low recognition accuracy and weak boundary representation for this class in the classification map.

Overall, these results demonstrate the KMeans clustering algorithm's ability to capture the coarse-grained structure of natural feature distributions under unsupervised conditions. However, the algorithm still has significant limitations in distinguishing fine-grained features, identifying mixed pixel boundaries, and maintaining patch-level stability.

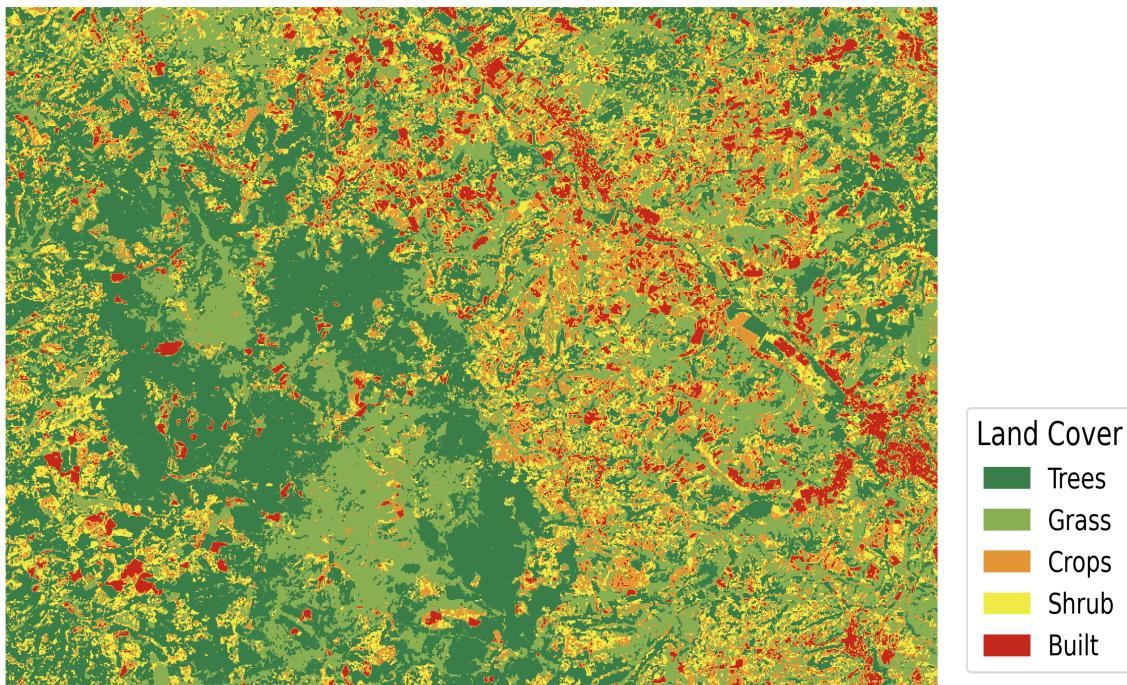


Figure 19. Tuscany KMeans.

5.1.2.2. Quantitative Accuracy Assessment and Discussion

As shown in the Accuracy Report, the Trees category demonstrated the best classification performance, with a producer accuracy (PA) of 0.95 and a user accuracy (UA) of 0.83, indicating that the vast majority of true trees pixels were accurately identified in the validation samples, and that most of the pixels labeled as trees in the classification map were indeed trees. This result is attributed to the stable and high reflectance characteristics of trees land in the red and near-infrared bands, enabling the formation of clear spectral boundaries in unsupervised clustering, thereby enhancing the clustering model's discriminative ability for this class. The PA for the Grass category is 0.90, indicating that the majority of true grass pixels were detected; but its UA is only 0.54, reflecting a significant "overclassification" issue for this category. According to the confusion matrix, among the pixels labeled as Grass, in addition to 272 true grass, there were also incorrect inclusions of Shrub (144 pixels), Crops (35 pixels), Built (50 pixels), and Trees (2 pixels). Such confusion primarily occurs in areas with mixed terrain types, such as hilly slopes and rural edges, indicating that grass and other low-lying vegetation exhibit high spectral similarity, leading to blurred cluster boundaries. The classification performance for Crops is at an intermediate level, with a PA of 0.58 and a UA of 0.68. Among the 288 validation samples, only 166 were correctly identified, with the remainder primarily misclassified as Trees (31 pixels), Grass (35 pixels), Built (22 pixels), and Shrub (34 pixels). This phenomenon is relatively common in Mediterranean hilly agricultural landscapes, indicating that crops and other vegetation categories overlap to some extent in the spectral dimension, and seasonal changes in April may also exacerbate this confusion.

Shrub classification accuracy was the lowest, with a PA of 0.33 and a UA of 0.50, with only 101 out of 302 validation samples accurately identified. Most shrub pixels were misclassified as Grass (144 pixels), Crops (46 pixels), Built (10 pixels), and Trees (1 pixel), revealing that this category not only has blurred spectral characteristics but also exhibits highly fragmented and irregular patchy spatial distribution, making it difficult for unsupervised models to capture its independent features. The Built category exhibited typical “low PA, high UA” characteristics, with PA at 0.54 and UA as high as 0.83, indicating that most pixels labeled as built were actual built, but a significant number of actual built pixels were not identified. According to the confusion matrix, these built pixels were primarily misclassified as Grass (50 pixels), Shrub (39 pixels), Crops (29 pixels), and Trees (17 pixels). These misclassifications were concentrated in crops edges or low-density building areas.

Overall, the KMeans clustering results for the Tuscany region achieved an overall classification accuracy of 0.66 and a Kappa coefficient of 0.57, indicating that the model has a certain ability to distinguish landforms without relying on supervised information, particularly for categories with significant spectral differences (such as Trees). However, given the region's significant topographic variations, complex intertwined land cover types, and strong spectral overlap among certain categories, the model exhibits notable shortcomings in boundary identification and intermediate category differentiation. The Kappa coefficient further suggests that part of the classification agreement may result from chance rather than true model performance, reflecting limited robustness of the clustering in spectrally ambiguous areas.

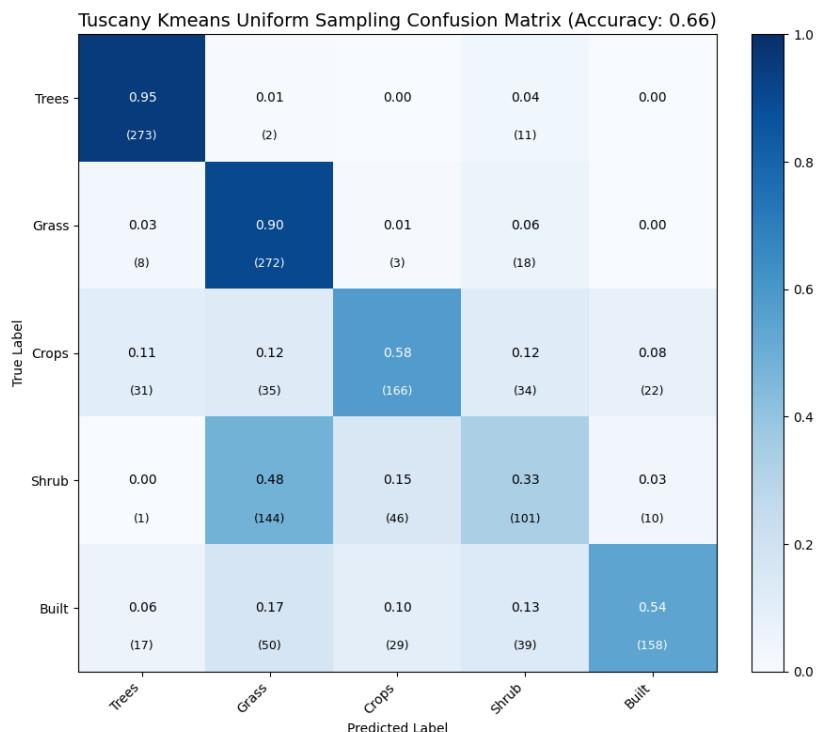


Figure 20. Tuscany KMeans Confusion Matrix.

Class	PA	UA
Trees	0.95	0.83
Grass	0.90	0.54
Crops	0.58	0.68
Shrub	0.33	0.50
Built	0.54	0.83
Overall Accuracy	0.66	
Kappa Coefficient	0.57	

Table 15. Tuscany Accuracy Report.

5.1.3. Aosta

5.1.3.1. Visual Interpretation of the Classification Map

In the KMeans clustering classification map of the Aosta region, Snow-ice is widely distributed in the northern and central parts of the image, forming large areas of continuous coverage, reflecting the seasonal phenomenon of permanent snow and glacier areas in the Alps mountain that have not yet melted by early October. Immediately below the Snow-ice layer, Bare areas exhibit continuous or patchy distributions, primarily concentrated on ridges, slopes, or bare rock zones on plateaus, closely adjacent to the snow-ice boundary. These represent natural vegetation-free zones where vegetation growth is restricted by high-altitude cold and shallow soil conditions. Transitioning downward, Shrub is primarily distributed between Bare and Grass, forming transitional patches with a fragmented mosaic distribution, concentrated in alpine meadows, transitional slopes, and sparsely vegetated areas on sunny slopes. Further extending to lower elevations, grass are widely distributed along the middle and lower slopes, intermingling with shrubs and trees in a mosaic pattern, particularly dense in the southern and western valleys of the image. Trees are primarily distributed in the southern and northwestern regions at medium to low elevations.

The unsupervised KMeans clustering applied to the Aosta Valley region effectively revealed distinct ecological zoning patterns, exhibiting a typical gradient sequence of “Snow-ice → Bare → Shrub → Grass → Trees.” This distribution generally aligns with regional topography and the natural seasonal state observed in early October. In the classification map, the boundary between snow-ice and bare areas is distinct, demonstrating the model's strong discrimination capability in regions with high reflectance and significant spectral differences. However, in intermediate transition zones—particularly between low-canopy vegetation types like shrubs, grasses, and trees—category boundaries appear blurred with high overlap, indicating the model's

limited ability to distinguish between spectrally similar classes. For instance, the boundary between grass and trees exhibits a relatively gradual transition. This may be attributed to the fact that understory herbaceous vegetation had not yet fully withered during the study period, reducing spectral response differences between grass and trees in the near-infrared band and thereby increasing classification difficulty.

Additionally, the classification map shows some snow-ice areas incorrectly identified as water, causing the spatial distribution of the water class to extend beyond actual river channels and standing water bodies. This phenomenon may relate to the strong reflectance characteristics of high-altitude snow-ice areas in blue and green light bands under specific observation angles, leading to their misclassification as water bodies with similar spectral signatures during unsupervised clustering. Although such misclassifications remain low in overall proportion, they indicate that the KMeans model, operating without training sample guidance, is susceptible to interference from spectral similarities between classes. This vulnerability is particularly pronounced in alpine environments where snow, water bodies, and ice surfaces are intermingled. Therefore, while this model demonstrates certain advantages in reconstructing macro-ecological zonation structures and is particularly suitable for tasks involving distinct major categories, it still exhibits significant limitations in handling edge-blended features, small-patch vegetation, and the subdivision of highly reflective classes.

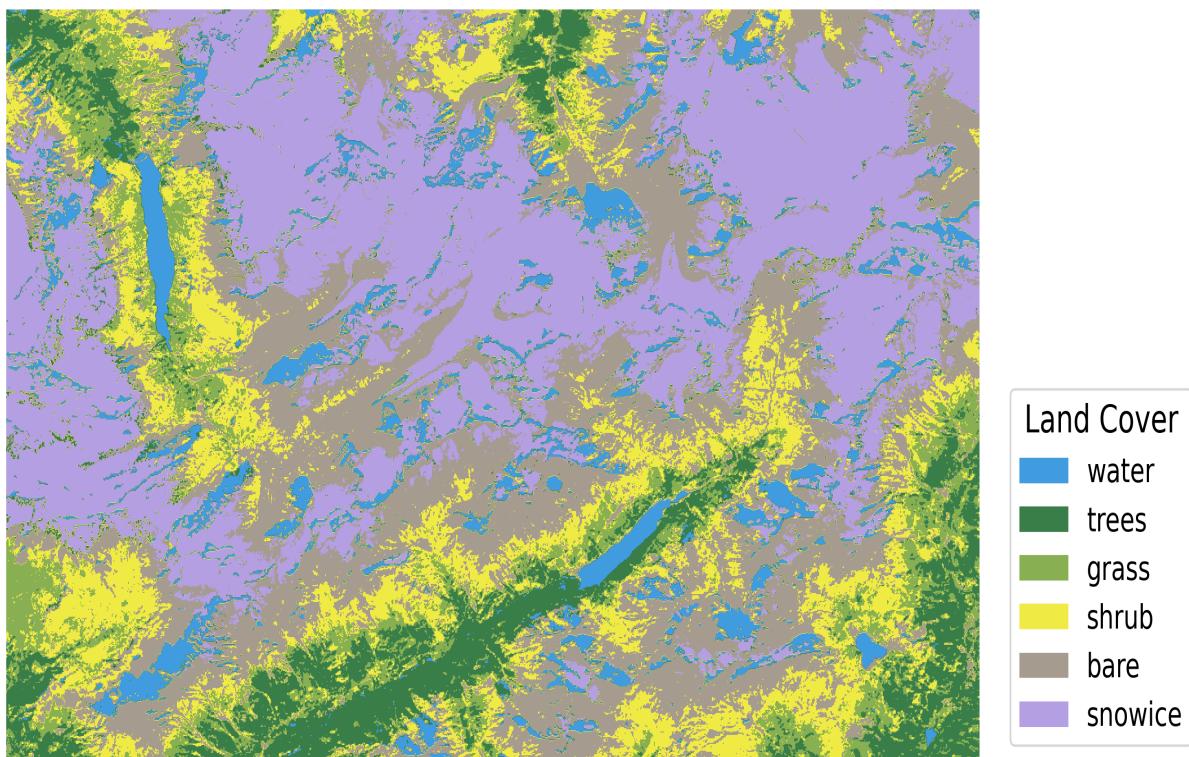


Figure 21. Aosta KMeans.

5.1.3.2. Quantitative Accuracy Assessment and Discussion

The KMeans clustering classification results for the Aosta region demonstrate high classification performance, with an Overall Accuracy (OA) of 0.92 and a Kappa coefficient of 0.90, indicating that the model still possesses significant feature recognition capabilities under unsupervised conditions without relying on prior training samples. The Kappa value further confirms that the classification results exhibit “nearly perfect” statistical consistency with the actual features, effectively eliminating interference from random consistency and enhancing the reliability of accuracy assessments.

Among all categories, Water performs most ideally, with a producer accuracy (PA) of 1.00, indicating that all true water bodies in the validation samples were accurately identified; and the user accuracy (UA) was 0.96, indicating that nearly all pixels classified as water in the classification map were actual water bodies. The Trees category also demonstrated outstanding classification performance, with a PA of 0.97 and a UA as high as 0.99, reflecting its prominent spectral performance in the near-infrared and red-edge bands, making it easy to form cluster centers with clear cluster boundaries. The PA for the Snow-ice category is 0.90, and the UA is 1.00. Its high reflectance characteristics result in stable clustering outcomes, but some pixels are still incorrectly classified as Water, Trees, and Bare, indicating that there is some spectral overlap between high-reflectance land cover types.

In contrast, some intermediate categories exhibit more pronounced confusion. The primary misclassification direction for the Grass category is Shrub, with 70 grass pixels incorrectly classified as shrub, indicating a certain degree of spectral or spatial distribution overlap between the two. Although the Shrub category has a high producer accuracy (PA = 0.97), indicating that the model can effectively identify true shrub pixels, its user accuracy is only 0.75, indicating that pixels classified as shrub still contain a significant number of large grass samples, with the primary source of error being the aforementioned 70 Grass pixels. This further indicates that in complex mountainous terrain, the spectral characteristics of shrubs are easily confused with those of grass, leading to overclassification issues. The Bare category performed relatively robustly, with a PA of 0.93 and a UA of 0.94. Misclassification primarily occurred in the snowline transition zone, where there was some overlap with Snow-ice and Shrub.

In summary, KMeans clustering performs exceptionally well in the Aosta region, particularly achieving classification accuracy comparable to supervised learning in categories with distinct spectral characteristics (such as Water, Trees, and Snow-ice). However, for categories with blurred spectral boundaries or high inter-class mixing (such as Grass and Shrub), there are still certain identification limitations, especially in areas with rugged terrain and complex, intertwined landforms. The model's performance in boundary identification and inter-class separation still needs

improvement. This further highlights the application boundaries and challenges of unsupervised methods in complex mountainous environments.

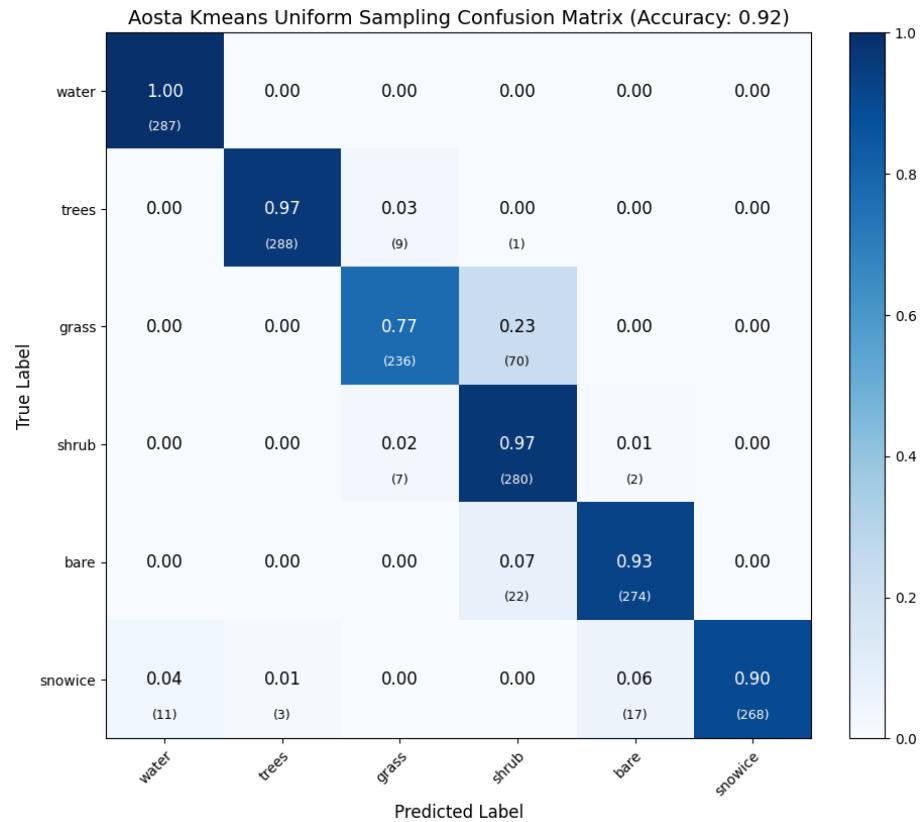


Figure 22. Aosta KMeans Confusion Matrix.

Class	PA	UA
Water	1.00	0.96
Trees	0.97	0.99
Grass	0.77	0.94
Shrub	0.97	0.75
Bare	0.93	0.94
Snow-ice	0.90	1.00
Overall Accuracy	0.92	
Kappa Coefficient	0.90	

Table 16. Aosta Accuracy Report.

5.2. Random Forest

5.2.1. Shanghai

5.2.1.1. Visual Interpretation of the Classification Map

Compared to the KMeans results, the Random Forest classifier produced a clearer and more structured classification map for the Shanghai region. The Water category was delineated with near-perfect consistency, accurately outlining the Yangtze River and its tributaries. Meanwhile, the Built-up areas in the urban core exhibited significantly improved spatial continuity and coherence, reflecting the classifier's enhanced ability to capture densely developed regions.

Also, Random Forest mitigated some of the confusion observed with KMeans between the adjacent "built" and "bare" classes. However, challenges persisted in transitional regions. Crops were occasionally misclassified as "bare" or "built," especially during winter, when fields lie fallow and their spectral signatures resemble those of impervious or bare surfaces. Similarly, small, scattered patches of trees were sometimes underestimated and confused with crops or built-up areas, reflecting the difficulty of capturing sparse vegetation in highly heterogeneous urban environments.

Overall, RF achieved more stable and reliable spatial structure than KMeans, particularly for the primary classes, although classification challenges remained for intermediate and fragmented features.

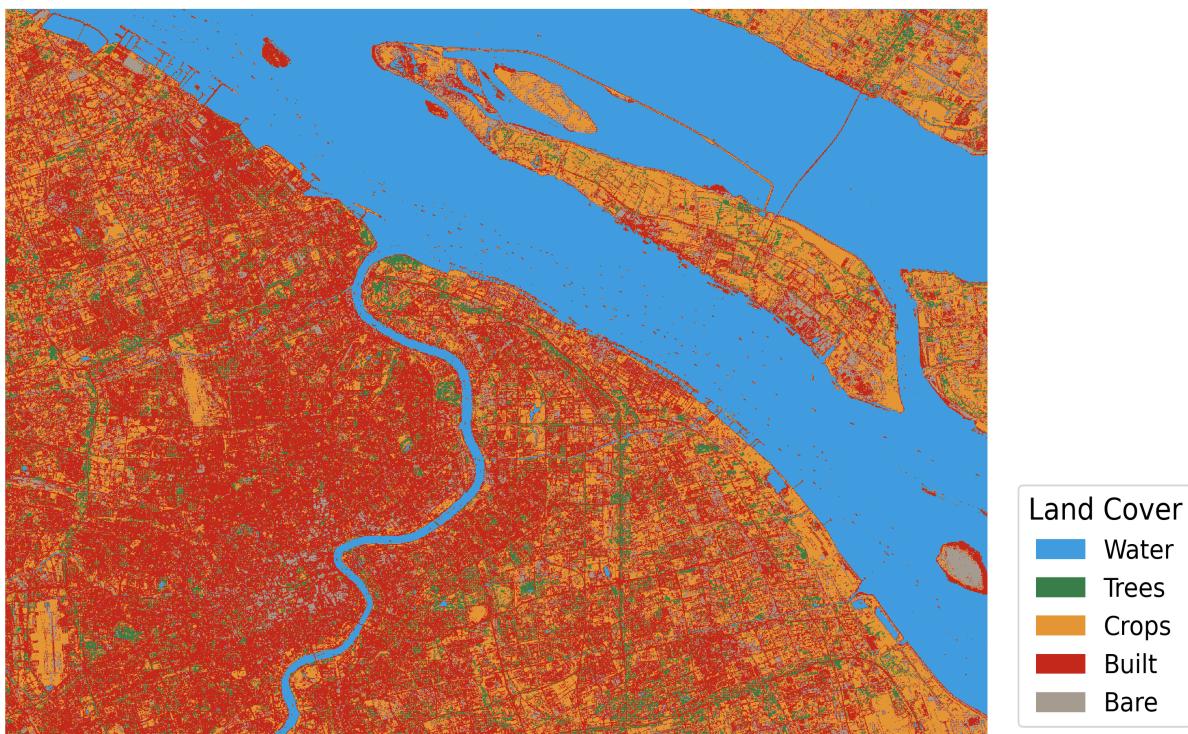


Figure 23. Shanghai RF.

5.2.1.2. Quantitative Accuracy Assessment and Discussion

Compared with KMeans (OA = 0.88, Kappa = 0.85), the Random Forest classifier achieved significantly higher overall classification performance in the Shanghai region, reaching an Overall Accuracy of 0.95 and a Kappa coefficient of 0.94. These results highlight the model's enhanced capability to capture complex urban surface features with better reliability and generalization.

The most notable improvement lies in the Built and Bare classes. The Built category, previously the worst-performing class in KMeans (PA = 0.79, UA = 0.72), improved substantially to PA = 0.90 and UA = 0.89, indicating a clear reduction in confusion with adjacent classes such as Crops and Bare. Similarly, Bare increased from PA = 0.83 , UA = 0.86 in KMeans to PA = 0.98, UA = 0.96. This reflects RF's better handling of spectral overlap in transitional areas such as construction sites and undeveloped land. Water remained perfectly classified in both methods (PA = UA = 1.00), confirming its strong spectral separability. Crops also benefited from improved control of commission and omission errors, with PA/UA rising slightly to 0.96 / 0.93 (vs. 0.93 / 0.90 in KMeans). For the Trees category, although the UA remained high (0.96), the PA improved only modestly from 0.86 to 0.90. Misclassification into Built persists, particularly in areas where vegetation and low-density structures are spatially mixed, this problem also exists in KMeans, but is better controlled in Random Forest.

The confusion matrix from RF shows a dominant diagonal structure with misclassified pixels sparsely, unlike the more clustered errors seen in KMeans. This pattern indicates that Random Forest handles high-dimensional feature interactions more effectively, especially in heterogeneous urban environments where traditional clustering struggles.

In summary, Random Forest outperforms KMeans across nearly all classes, especially in urban core and transitional zones. The improvements in Built and Bare highlight the advantages of supervised learning in resolving boundary ambiguity and reducing class confusion, confirming RF's suitability for urban land cover mapping.

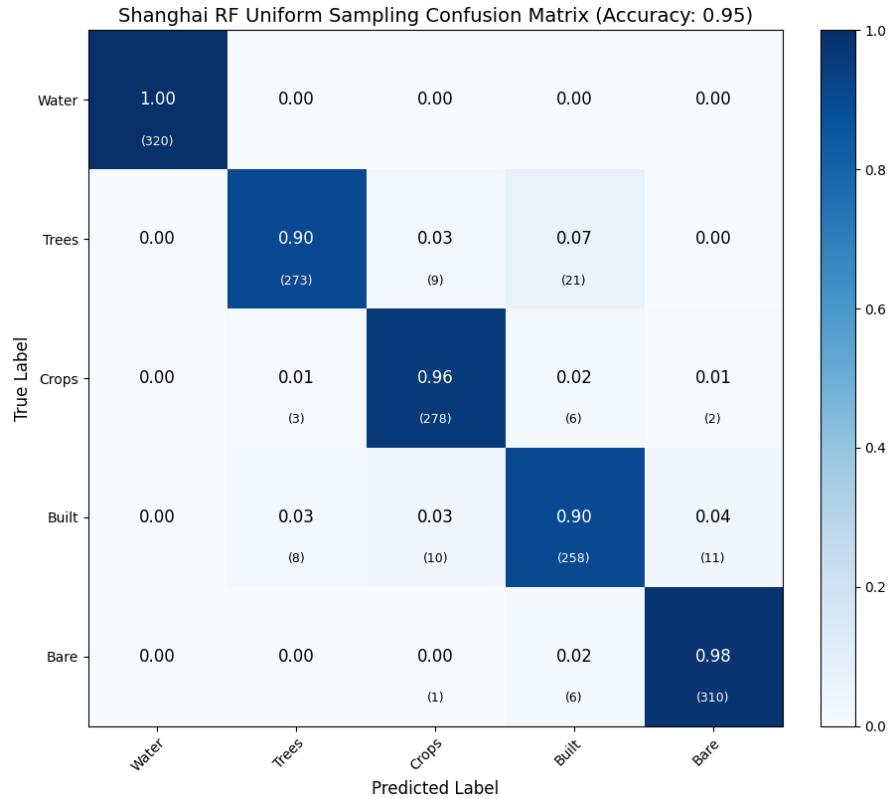


Figure 24. Shanghai RF Confusion Matrix.

Class	PA	UA
Water	1.00	1.00
Trees	0.90	0.96
Crops	0.96	0.93
Built	0.90	0.89
Bare	0.98	0.96
Overall Accuracy	0.95	
Kappa Coefficient	0.94	

Table 17. Shanghai Accuracy Report.

5.2.2. Tuscany

5.2.2.1. Visual Interpretation of the Classification Map

Compared with the KMeans results, the Random Forest classifier significantly enhances the spatial clarity and thematic consistency of the Tuscany classification map. Notably, Tree stands appear more coherent and compact, and Built-up areas are more accurately identified, with reduced omission in rural settlement zones.

One of the major limitations of KMeans was the extensive confusion between Grass, Crops, and Shrub, especially on sloped or mixed terrain. RF substantially mitigates this issue, with Grass and Shrub boundaries becoming more distinct, and fewer instances of overclassification between these low-lying vegetation types. Nevertheless, some residual ambiguity remains, particularly in transitional zones affected by spectral overlap, such as uncultivated plots or secondary vegetation growth.

Overall, the RF classification map captures the dominant structure of Tuscany's hilly agricultural mosaic more faithfully than KMeans, showing improved delineation of both spectrally distinct (Trees, Built) and spectrally similar (Grass, Shrub, Crops) land cover types.

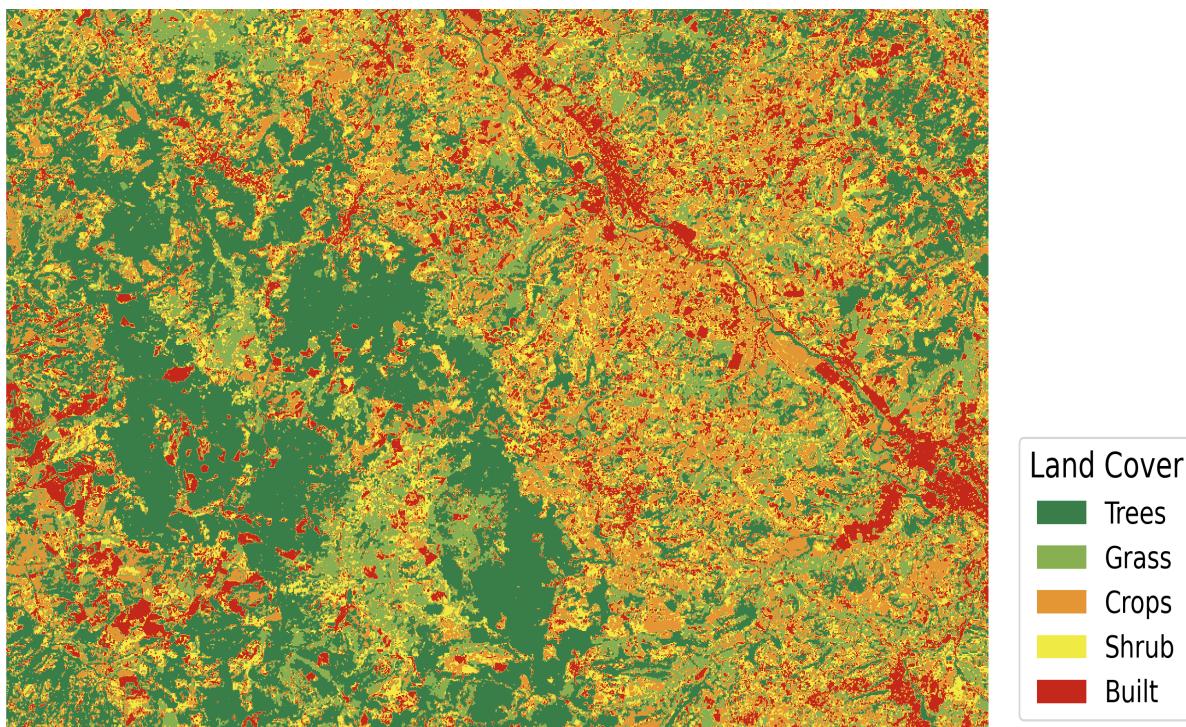


Figure 25. Tuscany RF.

5.2.2.2. Quantitative Accuracy Assessment and Discussion

Compared with the KMeans results ($OA = 0.66$, $Kappa = 0.57$), the Random Forest classifier significantly improved classification performance in the Tuscany region, achieving an Overall Accuracy of 0.86 and a Kappa coefficient of 0.82. This improvement reflects a substantial enhancement in class separability, particularly among spectrally similar vegetation types, and demonstrates the effectiveness of supervised learning in complex Mediterranean agricultural landscapes.

The Trees category, already the best-performing class in KMeans ($PA = 0.95$, $UA = 0.83$), saw further improvement to $PA = 0.99$, $UA = 0.96$, indicating that RF nearly eliminated classification errors for this category. Similarly, Grass improved significantly, especially in user accuracy (from 0.54 to 0.87), resolving much of the overclassification

issue that previously caused confusion with Shrub. Crops also saw clear gains (PA from 0.58 to 0.84, UA from 0.68 to 0.83). This suggests that the model more accurately distinguished crop fields, despite spectral overlap and phenological variability in April. Notably, Shrub, which had the weakest performance in KMeans (PA = 0.33, UA = 0.50), improved to PA = 0.74, UA = 0.78, though it remains the most ambiguous class. The improvements are meaningful but limited, primarily due to Shrub's inherently fragmented spatial distribution and spectral confusion with both Grass and Crops. Built also improved moderately, from PA = 0.54 to 0.78 and with UA remaining stable (0.82 in RF vs. 0.83 in KMeans), indicating better detection of rural built but residual confusion in low-density settlements.

The confusion matrix reveals a more concentrated diagonal structure in RF compared to KMeans, with substantially reduced cross-class errors, particularly among mid-vegetation classes. However, some boundary ambiguities persist in areas with heterogeneous terrain or land use mixing.

In summary, Random Forest substantially outperforms KMeans in Tuscany by improving the classification of both spectrally distinct and overlapping categories. The most notable gains are observed in Grass, Crops, Built and Shrub, previously problematic in the unsupervised results. While classification accuracy for Built and Shrub still requires refinement, the RF model demonstrates robust and balanced performance across classes under challenging agricultural conditions.

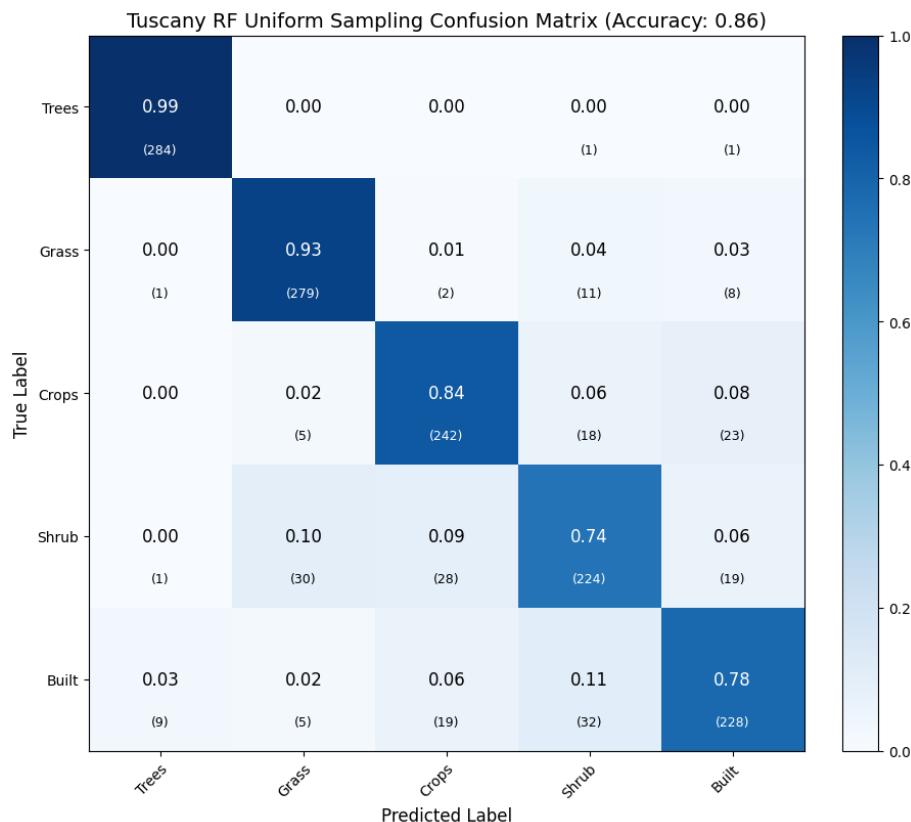


Figure 26. Tuscany RF Confusion Matrix.

Class	PA	UA
Trees	0.99	0.96
Grass	0.93	0.87
Crops	0.84	0.83
Shrub	0.74	0.78
Built	0.78	0.82
Overall Accuracy		0.86
Kappa Coefficient		0.82

Table 18. Tuscany Accuracy Report.

5.2.3. Aosta

5.2.3.1. Visual Interpretation of the Classification Map

Compared with the KMeans classification, the Random Forest classifier produces a more refined and consistent land cover map for the Aosta region, with notable improvements in boundary clarity and class separability, particularly across key elevation zones. The ecological stratification — from Snow-ice to Bare to Shrub, Grass, and Trees — remains clearly visible, but RF further enhances the definition and compactness of major classes.

The Snow-ice category, which was already well delineated in KMeans, becomes even more regular and distinct, with minimal confusion into Water. In contrast, KMeans had mistakenly extended Water into high-reflectance snow areas. Water bodies in RF are confined to their actual locations, showing narrow, elongated forms that align well with terrain. Tree clusters, particularly along river valleys and shaded slopes, are more accurately aggregated, reducing the fragmentation seen in the unsupervised map.

Moreover, RF reduces the overlap among spectrally similar vegetation types — especially Grass and Shrub, which were highly confused in KMeans. The boundaries remain gradual in some ecotonal zones, but the spatial patterns are less noisy and more consistent with topographic variation.

In summary, the Random Forest classifier demonstrates superior spatial coherence and reduced interclass confusion compared to KMeans, particularly for dominant land cover types such as Snow-ice, Trees, and Water. Although transitional vegetation boundaries remain somewhat ambiguous, RF offers a clearer and more ecologically consistent representation of alpine land cover gradients.

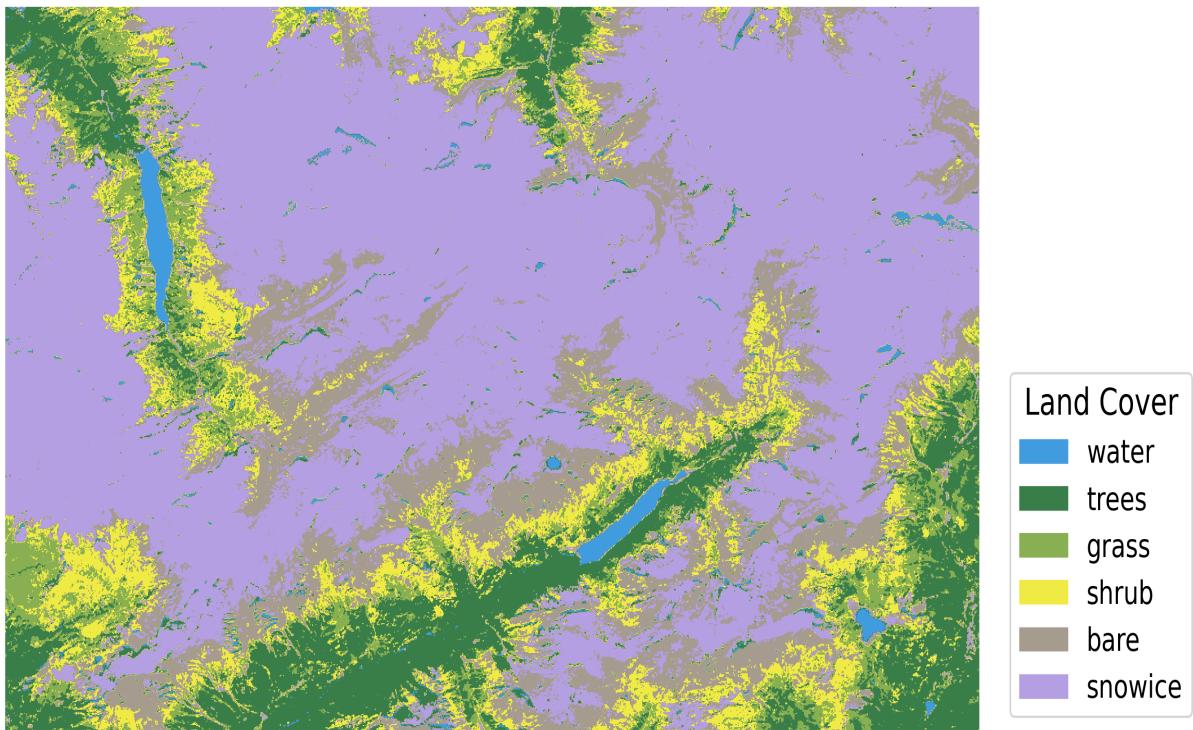


Figure 27. Aosta RF

5.2.3.2. Quantitative Accuracy Assessment and Discussion

Compared to the KMeans results ($OA = 0.92$, $Kappa = 0.90$), the Random Forest classifier yielded substantially improved performance in the Aosta region, achieving an Overall Accuracy of 0.98 and a Kappa coefficient of 0.97. This near-perfect consistency highlights the model's robustness in handling complex alpine terrain and confirms its superiority over unsupervised clustering in resolving class boundaries and transitional zones.

While both models performed well for spectrally distinct categories, RF achieved notable gains across all classes, especially in reducing confusion between Grass and Shrub—a major source of error in KMeans. For instance, Shrub improved from $UA = 0.75$ in KMeans to 0.94 in RF, indicating fewer misclassified Grass pixels. Likewise, Grass saw a PA increase from 0.77 to 0.97, suggesting improved recognition of vegetated mid-altitude slopes. The Snow-ice and Water categories maintained perfect or near-perfect accuracy in both models, but RF further reduced residual confusion between Snow-ice and Water, which had slightly affected KMeans results.

The confusion matrix of RF exhibited a sharper diagonal structure, with minimal off-diagonal errors and reduced overlap among spectrally adjacent classes such as Bare, Shrub, and Grass. In contrast to KMeans, where these intermediate categories exhibited higher mixing, RF produced cleaner boundaries, especially in elevation transition zones and ecotones.

In conclusion, the Random Forest classifier clearly outperforms KMeans in the Aosta case study. It maintains high accuracy in dominant classes and significantly improves the classification of intermediate land cover types in challenging high-altitude environments, demonstrating its strong adaptability, precision, and generalization ability in mountainous landscapes.

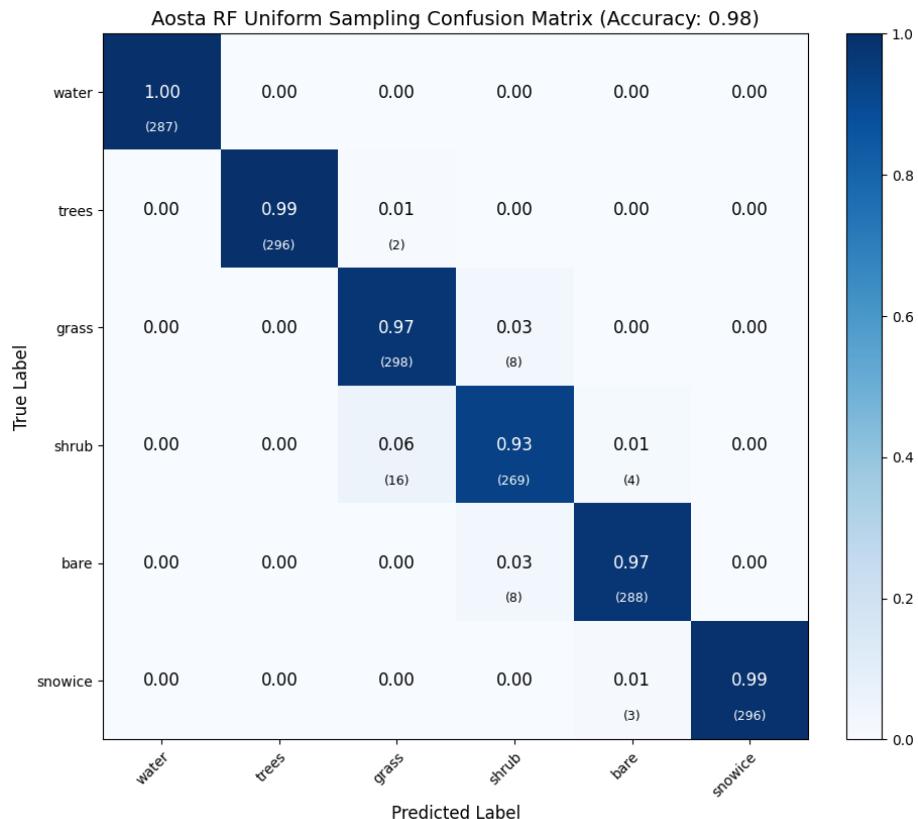


Figure 28. Aosta RF Confusion Matrix.

Class	PA	UA
Water	1.00	1.00
Trees	0.99	1.00
Grass	0.97	0.94
Shrub	0.93	0.94
Bare	0.97	0.98
Snow-ice	0.99	1.00
Overall Accuracy	0.98	
Kappa Coefficient	0.97	

Table 19. Aosta Accuracy Report.

5.3. Multi Layer Perceptron

5.3.1. Hyperparameter Configurations

The hyperparameter tuning procedure applied to the MLP classifier has been detailed in Chapter 2.3.3. Following this procedure, independent tuning experiments were conducted for Shanghai, Tuscany, and Val d'Aosta. The optimal hyperparameters are summarized in Table 20.

AOI	Activation	Alpha	Hidden Layer Sizes	Learning Rate
Shanghai	relu	0.05	(128, 64, 32)	constant
Tuscany	tanh	0.05	(100, 50)	constant
Aosta	tanh	0.0001	(128, 64, 32)	constant

Table 20. Optimal MLP Hyperparameters for Each AOI.

As shown in Table 20, the optimal parameters differ across AOIs. Shanghai and Aosta favored deeper three layer networks (128,64,32), indicating that their spectral distributions and class boundaries are more complex and require greater model capacity. Tuscany, however, achieved better results with a simpler two-layer architecture (100,50), suggesting relatively more regular land cover patterns. In terms of regularization, both Shanghai and Tuscany required stronger constraints ($\text{alpha}=0.05$) to suppress overfitting, while Aosta performed best with weaker regularization ($\text{alpha}=0.0001$), reflecting its demand for stronger fitting capacity. Notably, all three AOIs converged to constant learning rates, implying that under the adaptive mechanisms of the adam solver, a fixed base learning rate provided the most stable convergence.

5.3.2. Shanghai

5.3.2.1. Visual Interpretation of the Classification Map

Compared to the classification results obtained from KMeans and Random Forest, the Multi-Layer Perceptron classifier demonstrates a superior capacity to generate a highly structured and spatially coherent urban land cover map for the Shanghai region. Water bodies are consistently and precisely delineated, with the Yangtze River and its tributaries forming smooth, continuous boundaries that closely match the actual hydrological features. Unlike KMeans, which tends to introduce spectral confusion in high-reflectance zones, and Random Forest, which occasionally produces fragmented water edge artifacts, the MLP results display clean, uninterrupted river shapes and clearly separated water boundaries.

The representation of built-up areas is especially well-executed in the MLP classification map. High-density urban zones in the city center, along with linear transportation infrastructure such as major roads and railways, are distinctly and

compactly identified. Compared to KMeans, which suffers from class intermixing and patchy urban delineation, and RF, which captures the general built-up pattern but with less refined boundary details, MLP achieves more geometrically consistent and spatially continuous results. This performance is particularly notable in the southern and central portions of the image, where the urban form is densely packed and complex.

The MLP classifier also exhibits strong capability in recognizing minority and spectrally similar land cover types. Crops are accurately identified in the western and southeastern fringe areas, with improved spatial regularity and reduced noise compared to both KMeans and RF. Scattered patches of trees are effectively captured within both urban green spaces and peri-urban zones, maintaining internal consistency and clear boundaries despite the spectral similarity with crops and built features. Even bare land—typically challenging to extract due to spectral overlap and patch-size limitations—is discernible in the MLP map, particularly along riverbanks and undeveloped pockets, where it is often misclassified in KMeans results or over-smoothed in RF outputs.

Overall, the MLP classification delivers the most accurate and visually coherent land cover map among the three methods applied to the Shanghai region. It excels not only in dominant classes like water and built-up areas but also in minority categories such as crops, trees, and bare land. The model's ability to balance spectral discrimination with spatial continuity makes it especially suitable for complex urban environments, offering a clearer and more reliable foundation for downstream spatial analysis, urban planning, and ecological assessment.

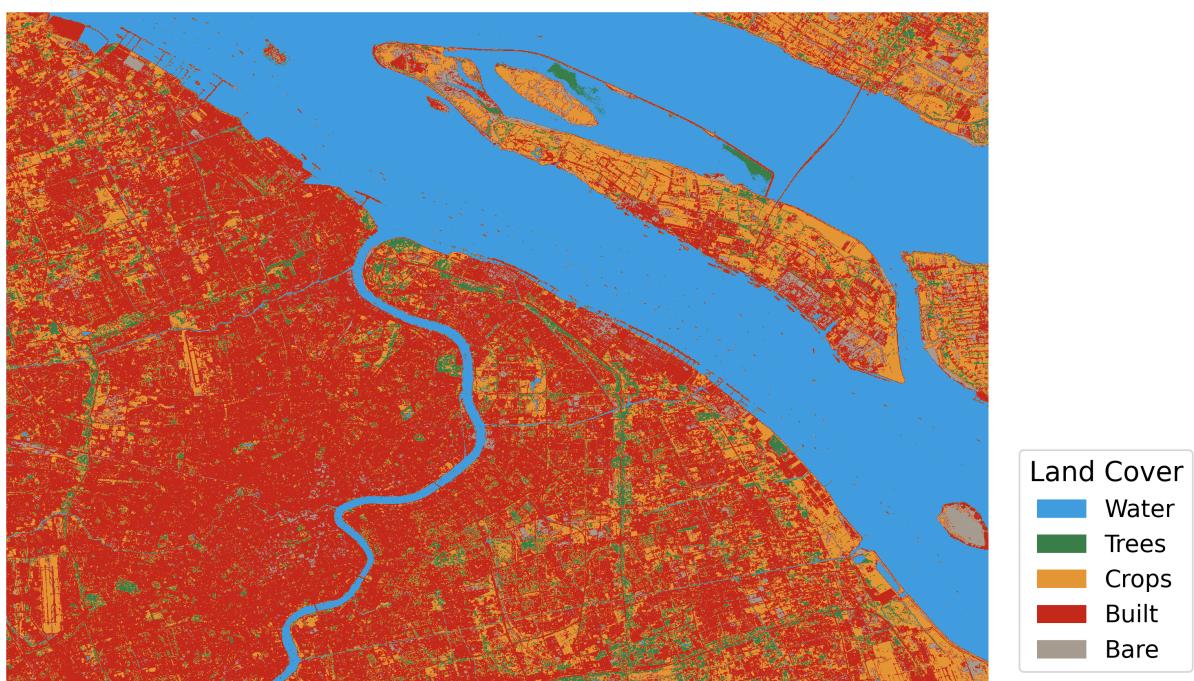


Figure 29. Shanghai MLP.

5.3.2.2. Quantitative Accuracy Assessment and Discussion

The MLP classifier achieved an Overall Accuracy of 0.98 and a Kappa coefficient of 0.98, outperforming both KMeans (OA = 0.88, Kappa = 0.85) and Random Forest (OA = 0.95, Kappa = 0.94) in the Shanghai region. These results highlight the MLP's strong generalization ability and its capacity to model complex spectral relationships in highly urbanized environments.

Across all land cover classes, MLP delivered the highest classification accuracies: Water was perfectly classified across all three models (PA = UA = 1.00), reaffirming its strong spectral separability. For Built, MLP achieved PA = 0.99 and UA = 0.96, significantly surpassing KMeans (PA = 0.79 / UA = 0.72) and slightly improving upon RF (PA = 0.90 / UA = 0.89). This indicates superior handling of spectral mixtures in dense urban cores and transition zones. Bare also achieved perfect classification in MLP (PA = UA = 1.00), outperforming RF (0.98 / 0.96) and KMeans (0.83 / 0.86), demonstrating robust recognition of exposed surfaces. Crops and Trees, typically more challenging due to spectral overlap and sparse distribution, were also classified with high accuracy. Crops reached PA = 0.98 / UA = 0.96 (vs. RF: 0.96 / 0.93; KMeans: 0.93 / 0.90), and Trees reached PA = 0.94 / UA = 0.98 (vs. RF: 0.90 / 0.96; KMeans: 0.86 / 0.95).

These results suggest that MLP not only improves boundary clarity but also reduces the confusion between spectrally similar or spatially fragmented classes—particularly Trees vs. Crops and Built vs. Bare—which were persistent issues in both KMeans and RF.

In summary, the MLP classifier demonstrated the highest overall and per-class accuracy among the three models, confirming its effectiveness in high-density urban land cover mapping under supervised learning frameworks. Its superior performance in both dominant and minority classes suggests strong modeling flexibility and robustness across complex urban landscapes.

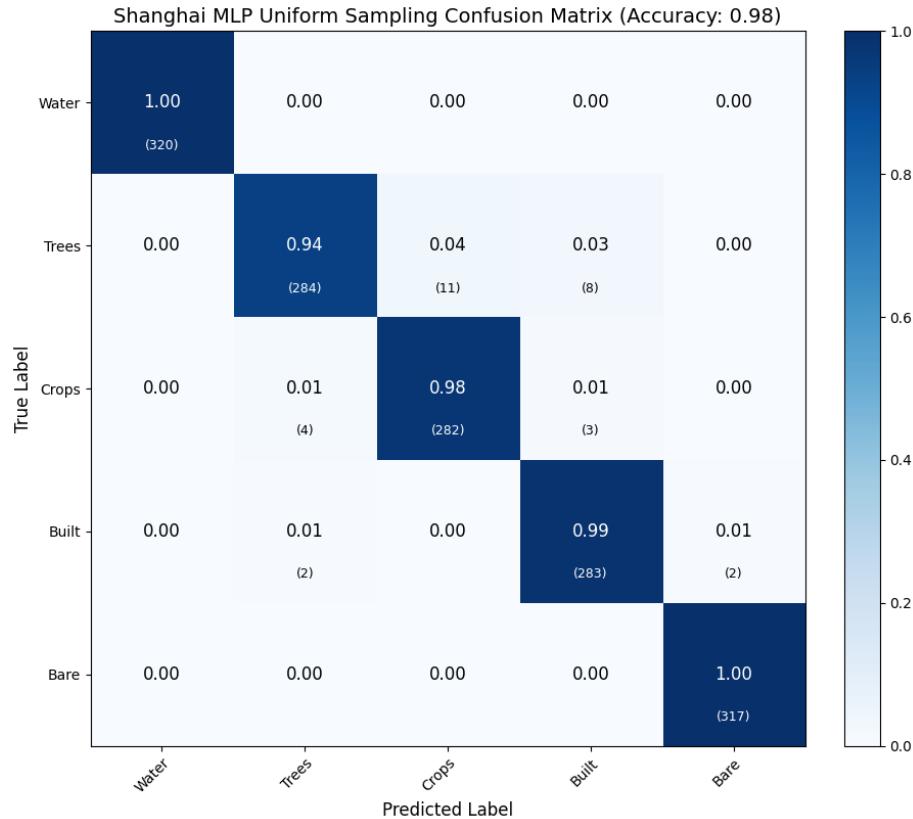


Figure 30. Shanghai MLP Confusion Matrix.

Class	PA	UA
Water	1.00	1.00
Trees	0.94	0.98
Crops	0.98	0.96
Built	0.99	0.96
Bare	1.00	1.00
Overall Accuracy	0.98	
Kappa Coefficient	0.98	

Table 21. Shanghai Accuracy Report.

5.3.3. Tuscany

5.3.3.1. Visual Interpretation of the Classification Map

Compared to KMeans and Random Forest, the Multi-layer Perceptron classifier produced a land cover map for the Tuscany region with superior spatial structure and classification consistency. Overall, the MLP outperformed RF in terms of boundary clarity and handling of small patches, particularly in complex transition zones and hilly terrain.

For dominant land cover types, MLP demonstrated notably accurate delineation of the Trees class, especially across the central-western hills of the image, where tree patches appeared continuous and compact. While RF showed clear improvement over KMeans—which exhibited fragmented and inconsistent tree clusters—MLP further enhanced internal consistency and border regularity.

The Crops class also showed significant improvement. In the KMeans result, agricultural features appeared as irregular and fragmented patches. In contrast, MLP produced more structured and coherent boundaries, aligning more closely with the typical Mediterranean agricultural parcel structure, such as olive groves and vineyards. Regarding the Built category, MLP clearly outperformed both KMeans and RF. Unlike the severe class mixing observed in KMeans, MLP reliably identified rural settlements and transportation corridors in both the eastern and western portions of the image. Compared to RF, MLP yielded cleaner and more coherent boundaries for built-up areas, particularly in small-scale residential clusters and linear infrastructure, demonstrating stronger spatial representation capacity. For low-lying and spectrally similar vegetation types such as Grass, Shrub, and Crops, MLP achieved higher class separability. In the eastern part of the image, shrublands were more accurately detected, and the confusion between shrub and crops was significantly reduced. While KMeans commonly produced blurred and ambiguous boundaries in these transition areas, and RF made partial improvements, MLP displayed sharper delineation and more natural spatial transitions, reflecting superior handling of spectral mixture effects.

In addition, the MLP classification map exhibited greater geometric regularity, substantially reducing the salt-and-pepper noise commonly observed in KMeans results. The overall output was smoother and more coherent than both KMeans and RF, which is particularly important for accurately characterizing the mixed agricultural-natural landscape of Tuscany.

In summary, the MLP classifier demonstrated outstanding performance in land cover mapping for the Tuscany region. It not only provided high accuracy for dominant classes but also effectively distinguished mixed pixels and maintained spatial stability at the patch level. Compared to KMeans, MLP significantly enhanced classification accuracy and map readability; relative to RF, it offered better boundary precision and feature structure representation. Therefore, MLP proves to be the most suitable classification method for complex hilly agricultural landscapes like Tuscany.

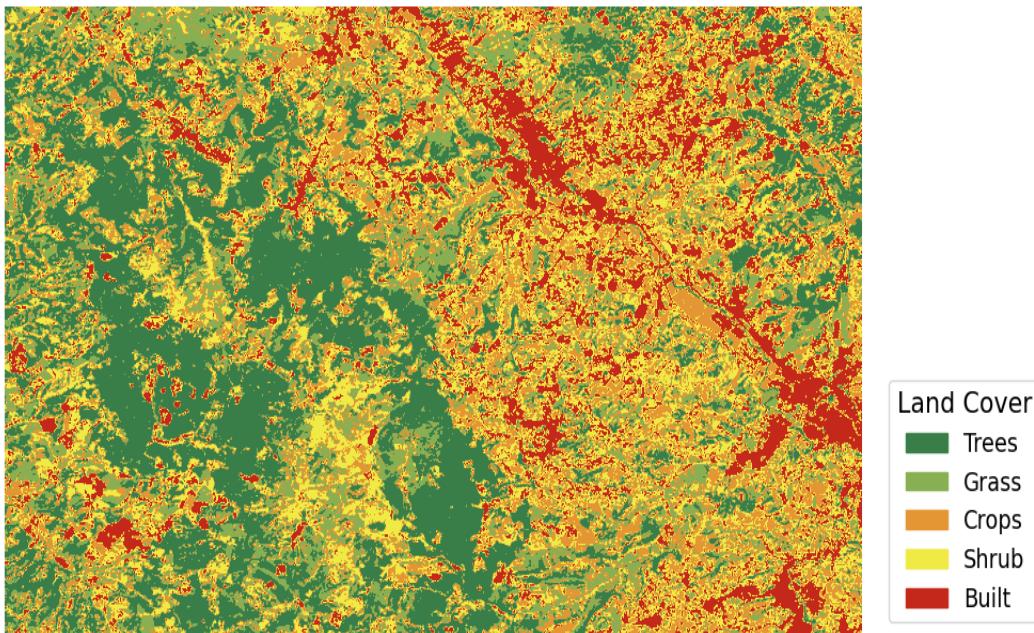


Figure 31. Tuscany MLP.

5.3.3.2. Quantitative Accuracy Assessment and Discussion

In the Tuscany study area, the Multilayer Perceptron classifier outperformed both KMeans and Random Forest models in terms of overall classification accuracy, class-level precision, and spatial consistency. Specifically, MLP achieved an Overall Accuracy (OA) of 0.91 and a Kappa coefficient of 0.89, indicating a strong agreement between the classified results and ground truth labels. These metrics surpass those of KMeans (OA = 0.66, Kappa = 0.57) and RF (OA = 0.86, Kappa = 0.82), suggesting that MLP is particularly well-suited for capturing complex land cover patterns in Mediterranean hilly agricultural landscapes.

The classification performance for the Trees category was nearly flawless with MLP, reaching both Producer Accuracy (PA) and User Accuracy (UA) of 0.99. While this category already achieved strong results under KMeans (PA = 0.95, UA = 0.83) and RF (PA = 0.99, UA = 0.96), MLP further reduced the remaining misclassifications, especially along mixed-pixel boundaries and transition zones. This improvement reflects the MLP model's superior ability to learn stable red and near-infrared spectral patterns typical of tree canopies. The Grass class showed the most significant improvement under MLP. In KMeans, it suffered from severe overclassification (PA = 0.90, UA = 0.54), often confusing Grass with Shrub, Crops, and Built-up pixels. RF mitigated this issue to some extent (UA = 0.87), but MLP achieved a substantial boost in both PA (0.98) and UA (0.91). The enhanced performance demonstrates MLP's advanced capacity to handle spectral ambiguity and terrain-induced reflectance variability in low-lying vegetation types. Crops classification also benefited from MLP's learning capabilities. Compared to KMeans (PA = 0.58, UA = 0.68) and RF (PA = 0.84, UA = 0.83), MLP reached PA = 0.90 and UA = 0.88. This improvement is

particularly relevant in Tuscany, where Mediterranean agricultural plots such as vineyards and olive groves are spectrally similar to other vegetation types. MLP produced more regular patch shapes and sharper boundaries, despite seasonal variability in crop phenology. Shrub was the weakest performing category in the KMeans model ($PA = 0.33$, $UA = 0.50$), primarily due to its fragmented spatial distribution and overlapping spectral signatures with Grass and Crops. RF significantly improved its classification ($PA = 0.74$, $UA = 0.78$), but MLP achieved the best results yet ($PA = 0.79$, $UA = 0.88$). This improvement suggests that MLP can better learn complex class boundaries and mitigate cross-class errors in spectrally ambiguous terrain. For the Built class, MLP yielded PA and UA values of 0.91, representing a significant advancement over KMeans ($PA = 0.54$, $UA = 0.83$) and RF ($PA = 0.78$, $UA = 0.82$). MLP effectively identified rural settlements, roads, and small-scale infrastructure with clean and consistent boundary representation. Misclassification in low-density built-up areas was further reduced, reflecting the model's robustness in identifying discrete anthropogenic features in mixed-use landscapes.

From both the class-wise and overall performance metrics, MLP consistently delivers superior results across all categories. In particular, it shows notable improvements in the Shrub classes, which were previously problematic under KMeans and RF. The high Kappa coefficient suggests that MLP classification outcomes are not merely the result of random agreement but represent genuine learning of class characteristics.

In conclusion, MLP stands out as the most effective classifier for the Tuscany region, especially in managing spectral overlaps, maintaining spatial coherence, and achieving accurate delineation of land cover types.

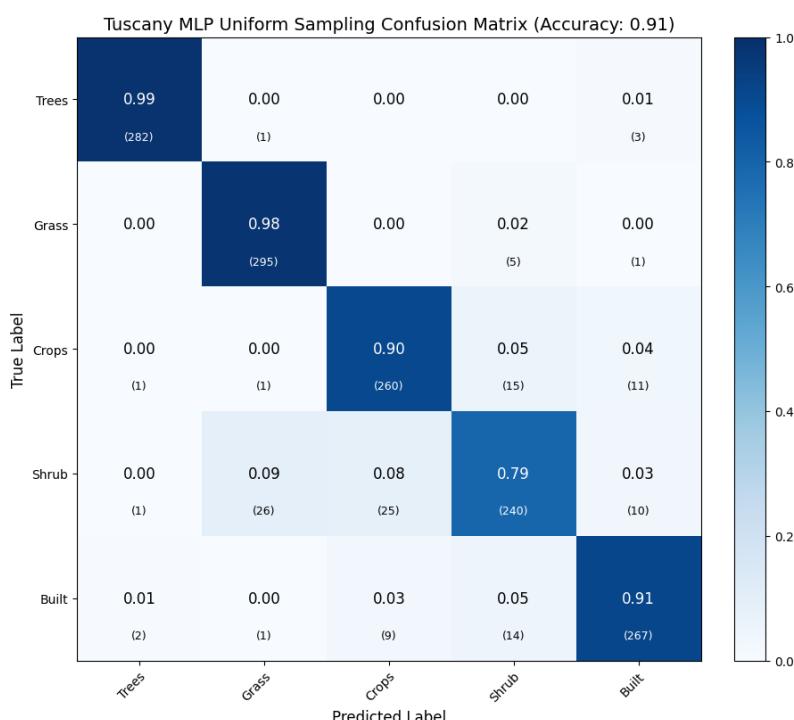


Figure 32. Tuscany MLP Confusion Matrix.

Class	PA	UA
Trees	0.99	0.99
Grass	0.98	0.91
Crops	0.90	0.88
Shrub	0.79	0.88
Built	0.91	0.91
Overall Accuracy		0.91
Kappa Coefficient		0.89

Table 22. Tuscany Accuracy Report.

5.3.4. Aosta

5.3.4.1. Visual Interpretation of the Classification Map

The Multi-Layer Perceptron classification map of the Aosta region presents a highly structured and ecologically coherent depiction of land cover, clearly reflecting the typical vertical zonation patterns characteristic of alpine environments. Compared with the K-means clustering and Random Forest classification methods, the MLP results demonstrate superior spatial consistency, smoother transitions between land cover types, and significantly reduced noise interference, particularly in areas with complex terrain.

At high elevations, snow and ice cover in the northern and central portions of the image are accurately mapped, forming continuous and compact glacier zones that closely correspond to the known perennial snowfields and glacial regions of the Alps. Unlike the KMeans classification—where spectral confusion in the visible bands led to portions of snow-ice being misclassified as water—and the RF result—which mostly corrected this issue but still showed slight overextensions, the MLP classifier successfully distinguishes snow-ice from water with high precision. Water bodies are strictly confined to valleys and alpine lakes, appearing as narrow, linear features that align well with the underlying topography. This notable improvement in water classification highlights the MLP model's enhanced ability to handle spectral ambiguity in high-reflectance environments.

The spatial distribution of bare, shrub, and grass classes is also more refined in the MLP map, closely matching the ecological gradient and topographic context of the region. Bare land is clearly delineated along ridges and rocky slopes, maintaining sharp transitions with adjacent classes. Shrub form fragmented but ecologically plausible patches between bare and grass zones, with noticeably less confusion than in the KMeans result and comparable clarity to that of the RF output. Grass, widely spread across mid-elevation slopes and valleys, are effectively separated from shrubs

and trees, with the MLP classifier maintaining clearer boundaries and more consistent internal structure across these vegetated types.

Overall, the MLP classification map exhibits the highest level of spatial coherence among the three approaches. It not only minimizes salt-and-pepper noise but also provides a clearer and more ecologically consistent representation of both large-scale ecological patterns and fine-scale land cover features. The boundaries of major categories such as snow-ice, bare ground, and vegetation types are sharp and topographically logical, while the transitions between spectrally similar vegetation types like shrubs and grass are handled with greater nuance. These results underscore the MLP classifier's superior capability in capturing the complex spatial and spectral characteristics of mountainous landscapes. By effectively distinguishing between spectrally overlapping classes and preserving landscape-level continuity, MLP proves to be the most suitable classification method for the heterogeneous terrain of the Aosta region.

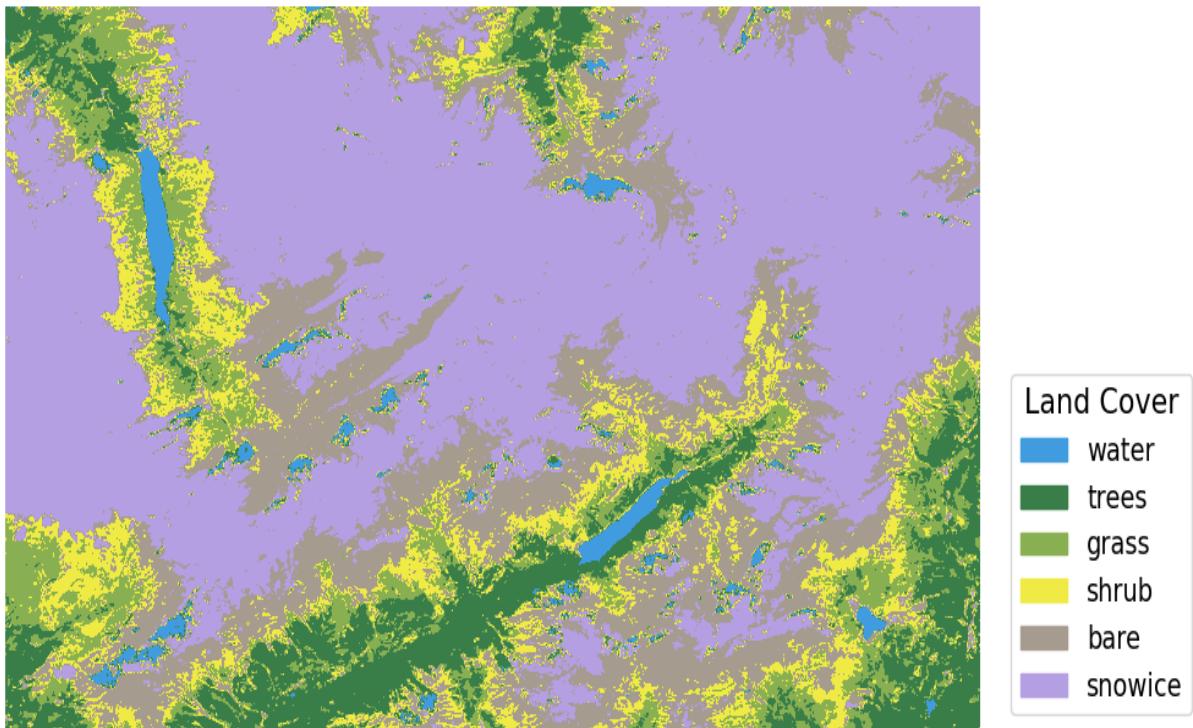


Figure 33. Aosta MLP

5.3.4.2. Quantitative Accuracy Assessment and Discussion

Compared to the results of KMeans ($OA = 0.92$, $Kappa = 0.90$) and Random Forest ($OA = 0.98$, $Kappa = 0.97$), the Multi-Layer Perceptron (MLP) classifier achieved an Overall Accuracy of 0.97 and a Kappa coefficient of 0.97 in the Aosta region, indicating a level of performance closely comparable to RF and clearly superior to the unsupervised KMeans approach. These results confirm the strong adaptability of supervised neural network models to complex alpine environments, while also demonstrating their

potential to provide balanced classification across both dominant and transitional land cover categories.

The spectrally distinct classes, such as Water and Snow-ice, were classified with perfect accuracy in MLP ($PA = UA = 1.00$). These results not only match the performance of RF but also eliminate the minor misclassifications observed in KMeans, where snow-ice areas were occasionally confused with water. The boundaries of lakes and glaciers are sharply delineated in the MLP classification, reflecting the model's robust discrimination ability in high-reflectance environments. For Trees, the MLP classifier further improved upon KMeans results ($PA = 0.97$, $UA = 0.99$) and achieved nearly flawless recognition ($PA = 0.99$, $UA = 1.00$), comparable to RF. Tree-covered slopes and valley forests were consistently mapped, with minimal spectral confusion with surrounding vegetation, highlighting the stability of the neural network in handling near-infrared and red-edge features. The greatest improvements relative to KMeans were observed in the Grass and Shrub categories, which had been the major sources of error in the unsupervised results. MLP significantly reduced misclassification, achieving $PA = 0.96$ and $UA = 0.90$ for Grass, and $PA = 0.89$ and $UA = 0.94$ for Shrub. Compared with RF, which also improved these categories, the MLP results are similar in magnitude but show slightly higher user accuracy for Shrub. This suggests that MLP effectively captures subtle spectral and spatial differences in transition zones, producing a more reliable separation between alpine meadows and shrub-dominated areas. The Bare category also performed strongly under MLP ($PA = 0.98$, $UA = 0.99$), surpassing KMeans ($PA = 0.93$, $UA = 0.94$) and aligning closely with RF. The model achieved clean, consistent boundaries between bare ground and surrounding vegetation.

Overall, the MLP classifier in the Aosta region demonstrates excellent classification performance, with accuracy metrics nearly equivalent to those of Random Forest and far surpassing KMeans. While RF slightly outperforms MLP in terms of grass producer accuracy, the MLP model achieves superior shrub user accuracy and maintains balanced results across all categories. These findings highlight that MLP not only matches the performance of ensemble-based supervised methods but also provides robust generalization in ecologically complex mountainous terrains.

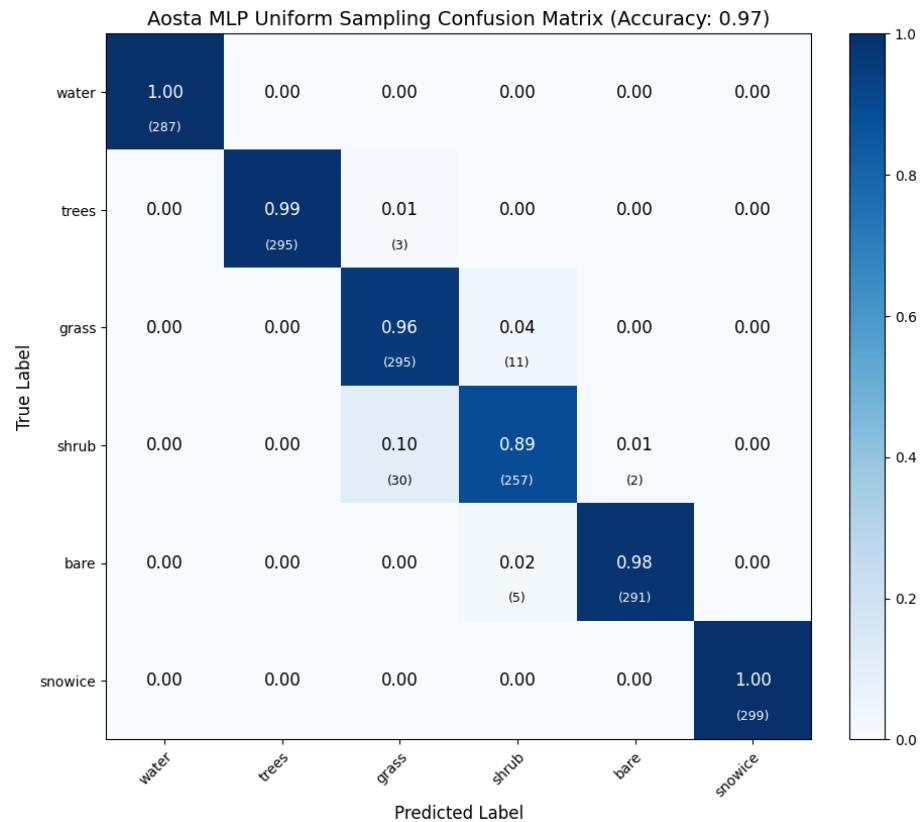


Figure 34. Aosta MLP Confusion Matrix.

Class	PA	UA
Water	1.00	1.00
Trees	0.99	1.00
Grass	0.96	0.90
Shrub	0.89	0.94
Bare	0.98	0.99
Snow-ice	1.00	1.00
Overall Accuracy	0.97	
Kappa Coefficient	0.97	

Table 23. Aosta Accuracy Report.

6 Comparison of Uniform and Proportional Sampling Strategies

To further validate the impact of sampling strategies on classification performance, last analysis designed and conducted a set of lightweight proportional sampling experiments as a comparison to the uniform sampling experiments. This experiment aims to systematically evaluate how different training strategies affect classifier accuracy and feature recognition capabilities, with particular focus on whether models develop classification bias due to imbalanced training sample proportions in scenarios where land cover types are highly unevenly distributed. In this experiment, training samples were extracted according to the actual area proportions of each land cover type within the study area. Specifically, training samples for different categories were proportionally extracted from the overall sample set based on their actual distribution.

6.1. Shanghai: comparison on RF

In the proportional sampling design for the Shanghai region, the distribution of training samples was strictly based on the actual area proportions of each type of land feature within the study area. As shown in Table 24, Built Land is the dominant class, accounting for 45.19% of the total samples, with 30,831 pixels and an area of 3.08 km². Water is the second most common class, accounting for 34.82%, with 23,757 pixels and an area of 2.38 km². Crops and Bare Land account for 9.51% and 7.18%, respectively. Trees is the smallest class, accounting for only 3.30%, with 2,251 pixels and an area of 0.23 km².

Type	Water	Trees	Crops	Built	Bare
Pixels	23,757	2,251	6,490	30,831	4,895
Area(km²)	2.38	0.23	0.65	3.08	0.49
Coverage	34.82%	3.30%	9.51%	45.19%	7.18%
Polygons	33	30	25	77	31

Table 24. Shanghai Proportional Sampling.

From the perspective of spatial distribution characteristics, Figure 34 (proportional sampling) reveals a typical “building-dominated” classification pattern. Extensive red areas represent the Built category, with high concentrations of this category observable in urban cores and their peripheries, urban-rural transition zones, and even extending into crops and bare patches. The classification results reveal a pronounced “dominant category overriding effect,” where the model tends to classify areas with ambiguous spectral features or blurred boundaries as the dominant category (Built). This leads to

systematic misclassification of non-dominant categories such as crops, bare, and trees at urban fringes. This bias correlates strongly with the high proportion of built areas in the training dataset (45.19%). During training, the model amplified its response to spectral information characteristic of built areas while reducing sensitivity to other categories.

In contrast, Figure 22 (uniform sampling) exhibits a more detailed spatial structure. Since the training phase allocated an equal number of samples to each land cover class (e.g., 1000 pixels per class), the model was compelled to learn spectral and textural features of all categories from a balanced perspective. This strategy significantly enhances the model's recognition capability and boundary discrimination for smaller categories such as crops, bare, and trees. The image reveals more scattered yet clearly defined feature patches, particularly in the eastern and western agricultural areas. Uniform sampling successfully extracts boundaries of crops and bare ground nested within built-up blocks, whereas proportional sampling broadly classifies these areas as Built, failing to capture local heterogeneity. Furthermore, in transition zones between trees, bare, and crops, uniform sampling better restores the true spatial patterns of patches, preserving transitional features and distributional integrity between land cover categories.

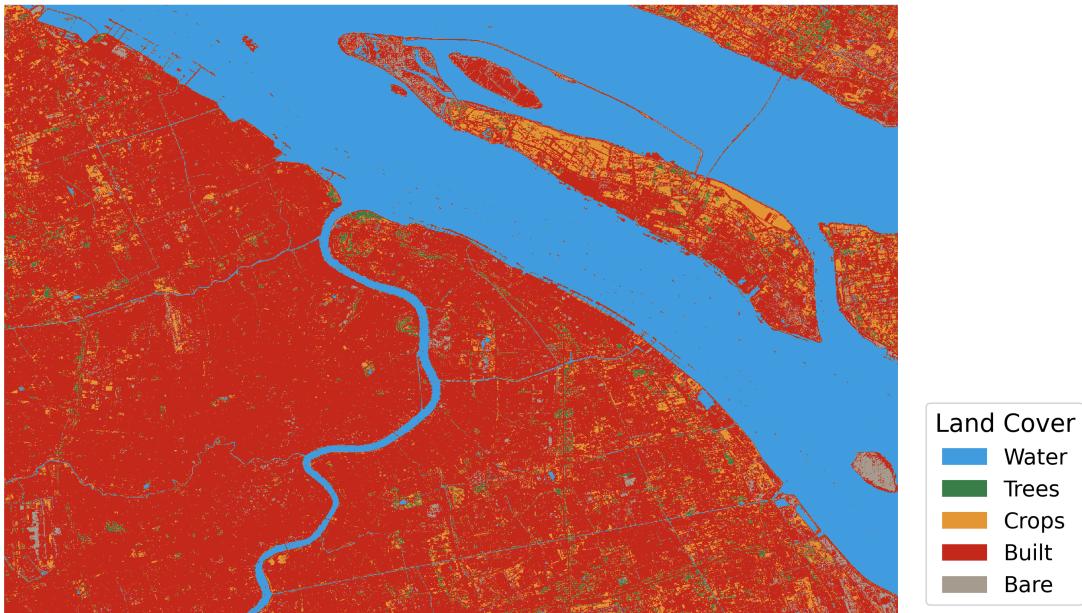


Figure 35. RF Proportional Sampling.

In terms of overall accuracy, the uniform sampling strategy significantly outperforms proportional sampling, improving the overall classification accuracy from 0.90 to 0.95. This result indicates that a balanced training sample configuration helps the model comprehensively learn the spectral characteristics of various land cover types, particularly enhancing its discrimination capability for minority classes.

Analyzing the confusion matrix reveals pronounced category imbalance under proportional sampling. The dominant Built class accounted for 45.19% of the training set, causing the model to exhibit significant “majority bias” during classification. For instance, the validation set accuracy for the Crops category was only 39%, with a staggering 52% of pixels misclassified as Built, indicating the model’s difficulty in accurately identifying complex agricultural features in urban-rural fringe areas. Under uniform sampling, however, the model’s recognition capability for the Crops category significantly improved due to providing equal training samples for each class. The classification accuracy rose to 96%, with a substantial decrease in misclassification rate, demonstrating higher sensitivity and discriminative power.

The improvement for the Bare category was even more pronounced. Under proportional sampling, its accuracy was only 36%, with 58% of samples misclassified as Built. Under uniform sampling, accuracy jumped to 98%, with misclassification rates drastically reduced. This significant improvement also benefited from balanced training sample allocation, enabling the model to effectively learn the spectral and textural features of bare and avoid being “overwhelmed” by strong interference from dominant classes. The classification accuracy for the Trees category reached 92% under proportional sampling and 90% under uniform sampling. Although the accuracy rate has decreased slightly, image boundaries and morphological features appeared more complete. Uniform sampling particularly preserved the fragmented structures in areas like urban green spaces, field ridges, and roadside green belts. Notably, trees maintained relatively high classification accuracy under both sampling strategies despite a comparatively smaller number of training samples. This is largely due to the distinct spectral profile of vegetation in multispectral imagery, particularly the strong reflectance in the near-infrared band and the pronounced NDVI response, which makes trees highly separable from other land cover types even with fewer samples.

It is worth noting that while uniform sampling achieved significant improvements in most class recognition, it also introduced a degree of “dominant class weakening.” In proportional sampling, the Built category achieved 97% classification accuracy due to its large training sample size; under uniform sampling, accuracy dropped to 90%, with some pixels misclassified as Trees (3%), Crops (3%), and Bare (4%). This performance decline primarily stems from the model no longer favoring dominant class features during training. When processing boundary or transitional areas, it tends toward balanced judgments, thereby affecting its overall coverage of built features.

In summary, uniform sampling offers significant advantages in enhancing overall classification balance and improving the model’s ability to recognize complex feature structures, particularly for Crops and Bare categories. However, this strategy may come at the cost of reduced local accuracy for dominant classes.

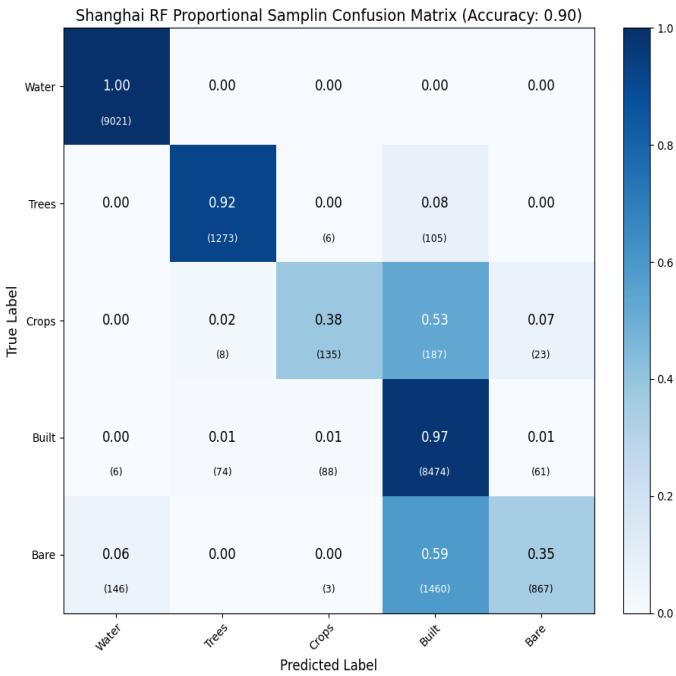


Figure 36. Proportional Sampling Confusion Matrix.

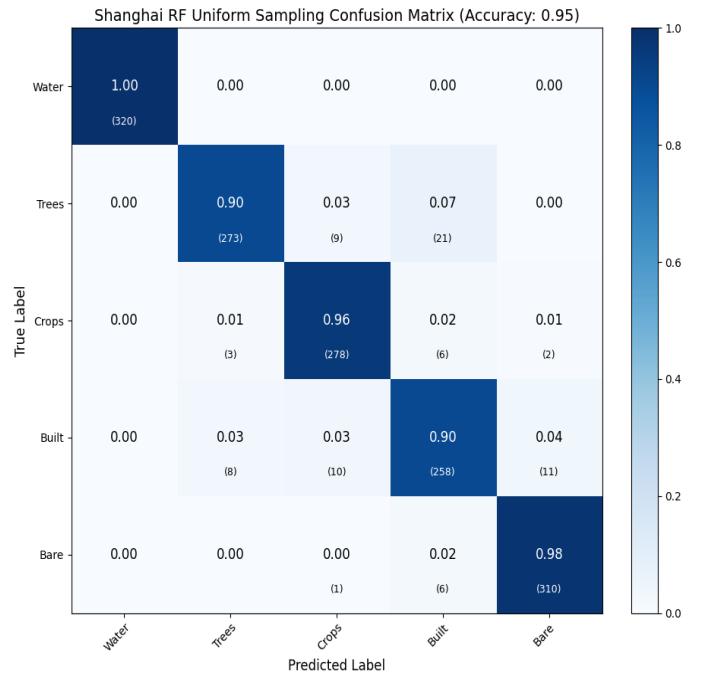


Figure 37. Uniform Sampling Confusion Matrix.

6.2. Tuscany: comparison on RF

The results are shown in Table 25. It is evident that significant disparities exist in sample quantities across categories, with Trees being the most prominent. A total of 37,473 training pixels were acquired, accounting for 50.28% of the total samples. This corresponds to a ground area of 3.75 km², demonstrating a typical dominant land cover feature in the remote sensing imagery. Crops represent a major land cover type in Tuscany, with 16,576 samples (22.24%) covering 1.66 km². Though lower than trees, this category still qualifies as moderately abundant, facilitating effective spectral feature learning by the model. The Shrub category comprises 8,014 pixels, accounting for 10.75% of the total, covering an area of 0.80 km². In contrast, the Grass and Built categories exhibit relatively smaller sample sizes, accounting for only 7.44% and 9.28% of the total, respectively. Their corresponding land cover areas are also limited, at 0.55 km² (5,542 pixels) and 0.69 km² (6,919 pixels).

Type	Trees	Grass	Crops	Shrub	Built
Pixels	37,473	5,542	16,576	8,014	6,919
Area(km²)	3.75	0.55	1.66	0.80	0.69
Coverage	50.28%	7.44%	22.24%	10.75%	9.28%
Polygons	60	38	42	30	32

Table 25. Tuscany Proportional Sampling.

Figure 37 shows the random forest classification results based on proportional sampling. The dominant class is Trees, which is primarily distributed in the center of the image, exhibiting a large, continuous distribution. Crops are primarily concentrated in the eastern portion of the image. Shrubs are scattered, primarily occurring at the edges of trees and crops. Built structures are distributed along roads and residential areas, with a relatively low concentration.

In contrast, the uniform sampling results shown in Figure 24 exhibit greater fragmentation at class boundaries and in detail. The distribution area of the Trees class is further subdivided into Grass and Shrub at the boundaries, resulting in a more fragmented shape in the tree area. Notably, the Built class shows a significant increase in recognitions with uniform sampling, with the red area in the image expanding significantly, capturing more scattered small built-up areas and road network structures.

Table 25 lists the number of pixels, coverage area, and polygon count for each class under proportional sampling. It can be seen that the Trees class accounts for over 50% of the sample count, while minority classes such as Grass, Shrub, and Built structures are significantly underrepresented. This imbalanced distribution causes proportional sampling to favor learning the spectral features of the majority class, resulting in clear boundaries for large classes and insufficient recognition of smaller classes in the classification results. For example, the Built class in Figure 37 has a limited distribution, potentially leading to missed detections. Uniform sampling improves the representation of minority classes by allocating an equal number of training samples to each class, resulting in a more fragmented distribution across classes.

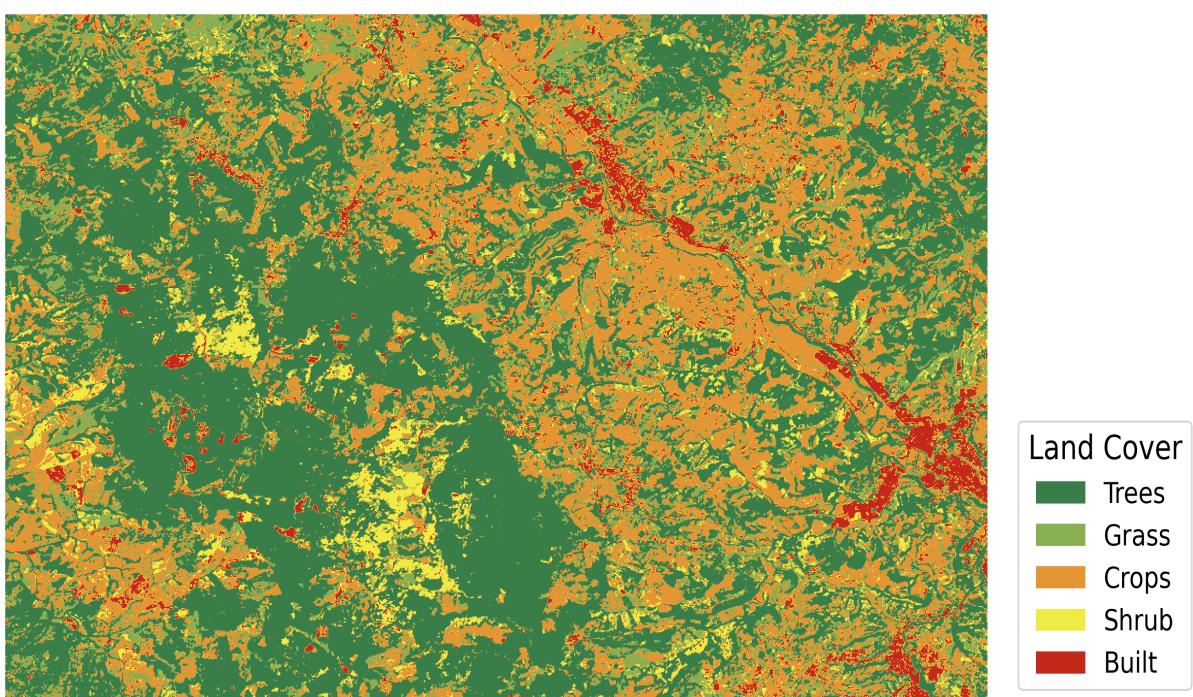


Figure 38. RF Proportional Sampling.

As shown in the figure, the overall accuracy of the model under the proportional sampling strategy is 0.78, while under the uniform sampling strategy, the overall accuracy improves to 0.86.

The different sampling strategies significantly affect the model's recognition ability. The Trees class exhibits extremely high classification accuracy under both sampling strategies, with a correct classification percentage of 0.99. This indicates that the spectral characteristics of this class are stable, the model is easy to recognize, and it is less affected by the sampling strategy. In contrast, Grass and Shrub are the classes that benefit most significantly: under proportional sampling, the correct classification percentage of Grass is only 0.49, and that of Shrub is 0.41, with both being misclassified as each other and Crops. Under uniform sampling, however, the recognition accuracy of these two classes significantly improves to 0.93 and 0.74, respectively, effectively alleviating the class confusion problem and verifying the effectiveness of minority class augmentation in improving the model's discriminative ability. The Crops category demonstrated high recognition accuracy under both sampling strategies, with correct classification rates of 0.81 and 0.84, respectively. This indicates that its spectral characteristics are relatively clear and less affected by the sampling method. For the Built category, the recognition accuracy under proportional sampling was 0.69, lower than the 0.78 under uniform sampling.

In summary, the uniform sampling strategy significantly enhances the model's learning ability for the minority class while maintaining recognition accuracy for the majority class, thereby effectively improving overall classification accuracy. Quantitative analysis of the confusion matrix shows that Grass and Shrub are the categories with the most significant improvements, with uniform sampling significantly alleviating the class confusion issues that exist in proportional sampling for both categories.

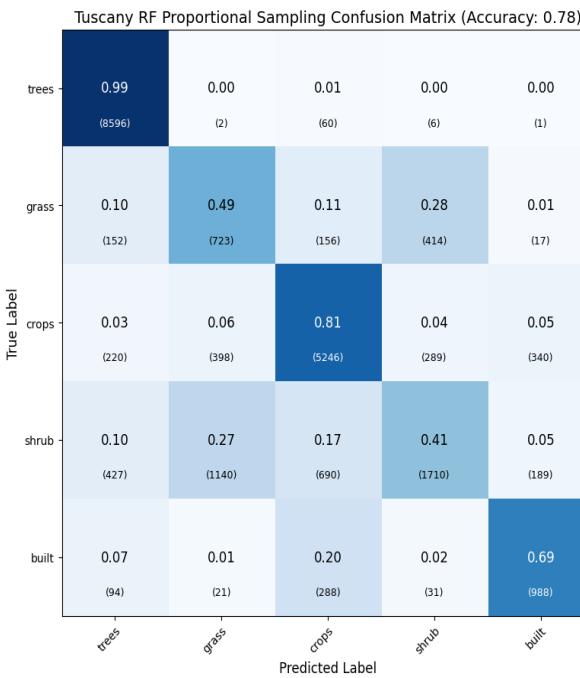


Figure 39. Proportional Sampling Confusion Matrix.

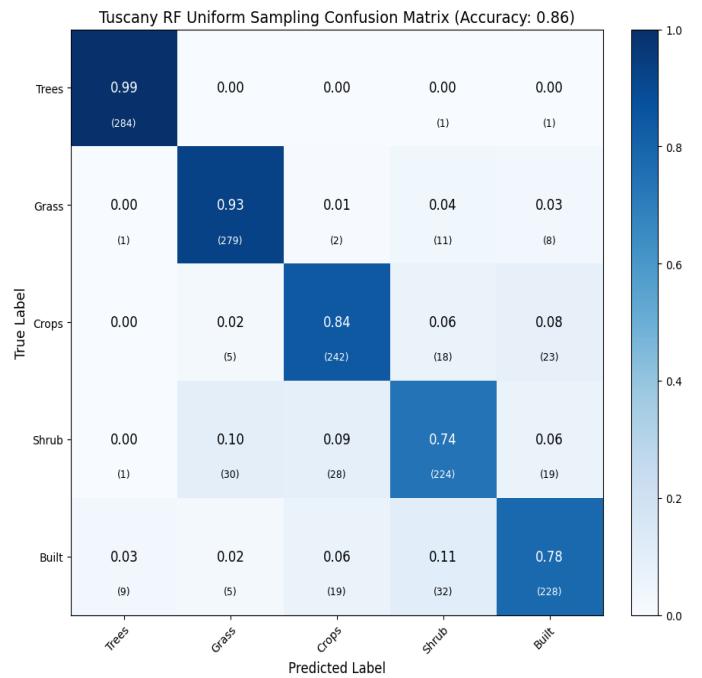


Figure 40. Uniform Sampling Confusion Matrix.

6.3. Aosta: comparison on RF

The statistical results of the training data are shown in Table 26. The six land cover classes exhibit significant unevenness in pixel count and surface area distribution, revealing a pronounced dominance by the Snow-ice category. The Snow-ice class comprises 27,640 training pixels, accounting for 50.32% of the total sample, with a corresponding surface area of 2.76 km². It is extensively distributed across high-altitude mountainous regions, constituting the dominant landform type in the region's remote sensing imagery. The Bare category follows, comprising 10,970 pixels (19.97% of total samples) covering 1.10 km², primarily concentrated in exposed mountain slopes and areas of glacial retreat. In contrast, Water and Grass categories had relatively fewer samples, with 1,580 and 2,328 pixels respectively, accounting for only 2.88% and 4.24% of the total. Their coverage areas were mainly distributed in valley lowlands and slope buffer zones. Shrub and Trees categories had moderate sample sizes, with 7,727 and 4,683 pixels respectively.

Type	Water	Trees	Shrub	Grass	Bare	Snow-ice
Pixels	1,580	4,683	7,727	2,328	10,970	27,640
Area(km²)	0.16	0.47	0.77	0.23	1.10	2.76
Coverage	2.88%	8.53%	14.07%	4.24%	19.97%	50.32%
Polygons	20	47	31	38	73	65

Table 26. Aosta Proportional Sampling.

In terms of overall feature classification accuracy, the random forest model trained using proportional sampling in the Aosta region demonstrated significant advantages in distinguishing dominant categories. Due to the large proportion of Snow-ice and Bare categories in the training samples, these classes exhibit extensive coverage, clear boundaries, and stable overall distribution in the classification map. This model tends to enhance recognition capabilities for dominant classes, consequently weakening its ability to identify minor categories like Trees, Grass, and Shrub within images. Under proportional sampling, the distribution of Trees and Grass in valleys and slopes noticeably contracted, while Shrub classification was similarly suppressed, exhibiting insufficient discrimination in marginal and mixed zones. In contrast, Bare and Snow-ice exhibit stable performance under both sampling strategies, showing virtually identical distribution patterns across images.

The uniform sampling strategy significantly improves classification performance for minor categories, enabling the model to learn more balanced capabilities across land cover types. Classification maps reveal enhanced continuity for Trees, Grass, and Shrub categories on sloped terrain under uniform sampling. Grass and Trees exhibit more complete classification bands along mountain edges, while Shrub spatial distribution becomes more reasonable. Overall, uniform sampling enhances the model's expressive power for fine-grained categories and improves spatial feature reproduction in images. This approach is particularly effective in complex mountainous regions like Aosta, characterized by intricate topography and highly uneven category distributions.

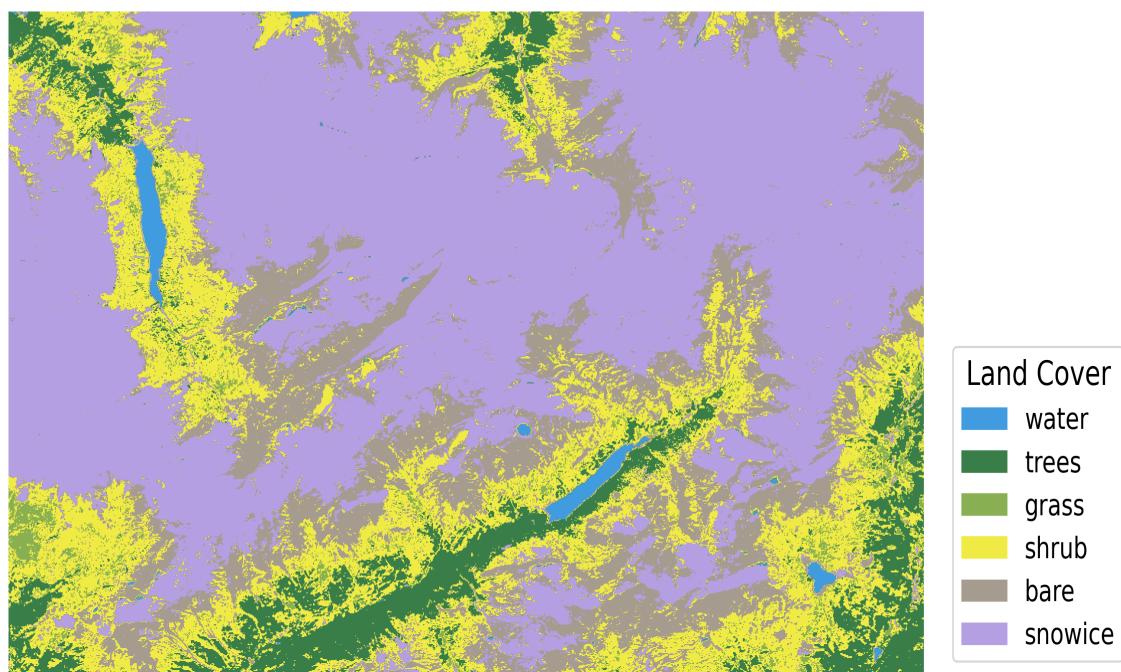


Figure 41. RF Proportional Sampling

In terms of overall accuracy, the random forest model for the Aosta region demonstrated high classification performance under both sampling strategies. When using proportional sampling, the model achieved an overall classification accuracy of 93.05%. Under the uniform sampling strategy, this metric further improved to 97.96%, an increase of nearly 5 percentage points. This indicates that in mountainous environments characterized by steep topography and complex feature distribution, uniform sampling significantly enhances the model's generalization capability and stability. The overall performance improvement primarily stems from the model's enhanced discrimination capability for non-dominant categories. Specifically, the recognition accuracy for Grass and Shrub categories markedly improved under uniform sampling, effectively suppressing the model's tendency to overfit dominant categories like Snow-ice. Notably, both Water and Snow-ice achieved classification accuracies close to 100% under both sampling strategies, demonstrating strong robustness in areas with distinct terrain structures or highly separable spectral features, where classification results are highly reliable.

The uniform sampling strategy notably enhances discrimination for easily confused categories. For instance, in proportional sampling, the Grass category achieved only 43% classification accuracy, with a high 57% of pixels misclassified as Shrub, revealing significant cross-confusion issues. In contrast, uniform sampling significantly mitigated this issue, boosting the Grass class recognition rate to 97%. Simultaneously, the accuracy rates for Trees and Shrub reached 99% and 93%, respectively, substantially outperforming the proportional sampling results.

In summary, despite differing sampling strategies, the random forest model for the Aosta region achieved high overall classification accuracy under both proportional and uniform sampling conditions, demonstrating robust classification performance and algorithmic adaptability. This high precision stems from the stable spectral characteristics and spatial distribution patterns of the Bare, Water and Snow-ice categories in the remote sensing data, enabling the model to perform stably in primary class discrimination. Meanwhile, uniform sampling demonstrated superior effectiveness in further enhancing the recognition capability of subcategories and mitigating inter-class confusion, making it more suitable for classification tasks in alpine environments characterized by significant terrain undulations and strong land cover diversity.

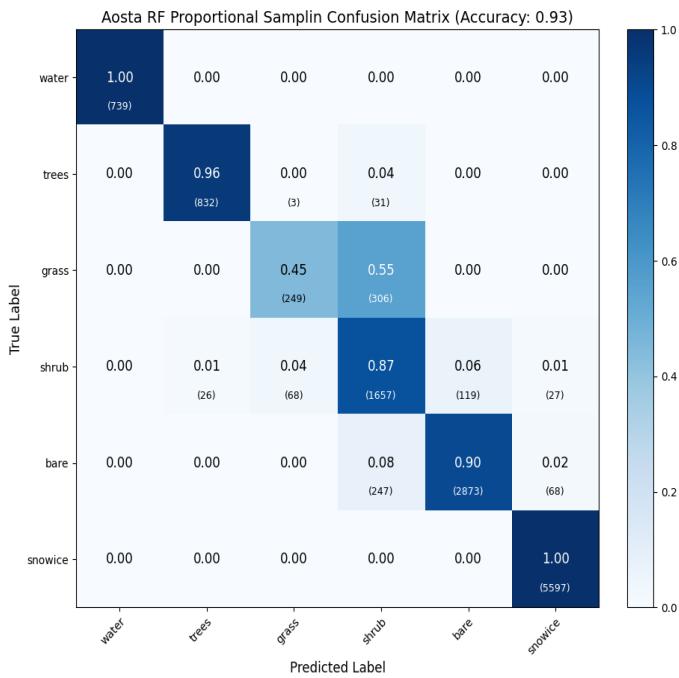


Figure 42. Proportional Sampling Confusion Matrix.

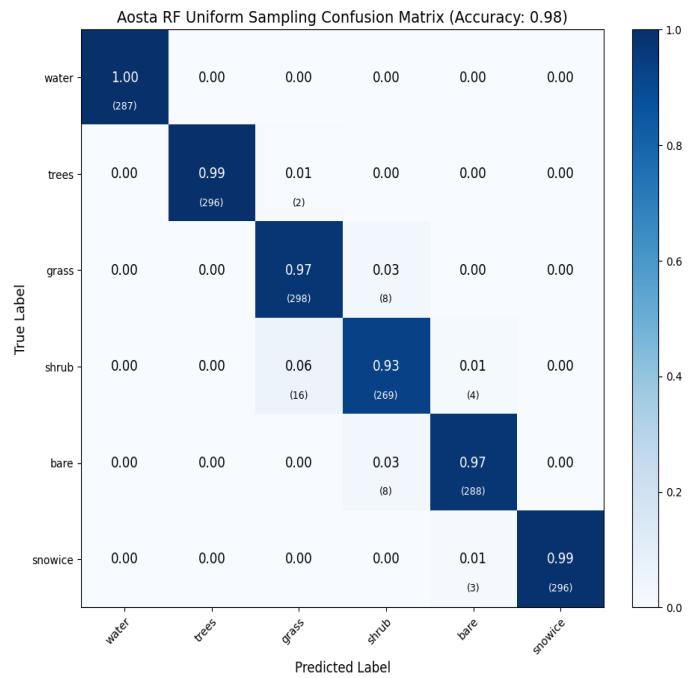


Figure 43. Uniform Sampling Confusion Matrix.

7 Conclusion

7.1. Comparative Analysis of Classification Models

This study utilizes Sentinel-2 multispectral remote sensing imagery as its data foundation, focusing on three representative terrains—urban (Shanghai), agricultural hills (Tuscany), and alpine snow-covered areas (Val d'Aosta). A standardized land cover classification workflow was developed, incorporating manual feature annotation, pixel-level feature extraction, model training, spatial visualization, and accuracy assessment. Three classification algorithms—KMeans, Random Forest (RF), and Multi-Layer Perceptron (MLP)—were systematically compared across the study regions. Performance was evaluated quantitatively through overall accuracy (OA), Kappa coefficients, confusion matrices, and producer/user accuracies, and qualitatively through spatial interpretation of classification maps.

The results demonstrate that supervised learning models consistently outperform the unsupervised KMeans algorithm, with the MLP classifier achieving the most accurate and spatially coherent results across all three areas. In Shanghai, Tuscany, and Aosta, MLP achieved overall accuracies of 0.98, 0.91, and 0.97, respectively, with Kappa coefficients close to or above 0.90, confirming its superior robustness and discriminative ability. RF also performed reliably, with OA values of 0.95, 0.86, and 0.98, but its boundary refinement and recognition of minority classes were slightly weaker than MLP. In contrast, KMeans captured only coarse spatial patterns, with substantially lower accuracies, particularly in Tuscany (OA = 0.66, Kappa = 0.57), where blurred boundaries and spectral overlap caused significant confusion among Grass, Crops, and Shrub.

From the perspective of spatial distribution, MLP not only preserved major landscape features such as water and urban areas in Shanghai or snow–bare–forest gradients in Aosta, but also demonstrated stronger performance in distinguishing minority and transitional classes. For example, in Tuscany's complex patchy terrain, MLP reduced misclassification between Grass, Crops, and Shrub compared to KMeans and RF. RF produced stable and interpretable results for dominant classes but lacked the same boundary precision. KMeans, meanwhile, suffered from high noise and frequent category overlap.

Overall, this study shows MLP as the most effective algorithm for multispectral remote sensing classification, followed by RF, while KMeans remains limited to providing coarse structural references. Future work should address residual misclassification in spectrally similar categories (e.g., Grass vs. Shrub, Trees vs. Crops) through multi-temporal data integration, topographic constraints (DEM, slope, aspect), and the

adoption of advanced deep learning architectures such as CNNs, U-Nets, or Transformers.

7.2. Impact of Sampling Strategies on Classification Performance

This study further compared the effects of two training sample construction strategies, proportional and uniform sampling, on Random Forest performance. The results showed that when the dominant class accounts for a large proportion of the dataset, uniform sampling significantly improves the model's recognition accuracy for the minority classes.

For example, in Tuscany, using proportional sampling, the RF model's accuracy for Grass and Shrub was only 49% and 41%, respectively, while using uniform sampling, these improved to 93% and 74%. In Aosta, the RF model's accuracy for Grass increased from 43% to 97%. In Shanghai, the RF model's accuracy for Crops increased from 38% to 96%. Spatially, uniform sampling produces more refined classification boundaries, particularly in transitional terrain zones and complex mixed areas, significantly enhancing the ability to represent feature distributions.

Therefore, from the perspective of model generalization ability and multi-class recognition balance, the uniform sampling strategy is more suitable for remote sensing classification tasks in highly heterogeneous and multi-type terrain areas. Especially when combined with the robust performance of the random forest model, it can effectively enhance the discrimination ability of the non-dominant class while maintaining the accuracy of the dominant class, thereby obtaining more reliable and more ecologically interpretable classification results.

Bibliography

- [1] Phiri, D., Simwanda, M., Salekin, S., Nyirenda, V. R., Murayama, Y., & Ranagalage, M. (2020). Sentinel-2 data for land cover/use mapping: A review. *Remote sensing*, 12(14), 2291.
- [2] Thanh Noi, P., & Kappas, M. (2017). Comparison of random forest, k-nearest neighbor, and support vector machine classifiers for land cover classification using Sentinel-2 imagery. *Sensors*, 18(1), 18.
- [3] Karra, K., Kontgis, C., Statman-Weil, Z., Mazzariello, J. C., Mathis, M., & Brumby, S. P. (2021, July). Global land use/land cover with Sentinel 2 and deep learning. In 2021 IEEE international geoscience and remote sensing symposium IGARSS (pp. 4704-4707). IEEE.
- [4] Ienco, D., Interdonato, R., Gaetano, R., & Minh, D. H. T. (2019). Combining Sentinel-1 and Sentinel-2 Satellite Image Time Series for land cover mapping via a multi-source deep learning architecture. *ISPRS Journal of Photogrammetry and Remote Sensing*, 158, 11-22.
- [5] Carrasco, L., O'Neil, A. W., Morton, R. D., & Rowland, C. S. (2019). Evaluating combinations of temporally aggregated Sentinel-1, Sentinel-2 and Landsat 8 for land cover mapping with Google Earth Engine. *Remote Sensing*, 11(3), 288.
- [6] Griffiths, P., Nendel, C., & Hostert, P. (2019). Intra-annual reflectance composites from Sentinel-2 and Landsat for national-scale crop and land cover mapping. *Remote sensing of environment*, 220, 135-151.
- [7] Phan, T. N., Kuch, V., & Lehnert, L. W. (2020). Land cover classification using Google Earth Engine and random forest classifier—The role of image composition. *Remote Sensing*, 12(15), 2411.
- [8] Praticò, S., Solano, F., Di Fazio, S., & Modica, G. (2021). Machine learning classification of mediterranean forest habitats in google earth engine based on seasonal sentinel-2 time-series and input image composition optimisation. *Remote sensing*, 13(4), 586.
- [9] Teluguntla, P., Thenkabail, P. S., Oliphant, A., Xiong, J., Gumma, M. K., Congalton, R. G., ... & Huete, A. (2018). A 30-m landsat-derived cropland extent product of Australia and China using random forest machine learning algorithm on Google

- Earth Engine cloud computing platform. *ISPRS journal of photogrammetry and remote sensing*, 144, 325-340.
- [10] Tamiminia, H., Salehi, B., Mahdianpari, M., Quackenbush, L., Adeli, S., & Brisco, B. (2020). Google Earth Engine for geo-big data applications: A meta-analysis and systematic review. *ISPRS journal of photogrammetry and remote sensing*, 164, 152-170.
- [11] Tustison, N. J., Shrinidhi, K. L., Wintermark, M., Durst, C. R., Kandel, B. M., Gee, J. C., ... & Avants, B. B. (2015). Optimal symmetric multimodal templates and concatenated random forests for supervised brain tumor segmentation (simplified) with ANTsR. *Neuroinformatics*, 13, 209-225.
- [12] Wang, Q., Wang, Y., Niu, R., & Peng, L. (2017). Integration of information theory, K-means cluster analysis and the logistic regression model for landslide susceptibility mapping in the Three Gorges Area, China. *Remote Sensing*, 9(9), 938.
- [13] Yang, Y., Yang, D., Wang, X., Zhang, Z., & Nawaz, Z. (2021). Testing accuracy of land cover classification algorithms in the qilian mountains based on gee cloud platform. *Remote Sensing*, 13(24), 5064.
- [14] Javadi, S., Hashemy, S. M., Mohammadi, K., Howard, K. W. F., & Neshat, A. (2017). Classification of aquifer vulnerability using K-means cluster analysis. *Journal of hydrology*, 549, 27-37.
- [15] Anderson, T. K. (2009). Kernel density estimation and K-means clustering to profile road accident hotspots. *Accident Analysis & Prevention*, 41(3), 359-364.
- [16] Gorsevski, P. V., Gessler, P. E., & Jankowski, P. (2003). Integrating a fuzzy k-means classification and a Bayesian approach for spatial prediction of landslide hazard. *Journal of geographical systems*, 5, 223-251.
- [17] Xu, H., Ma, C., Lian, J., Xu, K., & Chaima, E. (2018). Urban flooding risk assessment based on an integrated k-means cluster algorithm and improved entropy weight method in the region of Haikou, China. *Journal of hydrology*, 563, 975-986.
- [18] Pham, B. T., Bui, D. T., Prakash, I., & Dholakia, M. B. (2017). Hybrid integration of Multilayer Perceptron Neural Networks and machine learning ensembles for landslide susceptibility assessment at Himalayan area (India) using GIS. *Catena*, 149, 52-63.
- [19] Pijanowski, B. C., Brown, D. G., Shellito, B. A., & Manik, G. A. (2002). Using neural networks and GIS to forecast land use changes: a land transformation model. *Computers, environment and urban systems*, 26(6), 553-575.
- [20] Zare, M., Pourghasemi, H. R., Vafakhah, M., & Pradhan, B. (2013). Landslide susceptibility mapping at Vaz Watershed (Iran) using an artificial neural network

- model: a comparison between multilayer perceptron (MLP) and radial basic function (RBF) algorithms. *Arabian Journal of Geosciences*, 6, 2873-2888.
- [21] Li, C., Wang, J., Wang, L., Hu, L., & Gong, P. (2014). Comparison of classification algorithms and training sample sizes in urban land classification with Landsat thematic mapper imagery. *Remote sensing*, 6(2), 964-983.
- [22] Rwanga, S. S., & Ndambuki, J. M. (2017). Accuracy assessment of land use/land cover classification using remote sensing and GIS. *International Journal of Geosciences*, 8(04), 611.
- [23] Shalaby, A., & Tateishi, R. (2007). Remote sensing and GIS for mapping and monitoring land cover and land-use changes in the Northwestern coastal zone of Egypt. *Applied geography*, 27(1), 28-41.
- [24] Huang, X., & Jensen, J. R. (1997). A machine-learning approach to automated knowledge-base building for remote sensing image analysis with GIS data. *Photogrammetric engineering and remote sensing*, 63(10), 1185-1193.
- [25] Oxoli, D., Reyes, S. R., Peng, S., Brovelli, M. A., Coetzee, S., Ivanova, I., Leonardi, J. A., Rawal, D., Vacca, G., & Zlatanova, S. (2023). Capacity building for GIS-based SDG indicators analysis with global high-resolution land cover datasets. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLVIII-1/W2-2023, 559–564.
- [26] Zhou, B., Okin, G. S., & Zhang, J. (2020). Leveraging Google Earth Engine (GEE) and machine learning algorithms to incorporate in situ measurement from different times for rangelands monitoring. *Remote Sensing of Environment*, 236, 111521.
- [27] Bar, S., Parida, B. R., & Pandey, A. C. (2020). Landsat-8 and Sentinel-2 based Forest fire burn area mapping using machine learning algorithms on GEE cloud platform over Uttarakhand, Western Himalaya. *Remote Sensing Applications: Society and Environment*, 18, 100324.
- [28] Zhao, Z., Islam, F., Waseem, L. A., Tariq, A., Nawaz, M., Islam, I. U., ... & Hatamleh, W. A. (2024). Comparison of three machine learning algorithms using google earth engine for land use land cover classification. *Rangeland ecology & management*, 92, 129-137.
- [29] Loukika, K. N., Keesara, V. R., & Sridhar, V. (2021). Analysis of land use and land cover using machine learning algorithms on google earth engine for Munneru River Basin, India. *Sustainability*, 13(24), 13758.
- [30] Feizizadeh, B., Omarzadeh, D., Kazemi Garajeh, M., Lakes, T., & Blaschke, T. (2023). Machine learning data-driven approaches for land use/cover mapping and trend analysis using Google Earth Engine. *Journal of Environmental Planning and Management*, 66(3), 665-697.

List of Figures

Figure 1. K-means steps.....	10
Figure 2. Random Forest Construction.	15
Figure 3. MLP Structure.	17
Figure 4. Map of China with Shanghai position.	26
Figure 5. Shanghai Original Image.	26
Figure 6. Shanghai Training Samples.....	28
Figure 7. Map of Italy with Tuscany position.	29
Figure 8. Tuscany DTM Image.	30
Figure 9. Tuscany Original Image.....	30
Figure 10. Tuscany Training Samples.	31
Figure 11. Map of Italy with Aosta position.	32
Figure 12. Aosta DTM Image.....	32
Figure 13. Aosta Original Image.	33
Figure 14. AostaTraining Samples.	33
Figure 15. Military Grid Reference System.....	35
Figure 16. Experimental Workflow.....	39
Figure 17. Shanghai KMeans.	41
Figure 18. Shanghai KMeans Confusion Matrix.....	42
Figure 19. Tuscany KMeans.	44
Figure 20. Tuscany KMeans Confusion Matrix.....	45
Figure 21. Aosta KMeans.	47
Figure 22. Aosta KMeans Confusion Matrix.	49
Figure 23. Shanghai RF.	50
Figure 24. Shanghai RF Confusion Matrix.....	52
Figure 25. Tuscany RF.....	53
Figure 26. Tuscany RF Confusion Matrix.	54
Figure 27. Aosta RF	56
Figure 28. Aosta RF Confusion Matrix.....	57

Figure 29. Shanghai MLP	59
Figure 30. Shanghai MLP Confusion Matrix.....	61
Figure 31. Tuscany MLP.....	63
Figure 32. Tuscany MLP Confusion Matrix.	64
Figure 33. Aosta MLP	66
Figure 34. Aosta MLP Confusion Matrix.....	68
Figure 35. RF Proportional Sampling.	70
Figure 36. Proportional Sampling Confusion Matrix.....	72
Figure 37. Uniform Sampling Confusion Matrix.....	72
Figure 38. RF Proportional Sampling.	73
Figure 39. Proportional Sampling Confusion Matrix.....	75
Figure 40. Uniform Sampling Confusion Matrix.....	75
Figure 41. RF Proportional Sampling	76
Figure 42. Proportional Sampling Confusion Matrix.....	78
Figure 43. Uniform Sampling Confusion Matrix.....	78

List of Tables

Table 1. Comparison of Major Earth Observation Satellites.....	3
Table 2. S2_SR Bands.....	7
Table 3. K-means Usage.....	11
Table 4. K-means Parameter.....	12
Table 5. Random Forest Usage.....	14
Table 6. Random Forest Parameter.....	14
Table 7. MLP Usage.....	19
Table 8. MLP Parameter.....	20
Table 9. Confusion Matrix.....	23
Table 10. Kappa values.....	24
Table 11. Shanghai Training Samples.....	28
Table 12. Tuscany Training Samples.....	31
Table 13. Aosta Training Samples.....	34
Table 14. Shanghai Accuracy Report.....	43
Table 15. Tuscany Accuracy Report.....	46
Table 16. Aosta Accuracy Report.....	49
Table 17. Shanghai Accuracy Report.....	52
Table 18. Tuscany Accuracy Report.....	55
Table 19. Aosta Accuracy Report.....	57
Table 20. Optimal MLP Hyperparameters for Each AOI.....	58
Table 21. Shanghai Accuracy Report.....	61
Table 22. Tuscany Accuracy Report.....	65
Table 23. Aosta Accuracy Report.....	68
Table 24. Shanghai Proportional Sampling.....	69
Table 25. Tuscany Proportional Sampling.....	72
Table 26. Aosta Proportional Sampling.....	75

Acknowledgments

During my study abroad, I received generous support and assistance from many teachers, classmates, friends, and relatives. I would like to extend my heartfelt thanks to my supervisor, **Prof. Ludovico Giorgio Aldo Biagi**, for his insightful guidance and patient mentorship throughout the research process, which encouraged me to continuously explore and grow in the field of remote sensing classification.

I am also deeply grateful to **Dr. Nikolina Zallemi**, whose professional feedback and academic advice were instrumental in shaping the direction and refinement of this thesis.

My sincere appreciation goes to **Politecnico di Milano** for providing an excellent academic environment and a robust research platform, which laid a solid foundation for my graduate studies. I am particularly thankful to my classmates **Ellen Poli** and **Lorenzo Carlassara**, whose kindness and encouragement during my early days in Italy helped me quickly adapt to life and study at the Politecnico.

Finally, I would like to thank my family for their unwavering support, love, and understanding. Their encouragement has been my greatest source of strength and motivation throughout this journey.

E per finire, grazie Politecnico, grazie Milano.

Viva l'Italia!