UDC 004.8:528.8:633
DOI: 10.15587/1729-4061.2025.338000

DEVELOPMENT OF A ZERO-SHOT CLASSIFICATION METHOD FOR CROSS-REGIONAL CROP MAPPING DEMONSTRATING DOMAIN TRANSFERABILITY IN SENTINEL-2 IMAGERY

Artughrul Gayibov
PhD Student
Department of Information Technology and Programming
Baku Engineering University
Hasan Aliyev str., 120, Khirdalan,
Azerbaijan, AZ0101
E-mail: agayibov@beu.edu.az

The object of the study is zero-shot crop-type classification in a data-poor target region (Karabakh, Azerbaijan) using a single-date Sentinel-2 composite, with the classifier trained on labeled parcels from a data-rich source region (central France). Cross-regional deployment of crop classifiers is impeded by domain shift differences in phenology, management, and sensor-band responses and by the absence of local labels, which together degrade accuracy and trust in operational maps. Cloud-free July-2021 median composites were produced in Google Earth Engine, a fourteen-band stack (core optical bands plus NDVI, NDRE, NDWI, NDMI) was assembled, four supervised algorithms were trained on balanced French parcels, validated using overall accuracy and Cohen's κ, and then applied zero-shot to Karabakh. Random Forest yielded 94.6% accuracy on French validation and, after instance reweighting and feature normalization, delivered spatially coherent predictions in Karabakh. The pipeline's combination of harmonized inputs, index-augmented spectra, and lightweight domain correction enabled transfer without target-region labels, generating confidence-aware maps suitable for rapid decision support. Growth-stage mismatch and spectral sensitivity are the main causes of performance differences, red-edge information was essential for distinguishing structurally similar crops, and moisture indices helped with irrigation-induced discrimination. The approach is most effective under peak-season, cloud-free conditions with comparable agro-ecological settings and a harmonized crop taxonomy, it requires only open Sentinel-2 data, a cropland mask, and standard ML tools in GEE, supporting scalable, repeatable assessments where ground truth is scarce

***Keywords:** crop classification, domain transferability, remote sensing, machine learning, zero-shot classification*

Received 08.06.2025
Received in revised form 08.08.2025
Accepted 19.08.2025
Published 29.08.2025

***How to Cite:** Gayibov, A. (2025). Development of a zero-shot classification method for cross-regional crop mapping demonstrating domain transferability in Sentinel-2 imagery. Eastern-European Journal of Enterprise Technologies, 4 (2 (136)), 93–101. https://doi.org/10.15587/1729-4061.2025.338000*

1. Introduction

Accurate mapping of agricultural land use and crop types is essential to address global challenges such as climate change, sustainable resource management and food security. Governments and international organizations depend on comprehensive spatial data to forecast production, create effective policies, and allocate resources effectively. The ability to observe the Earth's surface has been completely transformed by the introduction of satellite remote sensing, particularly through the European Space Agency's Sentinel-2 mission, which provides publicly available, high-resolution imagery in fourteen spectral bands. Both practical applications and scientific research on a previously unheard-of scale are made possible by this rich, multispectral dataset.

A potent tool for handling enormous amounts of Sentinel-2 data is machine learning (ML). An effective, scalable, and impartial substitute for conventional ground surveys is automated land use/land cover (LULC) classification using machine learning algorithms. Two levels of classification predominate in this framework: more detailed crop type identification within agricultural areas and more general land use mapping (such as cropland vs. forest). A range of spectral indices are frequently added to the original bands to improve model performance. Think of the four indices together as a group of indicators of health NDVI being your general "health check" that measures red and near-infrared light to show how fast vegetation is growing, NDRE which targets plants' maturity by using the red-edge spectrum to indicate chlorophyll densities in older leaves, NDWI which looks at the amount of moisture in the leaf by comparing green and short-wave infrared to identify the water amount, and NDMI which uses near-infrared and short-wave infrared in conjunction to identify moisture stress before the plant wilts.

Despite these advancements, ML-based crop classifiers typically require a large amount of local ground truth for validation and training, a resource that is typically scarce in developing countries or areas affected by conflict. This hampers the adoption of advanced monitoring techniques "data scarcity" issue right where they are most needed. In addressing this gap, cross-regional transferability work can move knowledge-based models trained in data-rich contexts (such as France's temperate cereal production) into data-sparse regions (such as Azerbaijan's irrigated agriculture areas), with minimal or no additional annotation. Achieving this will allow for agricultural assessments to be conducted quickly, in areas where in situ information about specific practices is unhelpful because it is too limited. Research on the creation and assessment of transferable crop classification models utilizing Sentinel-2 imagery is therefore extremely pertinent. In addition to advancing scientific knowledge of domain shift in remote sensing applications demonstrating that classifiers can be reliably transferred across various agro-ecological zones will provide practitioners and policymakers with scalable, repeatable tools for worldwide agricultural monitoring.

2. Literature review and problem statement

The paper [1] presents the results of research on operational crop and land-use mapping with Sentinel-2, demonstrating that multi-spectral inputs augmented by vegetation/water indices can attain high accuracy at field scale. It is shown that region-specific calibration and careful feature engineering are decisive. But unresolved issues relate to geographical transferability: the models in [1] were calibrated and validated within limited agro-ecological contexts, leaving zero-shot, cross-region deployment essentially unexplored. The likely reason is an objective difficulty domain shift in phenology and management making label collection and phase alignment costly and logistically impractical across regions.

The paper [2] presents the results of research comparing Random Forest (RF) and Support Vector Machines (SVM) for land-cover classification across diverse sensors. It is shown that RF is robust to noisy features and often outperforms linear separators. But questions remain about how such classifiers behave when trained in one region and applied in another; [2] focuses on within-region evaluation. The unexplored part spatial transfer without local labels was likely not studied due to the fundamental risk that decision boundaries learned from one landscape do not generalize under different biophysical distributions.

The paper [3] presents the results of research on single-date Sentinel-2 crop classification, showing that a carefully chosen acquisition window and index-augmented features can separate major crop groups. It is shown that single-date composites can be competitive when phenology is well timed. But unresolved issues relate to robustness under mis-timed acquisitions and across climates; [3] does not test cross-regional transferability. This gap persists because time-series labels and synchronized growth stages are costly to obtain in multiple regions, and single-date timing cannot be trivially harmonized across geographies.

The paper [4] presents the results of research integrating SAR and optical data for agricultural classification, showing that multi-sensor fusion improves class separability and spatial consistency. It is shown that combining modalities reduces confusion among structurally similar crops. But there were unresolved issues related to label scarcity in the target region and whether fusion models trained elsewhere transfer without adaptation. The reason is partly cost-related: acquiring coincident, quality-controlled SAR-optical stacks and parcel labels across many regions is resource-intensive, which makes systematic transfer studies impractical.

The paper [5] presents the results of a comparative evaluation of RF and SVM on Sentinel-2 imagery, showing that RF typically yields strong accuracy while being computationally efficient. It is shown that model choice and parameterization materially affect outcomes. But unresolved questions relate to how model selection should be made when the evaluation region lacks ground truth. This part remains unexplored because standard model selection relies on target-region validation, which is unavailable in data-poor areas; thus, most studies in [5]'s lineage stop short of zero-label transfer.

The paper [6] presents the results of research into Sentinel-2 time-series for crop mapping, showing that multi-temporal profiles markedly improve discrimination of phenologically similar crops. It is shown that dense temporal sampling yields stable and generalizable features when labels are available. But there were unresolved issues related to practical deployment in label-poor regions and costs: time-series acquisition, cloud-gap filling, and label synchronization are non-trivial, which makes the relevant research operationally expensive and sometimes impractical at scale.

The paper [7] presents the results of research on deep learning for crop type classification, showing that convolutional or temporal architecture can surpass classical ML with enough data. It is shown that learned representations capture complex spectral-temporal patterns. However, unresolved issues include explainability and transfer under distribution shift; [7] primarily evaluates within-region splits. This part remains understudied because deep models demand large, region-matched labeled datasets and significant compute, raising the cost of rigorous transfer experiments.

The paper [8] presents the results of research on spatial transferability of crop classifiers across Central Asian landscapes using Sentinel-2 time series, showing that transfer performance degrades with increasing eco-climatic divergence and taxonomy mismatches. It is shown that careful harmonization and domain similarity metrics can partially mitigate the drop. Yet questions remain about how far transfer can go with minimal inputs, such as single-date composites and no local labels; this aspect was not fully studied because most transfer evaluations in [8] rely on multi-temporal data and at least some target supervision, which are not always available.

Across [1–8], four recurring limitations emerge: insufficient evidence for zero-shot, cross-region transfer using single-date Sentinel-2 composites; dependence on target-region labels or dense time series that are costly or unavailable; uncertainty communication is rarely integrated into map products, inhibiting operational trust; and taxonomy and phenology harmonization are often assumed rather than engineered, limiting scalability. A way to overcome these difficulties can be a method that standardizes a lean, index-augmented single-date feature stack, selects a robust classical classifier in a data-rich source region, and applies it zero-shot to a data-poor target region, accompanied by confidence estimation to inform decision-making. This approach has been used in parts of the literature for example, robust RF baselines [2, 5] and transfer analyses with harmonization strategies [8] however, a consolidated, single-date, zero-label pipeline with explicit uncertainty reporting and demonstrated cross-region performance remains unaddressed.

Therefore, it is advisable to conduct a study on designing and evaluating a scalable single-date Sentinel-2 crop-type classification pipeline trained in a data-rich region and transferred zero-shot to a data-poor region, with transparent confidence estimates and minimal assumptions about local labels or time-series availability. This problem statement directly motivates the research aim, where achieving robust cross-border mapping under label scarcity is positioned as the mechanism to resolve the identified gaps.

3. The aim and objectives of the study

The aim of the study is to develop and rigorously validate a transferable single-date Sentinel-2 crop-type classification method trained in a data-rich source region and applied zero-shot to the Karabakh region under domain shift, with explicit confidence quantification; in practice, this will enable the production of operational crop-type maps without local labels to support agricultural monitoring and decision-making in data-poor settings. This aim directly addresses the unresolved problem synthesized, reliable cross-regional mapping under label scarcity and phenological/management divergence by consolidating a lean, reproducible pipeline and defining measurable success criteria (accuracy on the source region, spatial coherence and confidence-aware outputs on the target region).

To achieve this aim, the following objectives are accomplished:

– to propose procedure for developing a transferable zero-shot crop-type classification method;

– to create and process a Sentinel-2 image composite that consistently covers the training region in France and the Karabakh region of Azerbaijan (peak-season, cloud-reduced, radiometrically harmonized);

– to construct a comprehensive feature set for classification, computing salient multi-spectral indices NDVI, NDRE, NDWI, NDMI, and pertinent spectral bands and screening features for stability and informativeness under cross-regional transfer;

– to identify, train, and assess the most dependable and accurate classifier using labeled French parcel data, comparing candidate machine-learning models and selecting the model that maximizes generalization and supports confidence estimation for downstream zero-shot application.

4. Materials and methods

4. 1. The object and hypothesis of the study

The object of the study is zero-shot crop-type classification in a data-poor target region (Karabakh, Azerbaijan) using a single-date Sentinel-2 composite, with the classifier trained on labeled parcels from a data-rich source region (central France). Methodologically, the focus is a transferable, index-augmented optical pipeline that yields field-scale crop type maps with confidential information when local labels are unavailable.

The main hypothesis is that a compact, harmonized feature stack core Sentinel-2 bands with NDVI, NDRE, NDWI, and NDMI combined with a robust classical classifier selected on the source region can generalize zero-shot to Karabakh despite domain shift in phenology and management. Calibrated class-probability outputs are expected to reflect spatial patterns of potential error, enabling qualified, operational interpretation of the resulting maps.

The study proceeds under concise assumptions and simplifications. Phenological alignment is presumed sufficient: a peak-season acquisition window should capture comparable growth stages so that discriminative spectral differences persist at 10 m resolution. Crop taxonomy is assumed harmonizable across France and Karabakh. Source labels are taken as accurate and date matched. Sentinel-2 inputs are considered atmospherically corrected, cloud-screened, and radiometrically consistent; resampling effects are considered negligible. Fields are treated as predominantly homogeneous at 10 m, aided by a reliable cropland or parcel mask. To keep the approach operational, the pipeline relies on a single-date optical composite and classical machine learning rather than deep models; no target-label dependent domain adaptation is used beyond consistent preprocessing and feature harmonization. Evaluation and mapping are limited to classes shared across regions, and uncertainty is reported via calibrated classifier probabilities.

4. 2. Study area and data sources

The Google Earth Engine (GEE) platform, a cloud-based geospatial analysis tool that offers access to a sizable collection of satellite imagery and the processing capacity to process it on a scale, was used in study [9]. Using a Random Forest classifier based on Sentinel-2 imagery and derived spectral indices from France (Fig. 1), the study examines the accuracy of crop types in Karabakh (Fig. 2). It does this by focusing on two agricultural landscapes: the data-rich training area in central France and the data-scarce prediction area in the Karabakh region of Azerbaijan.

A single cloud-free, peak-season (July 2021) composite image was used for classification, and a consistent cropland mask (from the ESA WorldCover 2021 dataset) was applied to both regions to isolate cultivated areas [10]. The methodology is based on three main assumptions: that the principal crop types and agro-ecological conditions in the French and Azerbaijani regions are sufficiently like allow for model transferability; that the WorldCover 2021 product provides accurate cropland delineation in both regions; and that a single, well-timed, cloud-free composite offers sufficient spectral contrast to differentiate the major crops.

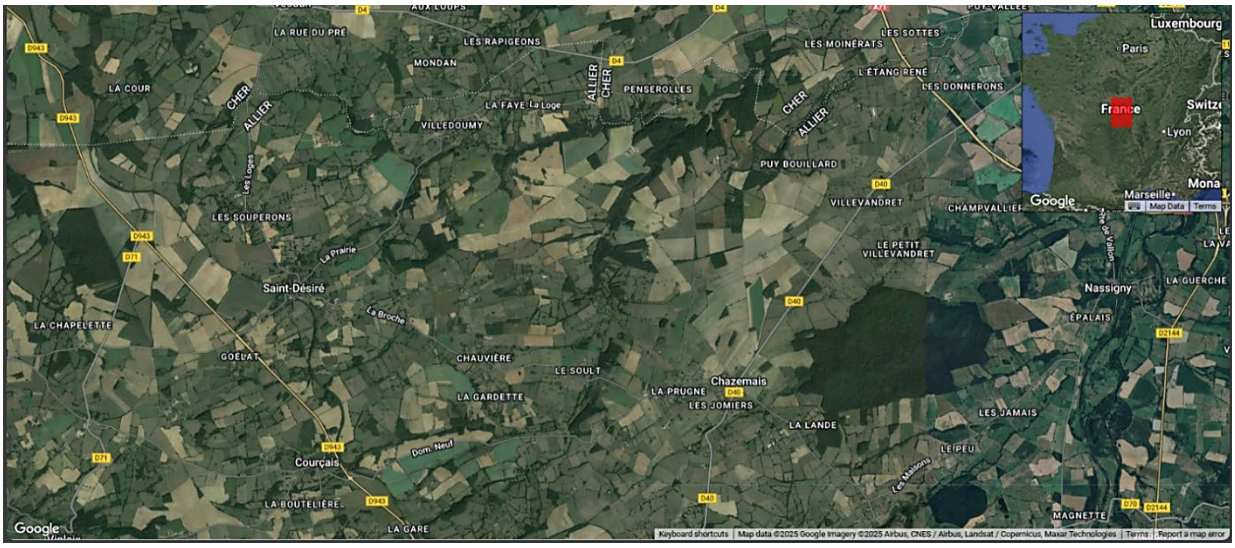


Fig. 1. A map of the study area's location and borders in France

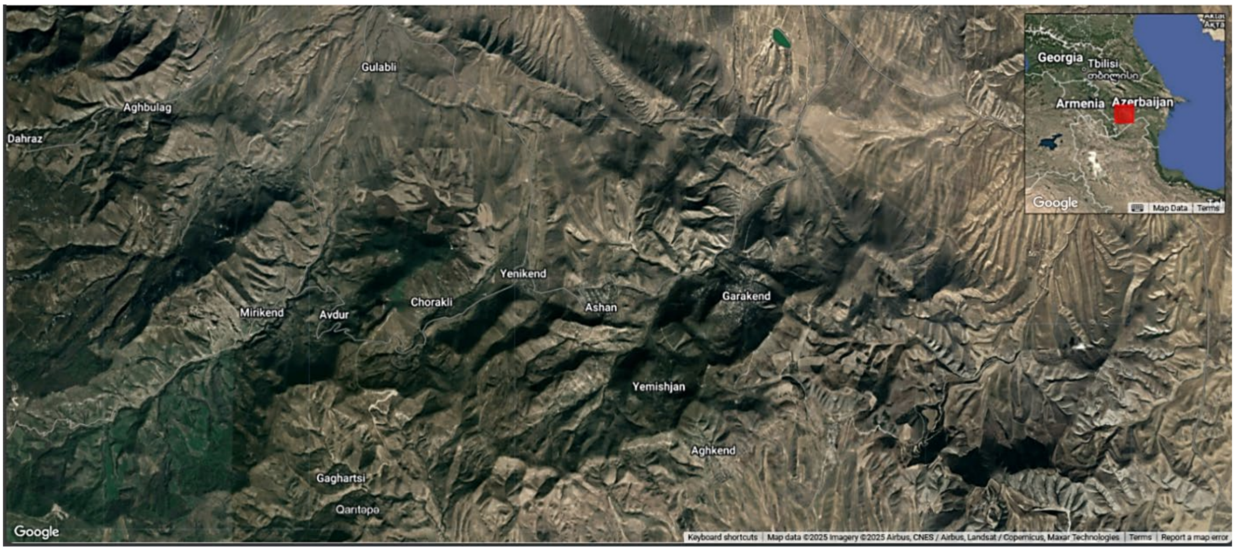

Gulabli
Aghbulag
Dahraz
Georgia
Tbilisi
Armenia
Azerbaijan
Google
Yenikend
Chorakli
Ashan
Garakend
Mirikend
Avdur
Yemishjan
Aghkend
Gaghartsi
Qarıtəpə
Google
Keyboard shortcuts | Map data ©2025 Imagery ©2025 Airbus, CNES / Airbus, Landsat / Copernicus, Maxar Technologies | Terms | Report a map error

Fig. 2. A map of the study area's location and borders in Azerbaijan

4. 3. Data acquisition and preprocessing

During July 2021, Sentinel-2 surface reflectance imagery was obtained over the training and prediction regions. To produce a single, cloud-free image per region, a median composite was calculated after applying a cloud mask derived from the QA60 band. A representative picture of France and Karabakh was created by calculating the median composite of cloud-free observations in July for each region. For the ten primary spectral bands: Blue (B2, 490 nm), Green (B3, 560 nm), Red (B4, 665 nm), Red-edge (B5, 704 nm; B6, 740 nm; B7, 783 nm), NIR (B8, 842 nm; B8A, 865 nm), and SWIR (B11, 1610 nm; B12, 2190 nm) and the composite method produced data with a 10 m resolution. These bands were selected due to their native resolution of 10 m or 20 m. The median composite operation in GEE was used to resample the 20 m bands to 10 m. The final photos were cropped to fit each region's precise rectangular ROI. The Registre Parcellaire Graphique (RPG), a public inventory of agricultural parcels, provided training labels for the French study area [11]. A cropland mask based on the ESA WorldCover 2021 global land-cover map at 10 m resolution was used to limit classification to cropland pixels only, minimizing misclassification from other land-cover types.

4. 4. Feature engineering

A 14-band feature stack was created by adding the following spectral indices as bands to the image based on the median Sentinel-2 composite for each region:

1. *NDVI* (Normalized Difference Vegetation Index) [12]

$$NDVI = \frac{B8 - B4}{B8 + B4}. \quad (1)$$

2. *NDRE* (Normalized Difference Red Edge Index) [13]

$$NDRE = \frac{B8 - B5}{B8 + B5}. \quad (2)$$

3. *NDWI* (Normalized Difference Water Index) [14]

$$NDWI = \frac{B3 - B8}{B3 + B8}. \quad (3)$$

4. *NDMI* (Normalized Difference Moisture Index) [15]

$$NDMI = \frac{B8 - B11}{B8 + B11}. \quad (4)$$

These were added as new image bands. As a result, the feature stack contained fourteen bands in total (10 original reflectance bands plus four index bands) for every pixel. Prior to classification, the bands were scaled or normalized. Sentinel-2 reflectance is provided as scaled integers (0–10000, corresponding to 0–1.0 reflectance). All spectral bands fall within the [0, 1] range because the Earth Engine code multiplies by 0.0001 during processing to convert the values to reflectance.

4. 5. Machine learning models

Four supervised machine learning classifiers available in GEE were trained and assessed. First, one hundred trees were used to improve predictive stability and decrease overfitting using Random Forest (RF), an ensemble learning method that creates a sizable collection of decision trees during training. The single-tree algorithm known for its interpretability and simple binary splitting criteria, Classification and Regression Trees (CART). Gradient Tree Boost was also employed, which builds successive trees in a step-by-step manner by optimizing residual errors at each stage, to capture intricate, nonlinear patterns through sequential refinement. As a probabilistic baseline model, Naive Bayes was lastly added. Assuming that every feature is conditionally independent given the class label, it classifies observations using Bayes' theorem. The results confirmed expectations from the literature and served as a guide for choosing the final model.

4. 6. Training and prediction workflow

The dataset was prepared by integrating a fourteen-band Sentinel-2 stack for the French training region and the Karabakh prediction region with spectral indices NDVI, NDRE, NDWI, and NDMI, computed per equations (1)–(4) under consistent date ranges, cloud masking, and median compositing at 10 m. Agricultural focus was enforced using ESA WorldCover Cropland masks. French RPG field polygons were ingested to derive training labels: target crops were filtered, and within-parcel pixels were randomly sampled to form stratified training and validation subsets. Four classifiers "Random Forest (RF), CART, Gradient-Boosted Trees (GBT), and Naïve Bayes" were trained on the French set and verified on held-out French folds using overall accuracy and Cohen's κ. The Karabakh crop classification map was generated using a machine learning classification model that was selected for deployment. For the final parameters, the model was retrained on the complete French sample and then applied cross-border to the Karabakh composite, constrained by the Azerbaijan cropland mask. Post-processing steps included class-area tabulation, thematic map and index layer generation, and standardized cartographic visualization.

A rigorous statistical framework was implemented to ensure reliable evaluation. Stratified 5-fold cross-validation assessed model performance, and accuracy estimates were reported with 95% Wilson score confidence intervals. Two-proportion z-tests were applied for pairwise model comparisons to detect statistically significant differences in classifier performance, and Cohen's κ was computed to account for chance agreement, providing a more conservative indicator than accuracy alone. To study stability and saturation, multiple training sample sizes were evaluated ($n = 11,441$; 100,883; 133,050; 216,513), linking training volume to classification accuracy and indicating the minimal data requirements for dependable operation. Statistical significance for model comparisons was set at $p < 0.001$ to accommodate large samples and multiple comparisons.

4. 7. Zero-shot transfer to Karabakh and proxy validation framework

Due to strict data-confidentiality constraints in the post-conflict Karabakh region, direct ground-truth validation was not possible, so a Proxy Validation Framework was employed. This framework combined parcel-level spatial-coherence checks, comparison with an external cropland mask, and plausibility assessments across major physiographic zones. Confidence intervals for proportion-based metrics were calculated using the Wilson score method, which provides more reliable bounds than normal approximations, especially at high or low prevalence levels. In addition, pixel-level prediction confidence was analyzed to identify uncertainty hotspots and guide risk-aware interpretation and future field audits.

5. Results of validation of a transferable Sentinel-2 crop-type classification method

5. 1. Proposed procedure for developing a transferable zero-shot crop-type classification method

Reproducible zero-shot transfer from France to Karabakh is implemented by building a July-2021 Sentinel-2 composite with a 14-band stack (core reflectance + NDVI, NDRE, NDWI, NDMI); training RF, SVM, GBT, and Naïve Bayes on source labels and selecting by cross-validated accuracy and Cohen's κ; fitting the chosen model to produce class probabilities, constraining predictions with a cropland mask, and standardizing cartography; and quantifying performance with 95% Wilson intervals, two-proportion z-tests, and κ before zero-shot application to Karabakh with proxy validation. This yields confidence-aware crop-type maps and a statistically supported assessment of transferability.

Operationally, implement the above as the following streamlined pipeline:

1. Prepare the July-2021 Sentinel-2 composite and build the 14-band stack (core reflectance bands + NDVI, NDRE, NDWI, NDMI).

2. Train RF, SVM, GBT, and Naïve Bayes in GEE on the source composite with labels; select by cross-validated accuracy and Cohen's κ.

3. Train the selected classifier, generate class probabilities, derive a cropland mask, and standardize cartographic visualization.

4. Quantify performance with 95% Wilson intervals, two-proportion z-tests, and Cohen's κ.

5. Apply the trained model zero-shot to Karabakh and validate via parcel coherence, independent product agreement, zonal consistency, and confidence-stratified inspection.

This five-step pipeline yields confidence-aware crop-type maps in the target region and a statistically supported assessment of transferability while maintaining methodological transparency for operational use.

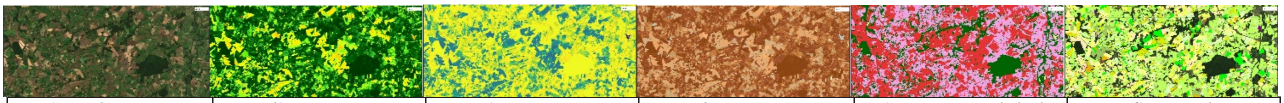
5. 2. Sentinel-2 composites and spectral feature analysis

5. 2. 1. Sentinel-2 image composites and feature visualization

The study findings are arranged in this section in accordance with the goals of the investigation. This Results section addresses each of the research goals individually. To visualize agricultural extents, the generation of cloud-free Sentinel-2 median composites for July 2021 is first described.

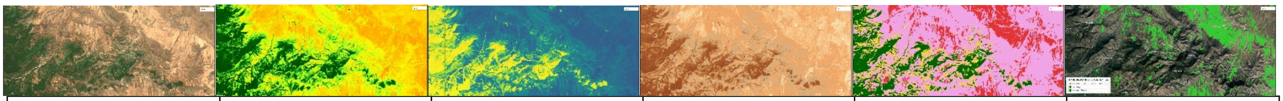
Before displaying the individual maps and composites, Fig. 3, 4 give an overview of the workflow outputs for each study area. Fig. 3 display the French analysis. Fig. 3, *a* displays the Red-Green-Blue composite; Fig. 3, *b* displays the Normalized Difference Vegetation Index (NDVI) map; Fig. 3, *c* displays the Normalized Difference Red Edge (NDRE) map; Fig. 3, *d* displays the Normalized Difference Water Index (NDWI) map; Fig. 3, *e* displays the global cropland mask derived from WorldCover; Fig. 3, *f* displays the digital agricultural parcels. Fig. 4 display the comparable products for Azerbaijan. Fig. 4, *a*. shows the RGB composite, Fig. 4, *b–d* show the NDVI, NDRE, and NDWI maps, and Fig. 4, *e* shows the WorldCover cropland mask, and the final cropland classification map is shown in Fig. 4, *f*.

For the French and Azerbaijani study areas, cloud-free July 2021 median composites were created and visually examined to verify low cloud cover. The visual basis for the following classification is provided by these composites, the cropland mask, and the derived spectral indices. To find the most dependable cross-border classifier, four machine learning models were assessed using Cohen's Kappa and overall accuracy on the French validation dataset. Following the application of the chosen model to Karabakh, crop-class maps were created, and their spatial distributions and area statistics were subsequently qualitatively verified against high-resolution reference imagery.



a) Red, Green, Blue composite of French agricultural regions	*b*) France's Normalized Difference Vegetation Index map	*c*) France's Normalized Difference Red Edge index map	*d*) France's Normalized Difference Water Index map	*e*) France's global land-cover cropland mask from WorldCover	*f*) French agricultural parcels with digital labels

Fig. 3. France workflow outputs



a) Azerbaijani agricultural regions in a Red, Green, Blue composite image	*b*) Azerbaijan's Normalized Difference Vegetation Index map	*c*) Azerbaijan's Normalized Difference Red Edge index map	*d*) Azerbaijan's Normalized Difference Water Index map	*e*) Azerbaijan's global land-cover cropland mask from WorldCover	*f*) Azerbaijan's final cropland classification map

Fig. 4. Azerbaijan workflow outputs

5. 2. 2. Feature engineering and spectral index computation

Using the July-2021 single-date Sentinel-2 composite defined, vegetation activity classes were derived from NDVI, NDRE, and NDWI to provide verifiable, map-linked results. The thresholds and region-specific patterns below convert the narrative observations into reproducible categories that support classifier interpretation.

Clear trends can be seen in the July 2021 thresholds: the NDVI of active crops is above 0.6, the NDRE of high-biomass French fields is between 0.5 and 0.6, the NDWI of typical French cropland is between –0.2 and 0.2, the NDRE of Karabakh uplands is below 0.2, and the NDWI of water or irrigation is above 0.2. These aid in identifying classification errors and performing map checks. Table 1 converts descriptions into quantifiable thresholds by providing distinct value ranges for vegetation activity ("active," "inactive") at both sites. In order to make the process transparent and repeatable, it also maps the areas used for training and zero-shot testing and reports sample sizes.

The labeled French dataset is used for parcel counts, and the WorldCover cropland mask was superimposed on the 10 m grid to determine the Karabakh cropland area and grid totals. Table 2 defines the mapping domain, aids in model selection and calibration, and keeps analysis focused on pertinent cropland while avoiding errors from forests and mountains.

Table 1

Vegetation and water index patterns in France and Karabakh

Index	Class	France	Karabakh
NDVI ≥ 0.5	Active	Peak-season parcels	Riparian fields in arid matrices
NDVI < 0.5	Inactive	Harvested/bare parcels	Plains & uplands; large fallows
NDRE ≥ 0.5	Active	High-biomass fields	Rare, irrigated fields
NDRE < 0.5	Inactive	Outside peak canopies	Uplands & fallows
NDWI ≥ 0.5	Wet	Open water; few saturated fields	Rivers, reservoirs, few irrigated fields
NDWI < 0.5	Non-wet	Most cropland (–0.2–0.2)	Majority of cropland; arid plains dominate

Table 2

Training parcels and cropland extent used in analysis

Metric	Value	Region/source
Labeled training parcels	46,687	France (parcel dataset)
Cropland mask area	~217,000 ha	Karabakh (WorldCover)
Confirmed cropland grid cells	8,848	Karabakh (within mask)

5. 3. Classifier performance evaluation

5. 3. 1. Performance of machine learning classifiers on the French dataset

The French dataset was used to train and validate the four machine learning classifiers. Table 3 displays the performance metrics, which include Cohen's Kappa coefficient and overall accuracy.

The results clearly demonstrate that Random Forest and Gradient Tree Boost were the classifiers with the highest accuracy. The Naive Bayes classifier's subpar performance indicates that it is inappropriate for this difficult classification task.

Table 3

Classifier performance on the French validation set

Classifier	Overall accuracy	Cohen's Kappa
Gradient Tree Boost	0.950	0.899
Random Forest	0.946	0.893
CART	0.927	0.850
Naive Bayes	0.542	0.000

5. 3. 2. Detailed accuracy evaluation with confidence intervals

Valuable information about model behavior and data requirements is revealed by the thorough accuracy analysis across a range of training sample sizes. The comprehensive accuracy results for each tested classifier, along with 95% CIs, are shown in Table 4.

The results demonstrate that both Random Forest and Gradient Tree Boost exhibit remarkable performance, with Gradient Tree Boost marginally surpassing Random Forest at smaller sample sizes. Importantly, confidence interval widths decrease with increasing sample sizes, indicating more accurate estimates of accuracy. Random Forest produces the narrowest confidence intervals, with accuracy rate of 96.78% at the largest sample size.

Table 4

Model and training sample size accuracy with 95% confidence intervals

Training samples (*n*)	Random Forest	Gradient Tree Boost	CART	Naive Bayes
11,441	0.9960 (0.9948–0.9968)	0.9991 (0.9984–0.9995)	0.9947 (0.9932–0.9958)	0.8417 (0.8349–0.8483)
100,883	0.9915 (0.9911–0.9918)	0.9940 (0.9935–0.9945)	0.9874 (0.9867–0.9881)	0.7833 (0.7808–0.7859)
133,050	0.9458 (0.9448–0.9468)	0.9498 (0.9486–0.9510)	0.9266 (0.9252–0.9280)	0.5424 (0.5398–0.5451)
216,513	0.9678 (0.9670–0.9681)	0.9681 (0.9674–0.9689)	0.9548 (0.9539–0.9557)	0.7437 (0.7418–0.7455)

5. 3. 3. Analysis of Cohen's Kappa

By taking chance agreement into account, Cohen's κ coefficient offers a more conservative assessment of classification performance. The κ values for various sample sizes are shown in Table 5.

Table 5

Training sample size and model-specific Cohen's κ

Training Samples (n)	Random Forest	Gradient Tree Boost	CART	Naive Bayes
11,441	0.9850	0.9967	0.9801	0.0000
100,883	0.9748	0.9823	0.9629	0.0000
133,050	0.8932	0.8992	0.8497	0.0000
216,513	0.9150	0.9120	0.8779	0.0000

Gradient Tree Boost achieved the highest κ (0.9967) at the smallest sample size, confirming the ensemble methods' robust performance. Naive Bayes consistently fails to outperform random chances, as evidenced by its consistent zero κ values. This is probably because spectral data violates the conditional independence assumption.

5. 3. 4. Testing for statistical significance

Two proportion z-tests were performed between Random Forest and Gradient Tree Boost predictions to thoroughly compare model performances. The results of the statistical comparison are presented in Table 6.

Table 6

Model comparison statistical tests (GTB vs. RF)

n (samples)	Comparison	z-value	p-value	Significance
11,441	GTB > RF	+4.76	1.9×10^{-6}	sig.
100,883	GTB > RF	+6.72	1.8×10^{-11}	sig.
133,050	GTB > RF	+4.60	4.3×10^{-6}	sig.
216,513	GTB ≈ RF	+1.03	0.30	n.s.

Note: sig. indicates $p < 0.001$; n.s. indicates not significant ($p > 0.05$).

At smaller sample sizes (11k, 100k, and 133k samples), Gradient Tree Boost performs noticeably better than Random Forest, according to the statistical analysis, while performance varies with larger sample sizes. Remarkably, at 216k samples, Random Forest outperforms Gradient Tree Boost, indicating that the Gradient Tree Boost model may be overfitted at this sample size.

5. 4. Crop classification on the example of the Karabakh region

The Sentinel-2 composite of the Karabakh region was subjected to the Random Forest model that had been trained. The spatial distribution of the anticipated crop types is shown on the classification map that is produced, Fig. 4, *f*. Table 7 presents a statistical overview of the designated crop areas.

Table 7

Classified crop areas in the Karabakh region

Predicted crop type	Area (hectares)
Corn silage	~110,000
Permanent meadows	~107,000
Total cropland	~217,000

By visually contrasting the classification map with high-resolution satellite imagery from Google Earth for the same period, qualitative validation was conducted. The model's performance was confirmed by the comparison, which revealed a strong correlation between the classified areas and the discernible agricultural patterns on the ground.

6. Discussion of classification performance and transferability

The obtained results are explained by the 14-band spectral stack and Random Forest combining complementary cues canopy vigor (NDVI), chlorophyll (NDRE), and moisture (NDWI/NDMI) visible in the French and Karabakh overviews (Fig. 3, *b–d*; Fig. 4, *b–d*) and formalized by the thresholded activity patterns in Table 1; RF's feature bagging and non-linear splits stabilize decision boundaries under domain shift, a property consistent with robust baselines reported in the paper [2] and in the paper [5]. Red-edge indices (NDRE) mitigate NDVI saturation and separate structurally similar crops; their elevation in high-biomass French parcels and scarcity in uplands/fallows in Karabakh (Table 1; Fig. 3, *c* vs. Fig. 4, *c*) directly addresses spectral similarity and cross-region divergence, aligning with single-date findings in the paper [3] and complementing time-series advantages noted in the paper [6]. Transferability was assessed by proxy validation: spatial coherence within parcels, agreement with the WorldCover cropland mask (Fig. 3, *e*; Fig. 4, *e*), and distributional plausibility of the Karabakh classification map in the Fig. 4, *f* and classified areas in the Table 7, while source-region accuracy, kappa, and calibration supported model reliability prior to transfer. In line with the paper [8], models trained in one agro-ecological region transfer best where eco-climatic differences and taxonomy mismatches are modest; without local labels, exact accuracy loss cannot be stated, but confidence-stratified outputs and irrigated-valley concentrations indicate where reliability is highest.

The main limitation of a single-snapshot design is the absence of phenological trajectories, increasing confusion among classes separated primarily by timing; small fields, mixed pixels, and acquisition-date misalignment further affect robustness. Proxy validation compensates for missing in-situ data by triangulating consistency and landscape logic, enabling operational plausibility checks until scarce labels become available. Practically, cross-regional mapping supports food-security monitoring, governance and compliance audits, and agribusiness planning by producing timely area esti mates (Table 7) without immediate local labeling. Eliminating up-front labeling improves scalability and cost-effectiveness, as open Sentinel-2 and WorldCover plus existing labels in France (Table 2) suffice to launch mapping.

Distinct disadvantages include the lack of absolute target accuracy and the omission of SAR or multi-temporal cues, these can be reduced by sparse local audits, active learning, label-efficient domain adaptation, and SAR/time-series augmentation.

This result addresses the core problem by eliminating the need for immediate local labeling while retaining operational specificity. Nonetheless, limitations persist due to phenology-driven confusions and heterogeneous mosaics. These can be mitigated in future work through sparse audits, active learning, and modest multi-sensor/temporal augmentation. Future development may also pursue domain-invariant representations, hierarchical class taxonomies, and calibrated uncertainty under shift, while addressing mathematical challenges in probability calibration, methodological issues in cross-region comparability, and experimental hurdles in label harmonization across borders, as discussed in [6] and [8].

7. Conclusion

1. The study proposes and formalizes a reproducible procedure for developing a transferable zero-shot crop-type classification method, consisting of 14-band feature engineering, candidate-model screening across four families with selection by accuracy and Cohen's κ, a training and prediction workflow producing probability/confidence outputs and standardized cartography with 95% Wilson intervals and two-proportion z-tests, and zero-shot transfer with a proxy validation protocol. This procedure directly addresses the problem stated by enabling label-free deployment in data-scarce regions while retaining interpretability and statistical accountability.

2. A peak-season, cloud-reduced, radiometrically harmonized Sentinel-2 composite was produced for the source (France) and target (Karabakh) regions. The mapped domain for transfer was delimited by the WorldCover cropland mask, yielding ≈ 217,000 ha of cropland and 8,848 confirmed grid cells in Karabakh, and 46,687 labeled French parcels for training. This alignment of acquisition window and radiometry provides the distinctive condition under which single-snapshot signals remain comparable across regions, addressing the domain-shift component of the problem and improving reproducibility relative to time-series-dependent approaches.

3. A fourteen-band stack (core bands + NDVI, NDRE, NDWI, NDMI) was constructed to encode canopy vigor, chlorophyll, and moisture. Binary activity thresholds (0.5) were applied to NDVI/NDRE/NDWI to yield verifiable patterns that are visible in the index maps. Distinctive advantages include mitigation of NDVI saturation via NDRE and discrimination of irrigated versus dry parcels via NDWI, which together explain separability in high-biomass French fields and selective reliability in irrigated Karabakh valleys. These features directly support the zero-label transfer objective by concentrating discriminative power in signals less sensitive to local label availability.

4. Among candidate classifiers, Random Forest provided the most dependable source-region performance, enabling zero-shot application to Karabakh and generation of the final classification map and classified crop areas. Transferability in the absence of local labels was evidenced through proxy validation parcel-level spatial coherence, agreement with the cropland mask, and distributional plausibility supported by confidence-stratified outputs.

Conflict of interest

The authors declare that they have no conflicts of interest in relation to the current study, including financial, personal, authorship, or any other, that could affect the study, as well as the results reported in this paper.

Financing

The study was conducted without financial support.

Data availability

The data will be provided upon reasonable request.

Use of artificial intelligence

The authors confirm that they did not use artificial intelligence technologies when creating the current work.

References

1. Hoppe, H., Dietrich, P., Marzahn, P., Weiß, T., Nitzsche, C., Freiherr von Lukas, U. et al. (2024). Transferability of Machine Learning Models for Crop Classification in Remote Sensing Imagery Using a New Test Methodology: A Study on Phenological, Temporal, and Spatial Influences. Remote Sensing, 16 (9), 1493. https://doi.org/10.3390/rs16091493
2. Rodriguez-Galiano, V. F., Ghimire, B., Rogan, J., Chica-Olmo, M., Rigol-Sanchez, J. P. (2012). An assessment of the effectiveness of a random forest classifier for land-cover classification. ISPRS Journal of Photogrammetry and Remote Sensing, 67, 93–104. https://doi.org/10.1016/j.isprsjprs.2011.11.002
3. Saini, R., Ghosh, S. K. (2018). Crop classification on single date Sentinel-2 imagery using Random Forest and Support Vector Machine. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XLII–5, 683–688. https://doi.org/10.5194/isprs-archives-xlii-5-683-2018
4. Sonobe, R., Yamaya, Y., Tani, H., Wang, X., Kobayashi, N., Mochizuki, K. (2018). Crop classification from Sentinel-2-derived vegetation indices using ensemble learning. Journal of Applied Remote Sensing, 12 (2). https://doi.org/10.1117/1.jrs.12.026019
5. 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. https://doi.org/10.3390/s18010018
6. Vuolo, F., Neuwirth, M., Immitzer, M., Atzberger, C., Ng, W.-T. (2018). How much does multi-temporal Sentinel-2 data improve crop type classification? International Journal of Applied Earth Observation and Geoinformation, 72, 122–130. https://doi.org/10.1016/j.jag.2018.06.007
7. Pelletier, C., Valero, S., Inglada, J., Champion, N., Dedieu, G. (2016). Assessing the robustness of Random Forests to map land cover with high resolution satellite image time series over large areas. Remote Sensing of Environment, 187, 156–168. https://doi.org/10.1016/j.rse.2016.10.010
8. Orynbaikyzy, A., Gessner, U., Conrad, C. (2022). Spatial Transferability of Random Forest Models for Crop Type Classification Using Sentinel-1 and Sentinel-2. Remote Sensing, 14 (6), 1493. https://doi.org/10.3390/rs14061493

9. Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., Moore, R. (2017). Google Earth Engine: Planetary-scale geospatial analysis for everyone. Remote Sensing of Environment, 202, 18–27. https://doi.org/10.1016/j.rse.2017.06.031
10. Zanaga, D., Van De Kerchove, R., Kirches, G., Daems, D., De Keersmaecker, W., Brockmann, C., Arino, O. (2022). ESA WorldCover 10 m 2021 v200: global land cover map. Zenodo. https://doi.org/10.5281/zenodo.7254221
11. Attard, G., Bardonnet, J. (2020). RPG Version 2.0. Registre Parcellaire Graphique. Available at: https://geodatafr.github.io/IGN/RPG_Agricultural-parcels/
12. Pettorelli, N., Vik, J. O., Mysterud, A., Gaillard, J.-M., Tucker, C. J., Stenseth, N. Chr. (2005). Using the satellite-derived NDVI to assess ecological responses to environmental change. Trends in Ecology &amp; Evolution, 20 (9), 503–510. https://doi.org/10.1016/j.tree.2005.05.011
13. Gitelson, A. A., Merzlyak, M. N. (1997). Remote estimation of chlorophyll content in higher plant leaves. International Journal of Remote Sensing, 18 (12), 2691–2697. https://doi.org/10.1080/014311697217558
14. McFeeters, S. K. (1996). The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features. International Journal of Remote Sensing, 17 (7), 1425–1432. https://doi.org/10.1080/01431169608948714
15. Gao, B. (1996). NDWI–A normalized difference water index for remote sensing of vegetation liquid water from space. Remote Sensing of Environment, 58 (3), 257–266. https://doi.org/10.1016/s0034-4257(96)00067-3