Ciências ULisboa Faculdade de Ciências da Universidade de Lisboa

Earth Sciences and Energy Department

MULTISPECTRAL REMOTE SENSING

Master's Degree in Geospatial Engineering

LABORATORY PROJECT 1

MACHINE LEARNING FOR LAND USE AND LAND COVER MAPPING USING SENTINEL-2 TIME SERIES	
OBJECTIVE	Develop a complete workflow for the production of a LULC map for a selected area in mainland Portugal using Sentinel-2 imagery from 2023 and machine learning algorithms implemented in Python (scikit-learn)
DURATION	3 March 2026 – 1 April 2026
EVALUATION	Oral presentation – 21 April 2026
BIBLIOGRAPHY	Machine learning for land use and land cover mapping in Google earth engine: A review of methods, applications, and challenges - ScienceDirect Sentinel-2 Land Cover Classification: State-of-the-Art Methods and the Reality of Operational Deployment—A Systematic Review A review of regional and Global scale Land Use/Land Cover (LULC) mapping products generated from satellite remote sensing - ScienceDirect Sentinel-2 Data for Land Cover/Use Mapping: A Review Land Use and Land Cover Mapping Using Sentinel-2, Landsat-8 Satellite Images, and Google Earth Engine: A Comparison of Two Composition Methods

LULC mapping workflow

STEP 1 **Area of interest (AOI) definition and data collection**	**1.1 Define the AOI** Identify a representative region within mainland Portugal Ensure land cover types diversity (urban, agriculture, forest, water, shrubs, etc.) **1.2 Download Sentinel-2 Level-2A Images (Copernicus/ESA)** Year = 2023 Cloud cover < 10% At least 8 acquisition dates well distributed across seasons (Winter/Spring/Summer/Autumn) **1.3 Additional datasets for reference and training** COS2023 COSc2023 OrtoSat 2023 LPIS 2023
STEP 2 **Sentinel-2 time series generation**	**2.1 Preprocessing** Atmospheric correction (already applied for Level-2A) Cloud and shadow masking (using SCL band) Resampling to a common spatial resolution (10 m) **2.2 Variables to include (spectral *bands and* derived features)** **Original spectral bands** B2, B3, B4, B8 (10 m) B5, B6, B7, B8A, B11, B12 (20 m) **Spectral indices** NDVI, EVI, NDWI, NBR, NDMI Red-edge indices (RENDVI, CIred-edge, NDVIre) Built-up index (NDBI) **Texture measures (useful for urban/agriculture separation)** GLCM variance, homogeneity, contrast (based on NDVI or specific bands*, such as B8 or B11*) **Temporal metrics (for each pixel over 2023)** min, max, mean, median*, standard deviation* of each index percentiles (10, 25, 75, 90) *NDVI annual* amplitude **2.3 Construction of the time series cube** Stack all dates *and all* derived layers into a single multidimensional dataset Export as a raster stack (GeoTIFF) or *NumPy* array for use in Python

STEP 3 **Nomenclature definition and samples generation**	**3.1 Nomenclature** Use 8 classes as in the Level 1/2 of the COSc2023 nomenclature[1]: 10. Artificial surfaces 21. Agriculture 30. Forest 41. Shrubs 42. Herbaceous spontaneous vegetation 50. Bare soil 61. Wetlands 62. Water **3.2 Sampling generation** Use polygons from COS/LPIS, but keep samples well inside polygons (buffer inward 20 m to avoid mixed pixels) Generate 2 random points inside each polygon to generate a minimum of 300 to 500 points per class (buffer outward 40 m and compute the median)
STEP 4 **Feature values extraction at samples**	**4.1 Verify that the samples are in the same CRS as the stack** Set the CRS as EPSG: 32629 - WGS84/UTM zone 29N **4.2 Use Python to sample the raster stack values at points** Overlay the sample points on the Sentinel-2 time-series stack to extract all feature values for every acquisition date (and any extra metrics) into a table **4.3 Create a table** Columns = features (as many as features in the stack) Target column= class label (10, 21, 30, 41, 42, …, 62) Optional columns = Easting, Northing Rows = points (as many as the sample points)
STEP 5 **Three machine learning (ML) models training (scikit-learn)**	**5.1 Choose at least 3 ML models** **Some ML model examples** RandomForestClassifier Gradient Boosting SupportVectorMachine (SVM): this model requires StandardScaler **5.2 Split rule** 75% train / 25% test, but do it in a spatially-aware way **5.3 Class weighting and cross-validation** Use class_weight="balanced" when classes are imbalanced Use cross-validation (even 3-fold) for hyperparameter sanity

[1] https://www.dgterritorio.gov.pt/sites/default/files/documentos-publicos/Nomenclatura_COSc.pdf

STEP 6 **Time series classification**	**6.1 Read the raster stack and reshape it to n_pixels x n_features** Convert the raster time series stack into a two-dimensional feature matrix to apply the trained machine learning models to all pixels in the study area **6.2 Write the prediction GeoTIFF** Convert the classification results back to raster format and export as a GeoTIFF to produce the final LULC map **6.3 Compute class membership (optional) and write probability GeoTIFF** Compute class membership probabilities using the trained model and export as a multi-band GeoTIFF, where each band represents the likelihood of a given land cover class for every pixel
STEP 7 **Accuracy assessment**	**7.1 Compute accuracy metrics based on the held-out 25% test samples** Confusion matrix Overall accuracy Precision, Recall, F1-score per class Macro and weighted F1-scores Samples per class
STEP 8 **Final map production**	**8.1 Choose the model with best macro F1** Select the model with the highest macro F1-score as the final classifier, ensuring balanced performance across all land cover classes regardless of class size **8.2 Import the final GeoTIFF to QGIS** Import the selected classification result into QGIS for visualization, cartographic refinement, and final map production Apply post-classification spatial filtering (majority or sieve filter) to reduce the salt-and-pepper effect and improve the visual coherence of homogeneous land cover areas **8.3 Use map layout in QGIS** Use the QGIS Layout Manager to design the final cartographic product, ensuring clear visualization and professional presentation of the classification results Legend with class names/colors Scale bar, north arrow, title Small inset of Portugal + AOI footprint Accuracy summary box (OA + macro F1)