Application of machine learning methods and Sentinel-2 data for multitemporal land-cover classification in conflict-affected areas

Kinga Karwowska, Aleksandra Sekrecka, Jakub Ślesiński, Magdalena Lewińska

Department of Imagery Intelligence, Faculty of Civil Engineering and Geodesy, Military University of Technology, 00-908 Warsaw, Poland – (kinga.karwowska, aleksandra.sekrecka, jakub.slesinski, magdalena.lewinska)@wat.edu.pl

Keywords: machine learning, classification, war damage, Sentinel, AlphaEarth

Abstract

In many regions of the world, especially those affected by armed conflicts, urbanization, or intensive environmental transformations, a high dynamic of land use and land cover changes is observed. Reliable monitoring of these processes requires the application of classification methods that ensure both high thematic accuracy and temporal consistency. This paper presents a multitemporal classification methodology based on Sentinel-2 optical data and machine learning models.

The research was conducted for the city of Sievierodonetsk (Luhansk Oblast, Ukraine) – an area that suffered significant destruction in 2022 as a result of military operations. The aim of the analysis was to identify land cover changes in the years 2021-2025 using three classifiers: k-Nearest Neighbors (kNN), Random Forest (RF), and Gradient Boosting Classifier (GBC), combined into an ensemble system based on dynamic confidence weighting.

Quality assessment using the recall metric showed that the fusion method outperformed individual classifiers, achieving average values of 0.87-0.96, while classical models obtained 0.81-0.89. The largest changes (39%) occurred in the years 2022-2023, coinciding with the period of greatest military activity. The proposed method achieved the highest classification quality indices (F1 = 0.93, Acc = 0.98 for 2021), surpassing global products and models based on AlphaEarth. In subsequent years, high stability was maintained (F1 ≥ 0.88), confirming the effectiveness and robustness of the approach under various environmental conditions.

1. Introduction

Advancements in satellite observations over the last decade, particularly within the Sentinel-1 and Sentinel-2 missions, have revolutionized the way land use and land cover changes are analyzed, enabling systematic observations at high spatial and temporal resolution. These data, acquired cyclically and made freely available by the European Space Agency (ESA), have become the foundation for modern classification methods and environmental change analysis at local and regional scales. At the same time, maintaining thematic and temporal consistency in multi-year classifications remains one of the most challenging issues in remote sensing, especially in areas undergoing rapid anthropogenic transformations or destructive war processes, such as eastern Ukraine. During armed conflicts, satellite classification becomes particularly significant, as it allows for objective, quantitative assessment of infrastructure damage, degradation of biologically active surfaces, and changes in land cover structure.

Contemporary land cover classification systems are based on machine learning methods that utilize both spectral and spatial features. Classical algorithms remain the basis of many operational satellite analysis systems due to their stability, low computational complexity, and high interpretability. Unlike deep learning models, they do not require massive training datasets or intensive computational power, making them especially useful for large-scale, multitemporal tasks. Despite the growing popularity of deep learning methods, classical approaches still offer advantages in terms of transparency and the ability for expert verification of results. In operational practice, this translates to greater trust in the outcomes, which is particularly important in the context of analyzing critical data, such as those related to armed conflicts or natural disasters.

2. Related works

Among the classical machine learning methods used in classification tasks (Maxwell et al., 2018) are k-Nearest Neighbors (kNN) (Fix, 1985; Pan et al., 2020), Support Vector Machines (SVM) (Pal and Mather, 2005), Random Forest (RF) (Belgiu and Drăguţ, 2016; Ho, 1995) and Gradient Boosting (Friedman, 2001) (e.g. LightGBM). These models are characterized by relatively low computational complexity, high interpretability, and resistance to overfitting when parameters are properly selected.

Another approach is object-based image analysis (OBIA), in which pixels are grouped into segments (superpixels) and then classified based on object statistics such as mean band values, texture indices, or shape features (Blaschke, 2010). This approach helps to reduce the so-called "checkerboard effect" at class boundaries, which is often observed in pixel-based classification.

More advanced methods, such as Markov Random Fields (MRF) and Conditional Random Fields (CRF), take into account spatial dependencies between pixels, enabling smoothing of classification results and improving spatial consistency at the cost of a risk of excessive generalization of class boundaries (Zhong et al., 2006).

In recent years, significant progress has been made through the use of deep neural networks (Deep Learning), particularly convolutional models (CNNs), employed for semantic image segmentation (Chen et al., 2018; Ronneberger et al., 2015). Architectures such as U-Net, SegNet, and DeepLab have become the standard in land cover classification tasks, offering high accuracy thanks to their ability to learn hierarchical feature representations (Zhang et al., 2018).

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-XI-3-2026-555-2026

Figure 1. Study area covering the city of Sievierodonetsk and its surroundings in the Luhansk Oblast, eastern Ukraine (including parts of the Sievierodonetsk Raion).

Nevertheless, their effective application requires large, well-annotated training datasets, the preparation of which is a time-consuming process, and the quality of labels has a crucial impact on the final model outcomes.

Simultaneously, methods for data fusion and model fusion are being developed to improve classification quality by integrating information from multiple sources or algorithms. Multisensor fusion (e.g., combining optical Sentinel-2 and radar Sentinel-1 data) allows for increased robustness of classification against cloud cover and enhances the distinguishability of classes with similar spectral signatures (Chen et al., 2017). Multimodel fusion, on the other hand, can be implemented at various levels: feature-level, by combining feature vectors; score-level, by averaging or weighting class probabilities; and decision-level, where the final result is the outcome of voting among different classifiers (Bischke et al., 2019; Du et al., 2012).

In land cover classification tasks, methods based on confidence-weighted voting (Foody, 2002; Opitz and Maclin, 1999) have gained particular popularity, representing a variant of classical majority voting. Other commonly used strategies include max-confidence decision (Foody and Mathur, 2004), where the class with the highest confidence level among all models is selected, and majority voting with confidence tie-break (Foody and Mathur, 2004), which incorporates a mechanism for resolving ties using confidence values. These mechanisms increase resistance to cases of erroneous but highly confident predictions by individual models, resulting in greater stability of fusion outcomes.

Considering the development of classification methods and the need to increase result reliability, the conducted research focused on developing an ensemble method that combines three classical, well-interpretable classification algorithms (kNN, RF, and GBC) into a single coherent decision system. The proposed method utilizes classification result fusion with normalized confidence levels (confidence normalization), enabling dynamic weighting of models depending on the local reliability of their decisions. In this work, we address the following research questions:

- Do classical classification algorithms, such as kNN, differ significantly in effectiveness compared to more complex machine learning methods?
- Does the use of classical ML models allow for more detailed and stable classification results compared to global deep learning-based products, such as AlphaEarth?
- Does the use of classification result fusion based on normalized model confidence levels significantly increase land cover classification accuracy in Sentinel-2 data?

The structure of the article is as follows: Section 3 describes the methodology used, Section 4 presents the research results, Section 5 discusses their interpretation, and Section 6 contains the final conclusions.

3. Methodology

The aim of this study was to analyze land cover changes between 2022 and 2025 based on supervised classification of Sentinel-2A optical imagery. The research focuses on an area where six land cover categories were distinguished: bare ground, built-up area, agricultural and green cover, grass and scrubs, trees, and water.

In this work, we applied an ensemble classifier utilizing the potential of selected machine learning-based classification methods.

3.1 Study area

To verify the effectiveness of the proposed methodology, the city of Sievierodonetsk (Ukr. Sievierodonetsk), located in the northeastern part of Ukraine, in the Luhansk region, was selected (Figure 1). This area was affected by intense military operations in 2022, resulting in significant destruction of urban infrastructure and substantial changes in land cover structure in adjacent areas. Due to the high dynamics of land cover changes and the presence of contrasting classes (built-up areas, bare land, vegetation, water), this area represents a challenging case for assessing the effectiveness of the proposed classification fusion method.

3.2 Data

This study used multispectral data from the Sentinel-2 satellite (Level-2A product). Since these data are freely provided by the European Space Agency (ESA) via the Copernicus platform (European Union, Copernicus Land Monitoring Service, 2025), the developed method can be easily applied to analyze other areas with dynamically changing environmental conditions.

All bands with a spatial resolution of 10 m (B02, B03, B04, B08) and 20 m (B05, B06, B07, B08A, B11, B12) were used for the research. Additionally, to enrich the representation of input data, commonly used remote sensing indices NDVI (Normalized Difference Vegetation Index) and NDWI (Normalized Difference Water Index) were calculated. Including these indices was part of feature engineering, providing the model with additional spectral information that significantly improves land cover classification accuracy, especially for the water and vegetation classes.

Due to the relatively large number of features describing a single pixel (10 spectral bands and 2 indices), an analysis of class separability and sample compactness within classes was conducted. The results of the Silhouette score analysis showed that the best class separation was achieved using the PCA (0.495), a slightly lower value was obtained for UMAP method (0.352), while the lowest score was recorded for t-SNE (0.172) (Figure 2).

Based on the value of the Silhouette coefficient, a decrease in data separability can be observed with the application of certain dimensionality reduction methods, which may negatively affect classification quality, especially in the case of objects with weak spectral distinctiveness. Therefore, for the subsequent classification stage, data without dimensionality reduction were used.

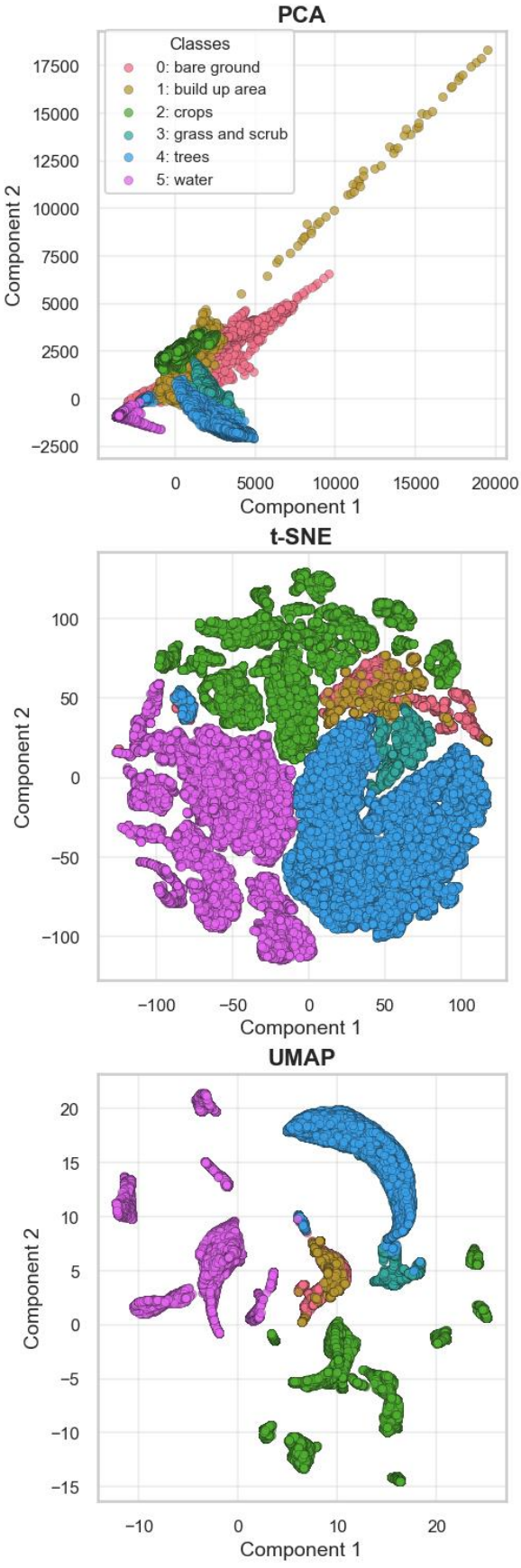

Figure 2. Visualization of class separability using dimensionality reduction methods: PCA, t-SNE, and UMAP. Colors correspond to land-cover classes

3.3 Proposed method

In this study, three machine learning algorithms were used to prepare land cover maps:

- kNN (k-Nearest Neighbors) (Fix, 1985; Pan et al., 2020) – a very simple and easily interpretable model. The model makes predictions based on the analysis of the nearest neighbors,
- RF (Random Forest) (Belgiu and Drăguţ, 2016; Ho, 1995) – an ensemble learning model. During the training process, a set of independent decision trees is built, each trained on independent data samples. The final class decision is made by majority voting,
- Gradient Boosting Classifier (GBC) (Friedman, 2001) – a representative of more advanced ensemble learning. Unlike RF, it operates sequentially, with each subsequent tree learning from the errors of its predecessors.

Due to the differing properties of the three models and the observed differences in classification quality for individual classes, a classification fusion procedure was developed as part of the study. The aim of this stage was to obtain a final map that is more stable and resistant to local classification errors. Each model, in addition to the class label, also returned a confidence value, interpreted as the degree of certainty of the prediction. However, the method for determining confidence varies depending on the model's characteristics. For kNN, it is defined as the inverse of the average distance to the k nearest neighbors, while for RF and GBC, it is the maximum class probability (i.e., the highest value among class probabilities).

Due to the different ranges and distributions of confidence values, min-max normalization in the range [0,1] was applied separately for each model (1):

$$\hat{c}^{(m)} = \frac{c^{(m)} - min(c^{(m)})}{max(c^{(m)}) - min(c^{(m)}) + \varepsilon} \quad (1)$$

Normalized values $\hat{c}^{(m)}$ enabled direct comparison of model confidence on the same scale. In alternative tests, median standardization using the median and median absolute deviation (MAD) was applied; however, variance analysis showed that simple min-max normalization provided greater stability in areas of low spectral contrast.

In the next stage, for each pixel, the combined score for each class k was calculated according to the formula (2):

$$S_k = w_{knn} \cdot P_{knn}(k) + w_{rf} \cdot P_{rf}(k) + w_{gb} \cdot P_{gb}(k) \quad (2)$$

where $P_m(k)$ denotes the normalized confidence of a given model (m) for class (k), w_m is the weight assigned to model m. The weights (w_m) were determined dynamically, proportionally to the local confidence values (3):

$$w_m = \frac{\hat{c}^{(m)}}{\sum_i \hat{c}^{(i)} + \varepsilon} \quad (3)$$

The use of dynamic weights enabled increasing the influence on the decision of the model that, at a given point, was characterized by higher prediction confidence.

The final classification decision for each pixel was determined as the class with the highest combined score (4):

$$y *= arg \max_k S_k \quad (4)$$

Meanwhile, the confidence (C*) was defined as the difference between the two highest values S_k (5):

$$C *= S_{(1)} - S_{(2)} \quad (5)$$

This approach enables quantitative assessment of the decisiveness of the classification decision. Pixels with low C* values (e.g., < 0.3) are classified as uncertain (potential areas of misclassification). Additionally, it allows identification of predictions that, despite high confidence, are incorrect. The last stage of the proposed method includes spatial filtering. To increase the spatial coherence of the map, a majority filter in a 3×3 pixel window was applied, which eliminates individual misclassified pixels while preserving boundary detail. In comparative tests, median filtering and morphological opening-closing filters were also analyzed, but their impact on accuracy improvement was minor relative to computational costs.

3.4 Evaluation metrics

To assess the quality of the classification performed, qualitative indicators commonly used in remote sensing were applied. The accuracy metric defines the proportion of correctly classified pixels among all analyzed cases, serving as a basic measure of model effectiveness. Precision indicates what percentage of pixels assigned to a given class were classified correctly, while recall shows to what extent the model can detect all pixels truly belonging to a given class. The F1-score, being the harmonic mean of precision and recall, allows for a balanced evaluation of classification quality in situations where both recognition accuracy and detection completeness are important. These indicators enable evaluation of classification effectiveness, taking into account both the overall correctness of the model and the recognition quality of individual thematic classes.

4. Experiments and Results

For the development of land cover classification models, three previously described algorithms were used: k-Nearest Neighbors (kNN), Random Forest (RF), and Gradient Boosting Classifier (GBC). To select the optimal hyperparameters, the Grid Search with Cross-Validation strategy was applied, enabling the exploration of the parameter space and the selection of the configuration that provided the highest accuracy metric. This approach allowed for the creation of model variants best suited to the data characteristics and the specifics of the studied area.

Table 1 presents the final sets of parameters used for classification. Although the hyperparameter tuning process was carried out independently for each analyzed year, the ultimately selected configurations turned out to be identical in all cases, indicating the stability of the adopted methodology.

To ensure the reliability of comparisons, the training and validation process was conducted in a repeatable manner, maintaining the same set of training fields and identical input data normalization procedures for all analyzed years.

Model	Best Hyperparameters
kNN	n_neighbors: 3, weights: distance
RF	max_depth: 20, n_estimators: 100
GBC	learning_rate: 0.1, max_depth: 5, n_estimators: 200

Table 1. Optimal sets of hyperparameters obtained for the applied classification models: k-Nearest Neighbors (kNN), Random Forest (RF), and Gradient Boosting Classifier (GBC).

For each of the three classifiers and their fusion, land cover maps were predicted for the subsequent five years (2021-2025), maintaining the same input data structure and set of training samples. The results were compared based on the recall metric, calculated independently for each class, year, and method (Table 2, Figure 3). The choice of this metric was dictated by its sensitivity to false negative errors, which are important in land cover analysis, where completeness of detection for

underrepresented classes is particularly crucial. High recall values indicate the model’s ability to fully reproduce the actual extent of a given class, while low values suggest problems with spectral separability or scene complexity in a given year.

Year	Method	0	1	2	3	4	5	MEAN
2021	kNN	0.89	0.78	0.80	0.95	0.97	1.00	0.89
	RF	0.84	0.80	0.59	0.96	0.98	1.00	0.86
	GBC	0.71	0.45	0.76	0.97	0.96	1.00	0.81
	our	0.99	0.96	0.83	1.00	1.00	1.00	0.96
2022	kNN	0.67	0.93	0.77	0.86	0.91	1.00	0.86
	RF	0.70	0.97	0.80	0.89	0.86	1.00	0.87
	GBC	0.74	0.95	0.82	0.90	0.85	1.00	0.88
	our	0.75	0.99	0.83	0.90	0.87	1.00	0.89
2023	kNN	0.58	1.00	0.67	0.73	0.92	0.96	0.81
	RF	0.58	1.00	0.69	0.74	0.93	0.96	0.82
	GBC	0.68	1.00	0.65	0.70	0.91	0.96	0.82
	our	0.79	1.00	0.74	0.77	0.94	0.96	0.87
2024	kNN	0.87	1.00	0.63	0.90	0.91	1.00	0.89
	RF	0.87	1.00	0.57	0.96	0.90	1.00	0.88
	GBC	0.78	1.00	0.64	0.99	0.89	1.00	0.88
	our	0.92	1.00	0.76	1.00	0.91	1.00	0.95
2025	kNN	0.78	1.00	0.94	0.70	0.84	1.00	0.88
	RF	0.78	0.99	0.92	0.62	0.84	1.00	0.86
	GBC	0.78	1.00	0.92	0.73	0.82	1.00	0.88
	our	0.83	1.00	0.95	0.74	0.88	1.00	0.90

Table 2. Classification accuracy results (recall) for individual land-cover classes for the years 2021-2025 obtained using the evaluated models and the proposed solution (*our*).

An analysis of the results indicates that the proposed fusion method (our) consistently achieved the highest recall values for all land cover classes in the years 2021-2025.
The average values of this indicator for the fusion method ranged from 0.87 to 0.96, while for classical algorithms, they ranged from 0.81 to 0.89.
An analysis of the results for individual classes shows that built-up areas (ID1) and water (ID5) achieved the highest and most stable recall values (0.95-1.00), regardless of the method used, confirming their good spectral separability. Tree-covered areas (ID4) also achieved very high values (0.85-1.00), although the difference between algorithms was greater. In contrast, low and cultivated vegetation (ID2 and 3) showed the greatest temporal variability, especially in the years 2022-2023, which coincides with the most intense military activity in this area. Bare surfaces (ID0) were also characterized by a high level of recall, with the fusion method significantly improving the results (from 0.71-0.89 to 0.99).
In Figure 4 (Sankey diagram), the interannual land cover transitions are visible. The diagram shows that the analyzed area was characterized by spatial dynamics, particularly in the years 2022-2023, which coincides with a period of intense military activity in the region. The share of pixels that retained the same class between consecutive years ranged from 61.8% (2021-2022) to 66.9% (2024-2025), with the highest share of changes (39.1%) occurring between 2022-2023. This means that during this period, nearly 40% of the study area changed land cover class, confirming the significant impact of infrastructure destruction, fires, and changes in land use caused by the armed conflict.

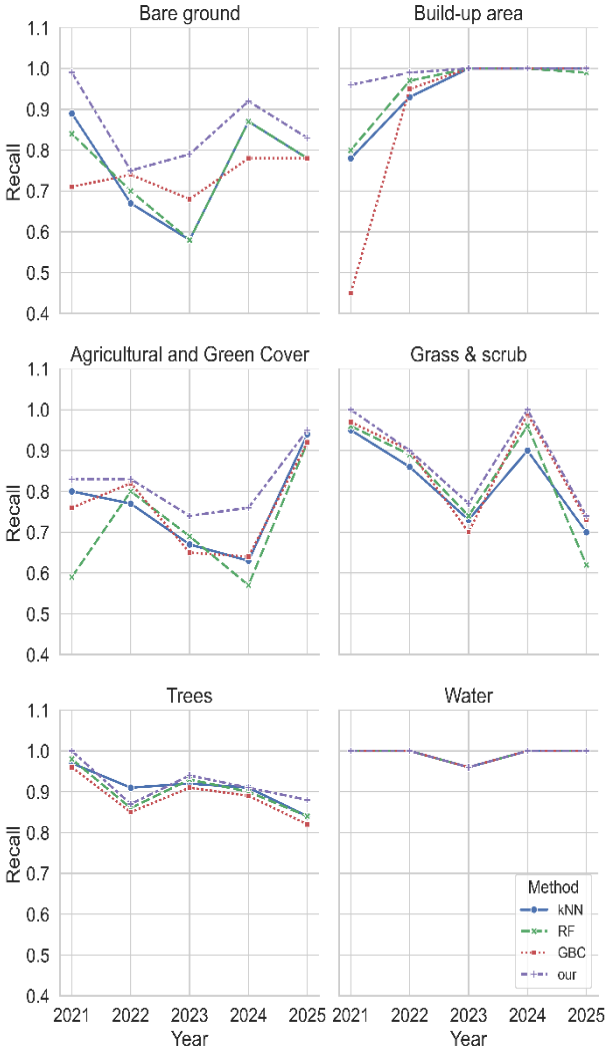

Figure 3. Changes in *recall* values for individual land-cover classes from 2021 to 2025 obtained using the kNN, Random Forest (RF), Gradient Boosting Classifier (GBC), and the proposed (our) model.

The most unstable category was agricultural land (from 17.9% to 32.2%). However, such high variability is typical for areas used seasonally. These changes are confirmed by the State Statistics Service of Ukraine, 2021-2023 (Verner, 2025). According to this data, in 2022 there was a sharp decline, where production dropped to 28.5% of the 2021 value, reflecting the effects of the full-scale invasion, loss of control over a significant part of agricultural land, and infrastructure destruction. In 2023, a partial rebound was recorded (81.5%), which still remains well below pre-war levels. These data confirm the deep crisis in agriculture in the frontline regions and its slow, uneven recovery, as illustrated by our Sankey diagram.
Built-up areas and tree-covered surfaces showed moderate stability, but there were also periodic changes ranging from 5-13%, which can be linked to the destruction of buildings and damage to forests. According to the United Nations Satellite Centre (UNOSAT) (“UNOSAT,” 2025), as of June 25, 2022 (the date of acquisition of the imagery analyzed by UNOSAT), 1 038 buildings with an area greater than 50 m² had been destroyed in the analyzed area.

This contribution has been peer-reviewed. The double-blind peer-review was conducted on the basis of the full paper.
https://doi.org/10.5194/isprs-annals-XI-3-2026-555-2026 | © Author(s) 2026. CC BY 4.0 License.

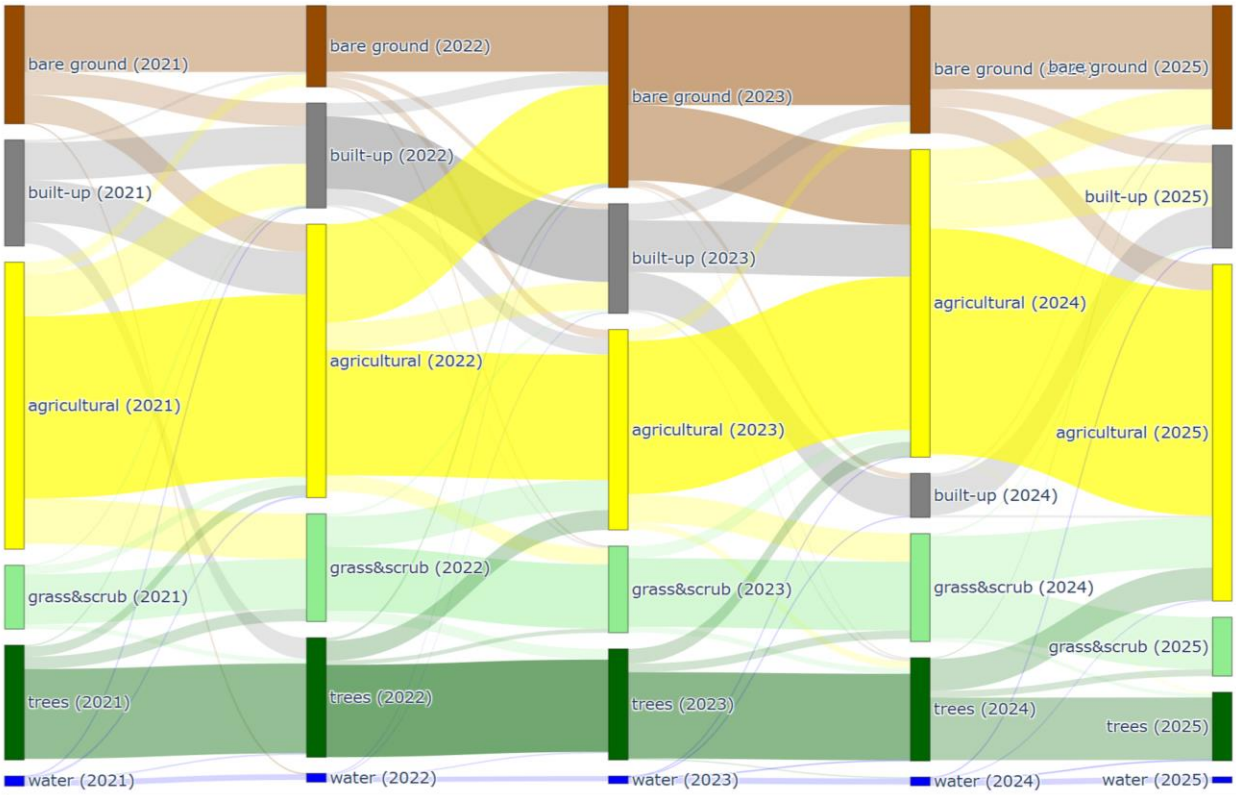

Figure 4. Sankey diagram illustrating interannual land-cover changes for the analyzed Sievierodonetsk area during 2021–2025. The width of each flow represents the number of pixels transitioning between land-cover classes, highlighting both the direction and magnitude of spatial transformations over the analyzed period.

year, metric / method	2021				2022				2023				2024				2025			
	F1	Prec.	Rec.	Acc.	F1	Prec.	Rec.	Acc.	F1	Prec.	Rec.	Acc.	F1	Prec.	Rec.	Acc.	F1	Prec.	Rec.	Acc.
ESRI													0.02	0.01	0.04	0.67				
DynamicWorld													0.59	0.66	0.64	0.88				
WorldCover									0.58	0.66	0.61	0.89								
AlphaEarth - RF	0.71	0.74	0.78	0.92	0.76	0.76	0.81	0.93	0.74	0.74	0.79	0.94	0.75	0.76	0.82	0.93				
AlphaEarth - GB	0.76	0.75	0.83	0.92	0.79	0.77	0.83	0.93	0.84	0.82	0.89	0.95	0.80	0.82	0.86	0.93				
AlphaEarth - kNN	0.83	0.83	0.88	0.94	0.82	0.81	0.89	0.95	0.83	0.81	0.89	0.95	0.83	0.83	0.88	0.94				
RF	0.84	0.84	0.89	0.96	0.83	0.81	0.87	0.95	0.82	0.81	0.84	0.94	0.84	0.82	0.88	0.95	0.83	0.82	0.86	0.95
GBC	0.78	0.78	0.81	0.95	0.84	0.83	0.88	0.95	0.80	0.82	0.81	0.94	0.84	0.82	0.88	0.95	0.84	0.83	0.87	0.95
kNN	0.88	0.87	0.90	0.97	0.82	0.81	0.85	0.95	0.79	0.79	0.81	0.94	0.87	0.86	0.89	0.96	0.87	0.87	0.89	0.96
our	0.93	0.93	0.94	0.98	0.85	0.84	0.89	0.96	0.84	0.83	0.90	0.95	0.88	0.87	0.91	0.96	0.88	0.87	0.89	0.96

Table 3. Land-cover classification accuracy assessment results for the Sievierodonetsk area from 2021 to 2025. The table presents the values of F1-score, Precision, Recall, and Accuracy metrics obtained for selected data sources (ESRI, Dynamic World, WorldCover), as well as classifications based on the AlphaEarth dataset using RF, GB, and kNN algorithms, and land-cover maps

5. Discussion

As presented in the Results subsection, our proposed method demonstrates significant qualitative superiority compared to individual methods. The results also showed how differently various algorithms infer, despite using the same input data.

- Comparison with other land cover products

To evaluate the effectiveness of the proposed method, a comparison of classification results was conducted against available global land cover products, such as ESRI Land Cover (2024) (Esri Land Cover, 2025), Dynamic World (2024) (Brown et al., 2022) and ESA WorldCover (v200, 2023) (Zanaga et al., 2022). Additionally, a set of classifiers trained on AlphaEarth (2024**)** (Google DeepMind, 2025) data was prepared, utilizing the same strategy and identical training fields as the proposed method.

Each of the mentioned products is characterized by a different classification strategy and temporal data coverage. The ESRI Land Cover product is based on a CNN model trained on Sentinel-2 data, covering global data for the year 2024. Dynamic World is a dynamic product based on neural networks, providing daily Sentinel-2 classifications from 2020 onwards, enabling the analysis of seasonal changes and long-term trends. ESA WorldCover (v200, 2023), on the other hand, is an annual classification product with 10 m resolution, developed by the European Space Agency (ESA) using Sentinel-1 and Sentinel-2

data, representing the average annual land cover status for the year 2023.

A new and particularly interesting data source is AlphaEarth (2024) – a global product generated using deep learning models and satellite embeddings. However, its limitation is availability only for selected time intervals – maps are generated by aggregating data from approximately one year, which prevents the analysis of short-term seasonal changes.

A detailed summary of the classification quality of individual products is presented in Table 3.

The results indicate a clear advantage of the proposed method over both global products and classical baseline models. For the year 2021, the highest quality metric values were achieved (F1 = 0.93, Precision = 0.93, Recall = 0.94, Accuracy = 0.98), confirming high classification consistency with reference data. In subsequent years, despite natural fluctuations in class distribution and image quality, the method maintained stability – F1 and Accuracy values remained at ≥ 0.88 and ≥ 0.96, respectively.

For comparison, global products (ESRI, Dynamic World, WorldCover) achieved significantly lower results – F1 in the range of 0.58-0.59 and Accuracy ~0.88-0.89 – confirming the limited effectiveness of global-scale models for local analyses. Classifiers trained on AlphaEarth data (RF, GB, kNN) achieved moderately high values (F1 = 0.79-0.84), but still lower than the proposed method, demonstrating the effectiveness of the designed pipeline and the use of contextual data.

- Quality of Building Detection

In the conducted studies, particular attention was paid to the quality of built-up area detection. A comparative analysis was performed for the two highest-rated land cover maps (Table 3) for the year 2023. The choice of this year was deliberate – for this period, vector reference data on buildings from the Microsoft Building Footprints (MS Buildings) database were available. OpenStreetMap (OSM) data were not used due to numerous gaps in the building layer, likely related to infrastructure damage resulting from warfare. Figure 5 presents examples of built-up area detection results (marked in gray) for AlphaEarth (b) and the method proposed in this study (c).

The analysis shows that the developed method correctly identified 87.4% of buildings with an area exceeding 50 m², while the AlphaEarth product achieved higher effectiveness at 95%. However, it should be emphasized that AlphaEarth results – similar to other global products – tend to over-interpret, i.e., classify some non-built-up areas as built-up (as seen in Figure 5b).

The proposed method features greater precision in locating built-up objects. This is possible due to the use of a small spatial filter (3×3), which limits excessive generalization and allows better preservation of object boundary details.

- Limitations

The comparison with global land cover products involves certain limitations related to the characteristics of these datasets. Products such as ESRI Land Cover, Dynamic World, and ESA WorldCover are designed for global-scale applications and therefore rely on generalized classification models trained on highly diverse datasets. As a result, these models may be less adapted to local environmental conditions and the specific landscape characteristics of the study area.

One of the key challenges is spectral confusion between land cover classes. In the analyzed area, there is a noticeable similarity in spectral signatures between building rubble, exposed soil, and degraded urban surfaces, which may lead to misclassification in models trained on global datasets.

Another limitation concerns the temporal aggregation of imagery. Products such as WorldCover and ESRI Land Cover are based on annual Sentinel composites, which average the spectral signal over long periods and reduce the ability to detect short-term changes. A similar limitation applies to AlphaEarth, where embedding representations are generated from data aggregated over approximately one year.

In contrast to global products, the proposed method was tailored to the characteristics of the study area by using local training data and contextual analysis, which helped reduce classification errors and improve the precision of built-up area detection.

6. Conclusion

The conducted research confirmed the effectiveness of classical machine learning methods in analyzing land cover changes on Sentinel-2 imagery, especially in regions affected by dynamic transformation processes, such as areas impacted by warfare. The results indicated that kNN, RF, and GBC models allow for high quality metric values at relatively low computational costs and high interpretability of results.

The proposed classifier fusion method, based on dynamic confidence weighting, enabled a significant improvement in

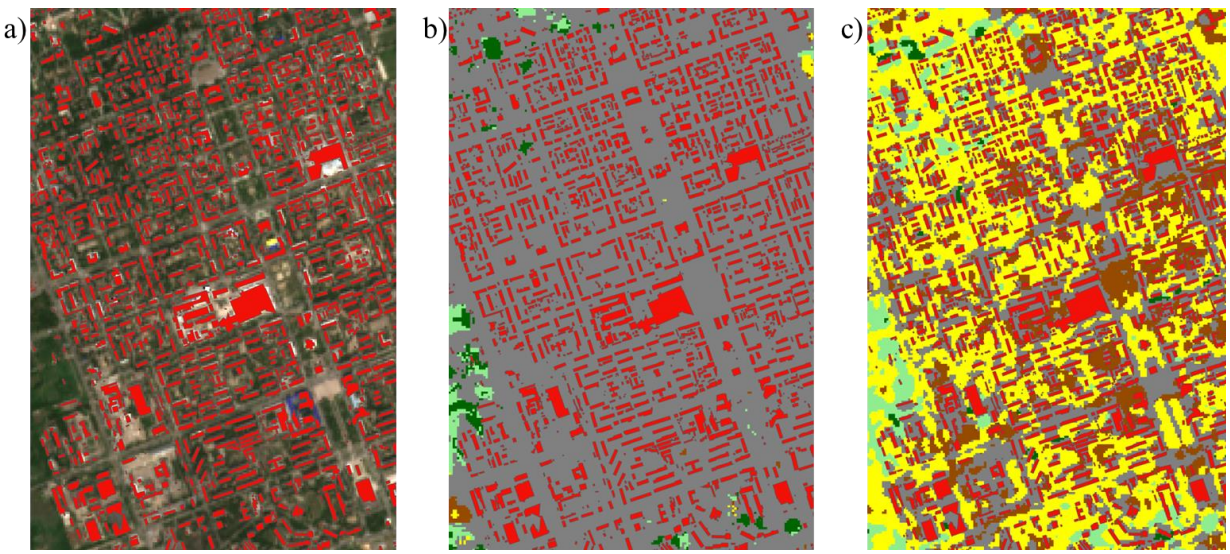

Figure 5. Comparison of built-up area detection (a) results using (b) AlphaEarth and the proposed method (c).

classification stability and accuracy compared to individual models. The introduction of an adaptive model weighting mechanism allowed for more effective utilization of model properties and reduction of misclassifications in areas of low spectral contrast.

Multitemporal analysis showed that the greatest changes in land cover structure occurred in 2022-2023, which corresponds to the period of intense military activity in the region. The share of pixels retaining the same class between consecutive years increased from 61.8% in 2021-2022 to 66.9% in 2024-2025, indicating gradual stabilization and reconstruction. Agricultural areas exhibited the greatest variability, while the “built-up” and “trees” classes maintained relatively high spatial stability. Research results confirm that classical ML models, when properly prepared and combined into an ensemble system, can serve as an effective alternative to complex deep learning methods, especially in cases of limited training data and in multi-temporal analyses. The developed fusion method can be easily integrated with existing remote sensing data processing systems and used operationally for environmental change monitoring tasks.

In future research, it is planned to extend the approach by integrating Sentinel-1 radar data in both classification and multi-temporal analysis. Additionally, it is planned to account for conditions of classification ambiguity, enabling the identification of spatial error patterns resulting from spectral similarities between classes. The integration of a confusion-aware learning module with confidence-level-based fusion would allow for adaptive enhancement of decision reliability in problematic areas.

From a broader perspective, the developed methodology may provide a foundation for building continuous environmental monitoring systems in crisis conditions, supporting the analysis of the impacts of armed conflicts, natural disasters, and infrastructure reconstruction processes.

References

Belgiu, M., Drăguţ, L., 2016. Random forest in remote sensing: A review of applications and future directions. ISPRS J. Photogramm. Remote Sens. 114, 24–31. https://doi.org/10.1016/j.isprsjprs.2016.01.011

Bischke, B., Helber, P., Folz, J., Borth, D., Dengel, A., 2019. Multi-Task Learning for Segmentation of Building Footprints with Deep Neural Networks, in: 2019 IEEE International Conference on Image Processing (ICIP). Presented at the 2019 IEEE International Conference on Image Processing (ICIP), pp. 1480–1484. https://doi.org/10.1109/ICIP.2019.8803050

Blaschke, T., 2010. Object based image analysis for remote sensing. ISPRS J. Photogramm. Remote Sens. 65, 2–16. https://doi.org/10.1016/j.isprsjprs.2009.06.004

Brown, C.F., Brumby, S.P., Guzder-Williams, B., Birch, T., Hyde, S.B., Mazzariello, J., Czerwinski, W., Pasquarella, V.J., Haertel, R., Ilyushchenko, S., Schwehr, K., Weisse, M., Stolle, F., Hanson, C., Guinan, O., Moore, R., Tait, A.M., 2022. Dynamic World, Near real-time global 10 m land use land cover mapping. Sci. Data 9, 251. https://doi.org/10.1038/s41597-022-01307-4

Brown, C.F., Kazmierski, M.R., Pasquarella, V.J., Rucklidge, W.J., Samsikova, M., Zhang, C., Shelhamer, E., Lahera, E., Wiles, O., Ilyushchenko, S., Gorelick, N., Zhang, L.L., Alj, S., Schechter, E., Askay, S., Guinan, O., Moore, R., Boukouvalas, A., Kohli, P., 2025. AlphaEarth Foundations: An embedding field model for accurate and efficient global mapping from sparse label data. https://doi.org/10.48550/arXiv.2507.22291

Chen, B., Huang, B., Xu, B., 2017. Multi-source remotely sensed data fusion for improving land cover classification. ISPRS J. Photogramm. Remote Sens. 124, 27–39. https://doi.org/10.1016/j.isprsjprs.2016.12.008

Chen, K., Fu, K., Yan, M., Gao, X., Sun, X., Wei, X., 2018. Semantic Segmentation of Aerial Images With Shuffling Convolutional Neural Networks. IEEE Geosci. Remote Sens. Lett. 15, 173–177. https://doi.org/10.1109/LGRS.2017.2778181

Du, P., Xia, J., Zhang, W., Tan, K., Liu, Y., Liu, S., 2012. Multiple Classifier System for Remote Sensing Image Classification: A Review. Sensors 12, 4764–4792. https://doi.org/10.3390/s120404764

Esri Land Cover, A.L.A. of the, 2025. Esri Land Cover [WWW Document]. URL https://livingatlas.arcgis.com/ (accessed 10.30.25).

European Union. (2025). Copernicus Land Monitoring Service. https://www.copernicus.eu/en (accessed 10.30.25).

Fix, E., 1985. Discriminatory Analysis: Nonparametric Discrimination, Consistency Properties. USAF School of Aviation Medicine.

Foody, G.M., 2002. Status of land cover classification accuracy assessment. Remote Sens. Environ. 80, 185–201. https://doi.org/10.1016/S0034-4257(01)00295-4

Foody, G.M., Mathur, A., 2004. A relative evaluation of multiclass image classification by support vector machines. IEEE Trans. Geosci. Remote Sens. 42, 1335–1343. https://doi.org/10.1109/TGRS.2004.827257

Friedman, J.H., 2001. Greedy Function Approximation: A Gradient Boosting Machine. Ann. Stat. 29, 1189–1232.

Google DeepMind. (2025). AlphaEarth Foundations helps map our planet in unprecedented detail. https://deepmind.google/discover/blog/alphaearth-foundations-helps-map-our-planet-in-unprecedented-detail/

Ho, T.K., 1995. Random decision forests, in: Proceedings of 3rd International Conference on Document Analysis and Recognition. Presented at the 3rd International Conference on Document Analysis and Recognition, pp. 278–282 vol.1. https://doi.org/10.1109/ICDAR.1995.598994

Maxwell, A.E., Warner, T.A., Fang, F., 2018. Implementation of machine-learning classification in remote sensing: an applied review. Int. J. Remote Sens. 39, 2784–2817. https://doi.org/10.1080/01431161.2018.1433343

Opitz, D., Maclin, R., 1999. Popular Ensemble Methods: An Empirical Study. J. Artif. Intell. Res. 11, 169–198. https://doi.org/10.1613/jair.614

Pal, M., Mather, P.M., 2005. Support vector machines for classification in remote sensing. Int. J. Remote Sens. 26, 1007–1011. https://doi.org/10.1080/01431160512331314083

Pan, Z., Wang, Y., Pan, Y., 2020. A new locally adaptive k-nearest neighbor algorithm based on discrimination class. Knowl.-Based Syst. 204, 106185. https://doi.org/10.1016/j.knosys.2020.106185

Ronneberger, O., Fischer, P., Brox, T., 2015. U-Net: Convolutional Networks for Biomedical Image Segmentation, in: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (Eds.), Medical Image Computing and Computer-Assisted Intervention

– MICCAI 2015. Springer International Publishing, Cham, pp. 234–241. https://doi.org/10.1007/978-3-319-24574-4_28

UNOSAT [WWW Document], 2025. URL https://unosat.org/products/3300 (accessed 10.30.25).

Verner, I., 2025. STATE STATISTICS SERVICE OF UKRAINE.

Zanaga, D., Van De Kerchove, R., Daems, D., De Keersmaecker, W., Brockmann, C., Kirches, G., Wevers, J., Cartus, O., Santoro, M., Fritz, S., Lesiv, M., Herold, M., Tsendbazar, N.-E., Xu, P., Ramoino, F., Arino, O., 2022. ESA WorldCover 10 m 2021 v200. https://doi.org/10.5281/zenodo.7254221

Zhang, C., Sargent, I., Pan, X., Li, H., Gardiner, A., Hare, J., Atkinson, P.M., 2018. An object-based convolutional neural network (OCNN) for urban land use classification. Remote Sens. Environ. 216, 57–70. https://doi.org/10.1016/j.rse.2018.06.034

Zhong, Y., Zhang, L., Huang, B., Li, P., 2006. An unsupervised artificial immune classifier for multi/hyperspectral remote sensing imagery. IEEE Trans. Geosci. Remote Sens. 44, 420–431. https://doi.org/10.1109/TGRS.2005.861548