applied sciences

Article

Assessing Land Cover Changes Using the LUCAS Database and Sentinel Imagery: A Comparative Analysis of Accuracy Metrics

Beata Hejmanowska * and Piotr Kramarczyk

Department of Photogrammetry, Remote Sensing, and Spatial Engineering, Faculty of Geo-Data Science, Geodesy and Environmental Engeneerinf, AGH University, al. A. Mickiewicza 30, 30-059 Krakow, Poland; piotrkramarczyk@proton.me

* Correspondence: galia@agh.edu.pl

Featured Application: This research underscores the feasibility of utilizing the Copernicus LUCAS 2018 database, including in its original form, to classify basic land cover and land use types across Europe, with overall accuracy ca. 80%. The precise use of classification accuracy metrics enables reliable accuracy analyses across various ML models. The "overall accuracy OA" metric should be reported instead of, or alongside, "average accuracy ACC" to ensure clarity and comparability.

Abstract: Classification of remote sensing images using machine learning models requires a large amount of training data. Collecting this data is both labor-intensive and time-consuming. In this study, the effectiveness of using pre-existing reference data on land cover gathered as part of the Land Use–Land Cover Area Frame Survey (LUCAS) database of the Copernicus program was analyzed. The classification was carried out in Google Earth Engine (GEE) using Sentinel-2 images that were specially prepared to account for the phenological development of plants. Classification was performed using SVM, RF, and CART algorithms in GEE, with an in-depth accuracy analysis conducted using a custom tool. Attention was given to the reliability of different accuracy metrics, with a particular focus on the widely used machine learning (ML) metric of "accuracy", which should not be compared with the commonly used remote sensing metric of "overall accuracy", due to the potential for significant artificial inflation of accuracy. The accuracy of LUCAS 2018 at Level-1 detail was estimated at 86%. Using the updated LUCAS dataset, the best classification result was achieved with the RF method, with an accuracy of 83%. An accuracy overestimation of approximately 10% was observed when reporting the average accuracy ACC metric used in ML instead of the overall accuracy OA metric.

Keywords: LUCAS; Google Earth Engine; Sentinel-2

check for updates

Academic Editors: Samuel Adelabu, Romano Lottering and Kabir Peerbhay

Received: 4 November 2024
Revised: 25 December 2024
Accepted: 26 December 2024
Published: 30 December 2024

Citation: Hejmanowska, B.; Kramarczyk, P. Assessing Land Cover Changes Using the LUCAS Database and Sentinel Imagery: A Comparative Analysis of Accuracy Metrics. *Appl. Sci.* **2025**, *15*, 240. https://doi.org/10.3390/app15010240

1. Introduction

Europe's Land Use–Land Cover Area Frame Survey (LUCAS) [1] is a field survey conducted every three years, covering the entire European Union and providing standardized data on land use, land cover, agro-environmental topics, and soil. The data for LUCAS are collected by field surveyors at sample points arranged in a 2 km grid, with only a subset of points surveyed each year [2–4]. The collected information supports the production and validation of Copernicus program products, including land use classification and data quality control. Initiated by EUROSTAT, LUCAS is co-funded by the European Commission and updated in collaboration with EU member states. In light of increasing satellite data volumes, the challenge lies in the need for more frequent updates to effectively support the Copernicus program.

In 2022–2023, the European Environment Agency (EEA) achieved significant milestones in managing in situ data [5], key for supporting Copernicus Services, ESA, and EUMETSAT. A major accomplishment was launching the Copernicus In Situ Information System (CIS2), which tracks data needs and sources, complemented by new Country Reports emphasizing each nation's contributions to Copernicus. The EEA also reviewed challenges in geospatial data access, such as harmonizing land data, and began utilizing crowdsourced and citizen science data for future applications [6]. Looking ahead, 2023–2024 plans include updating the State of Play Report, enhancing CORDA's functionalities for data access, and expanding collaborations with data providers, including EuroGeographics.

A query in the Scopus database using TITLE-ABS-KEY (LUCAS 2018) and limiting the subject areas to "AGRI", "ENVI", "EART", and "BIOC" yielded 77 publications. When limited to the years 2024–2025, this number was reduced to 15 documents ([7–21] after the exclusion of one incorrect answer). Among these, only two publications focused on LULC analysis [7,8], while nine documents pertained to soil studies [9–17] and four to crop recognition [18–22].

Particular attention should be given to one publication [7], which presents LULC classification results using machine learning techniques and the Sen4Map reference database in the context of LUCAS 2018. The authors classified Sentinel-2 imagery from 2018 and compared the results with LUCAS 2018 data, achieving a maximum accuracy of 0.93 with the Video Vision Transformer method and 0.89 using the Random Forest (RF) method.

Another noteworthy publication from 2024 describes a semi-supervised approach for mapping European forest types using Sentinel-1 and Sentinel-2 data, with training and validation supported by the LUCAS Copernicus 2018 and 2022 datasets [22]. The 2022 forest type map distinguishes broadleaved, coniferous, and mixed forests at a 10 m resolution, achieving 93% accuracy.

Additionally, two earlier publications deserve mention [23,24]. In one study [23], LUCAS in situ reference data were utilized to classify high-resolution Sentinel-2 imagery on a large scale. Several pre-processing schemes (PSs) were tested, involving positioning and semantic selection approaches to enhance automated classification at the national level. The impact of these schemes on a Random Forest classifier was evaluated in an iterative training process. Results showed that positional corrections of LUCAS points significantly improved overall classification accuracy, with all PSs achieving accuracies above 80%, with an average of 93.1%. This demonstrates that LUCAS data represent a robust reference source for large-scale, high-resolution land cover mapping using Sentinel-2 imagery. Another frequently cited publication [24] explored the integration of LUCAS land survey data with Landsat-8 imagery to map 12 land cover and land use classes across Europe. The study leveraged spectral–temporal metrics to capture seasonal variations and mitigate cloud interference. The resulting pan-European land cover map achieved 75.1% accuracy, outperforming the CORINE dataset and showcasing the potential of combining large-scale survey datasets with advanced remote sensing techniques.

In the literature review on the use of the LUCAS database for LULC classification, two research scales can be distinguished: national-scale classification [25–29] and Pan-European mapping [22,24,30–32]. In all reviewed studies, classifications were performed using remote sensing imagery corresponding to the temporal coverage of the LUCAS database. Furthermore, all of the above-mentioned papers provide sufficient information regarding the accuracy assessment of classification results. However, not all publications include adequate details on this topic. This issue is discussed in greater depth in the discussion section.

Through our literature review, we identified two key gaps in the field:

- A lack of studies exploring the broader potential of in situ measurement data collected in the LUCAS database, particularly archival datasets, for land use classification.
- A critical gap in understanding how different accuracy metrics (e.g., "accuracy" in machine learning versus "overall accuracy" in remote sensing) may lead to misinterpretation when evaluating methods discussed in the literature.

This article fills these gaps by demonstrating the practical use of archival LUCAS 2018 data for LULC classification in 2023 [1–32]. It also provides a thorough examination and practical application of accuracy metrics commonly used in ML and remote sensing, illustrating how these metrics can lead to misinterpretations in comparative analyses. Section 2 provides an overview of the study area as well as the data utilized and their processing and analysis. The study area is located in Poland near the large urban agglomeration of Kraków and covers approximately 5000 square kilometers. The available LUCAS database for the analyzed area was reviewed. The workflow of the analyses is presented schematically, along with a detailed methodology for preparing the LUCAS database for further analysis and the preprocessing and classification of Sentinel-2 remote sensing data [33–36]. Various accuracy metrics are discussed in detail [37–40]. Section 3 presents the outcomes of the analyses conducted according to the described methodology. In Section 4, the findings of our research are compared and contextualized with the existing literature [41–47]. Finally, the Conclusions section summarizes the key insights derived from our study.

2. Materials and Methods

The study focuses on land use and land cover (LULC) classification in a test site near Krakow, Poland, a region characterized by its heterogeneous landscape of agricultural fields, grasslands, woodlands, and urban zones. This diversity makes the area an ideal location for testing classification methods capable of distinguishing complex LULC types. Furthermore, the region exemplifies transitions typical of Central Europe, such as urban expansion, shifts between agricultural and grassland zones, and the persistence of mixed-use landscapes. Investigating these dynamics provides insights relevant to broader geographic and environmental studies. The LUCAS database was selected as the reference dataset due to its unique in situ measurements as derived from a highly dense sampling network across Europe. The primary source of remote sensing data was Sentinel-2, currently the most spatially detailed free satellite imagery available. The cloud-based platform Google Earth Engine (GEE) was employed for this research, serving both as a repository for the Sentinel-2 dataset and as a tool for classification and basic accuracy analyses. Machine learning algorithms available within GEE were utilized for image classification: Random Forest (RF), Support Vector Machine (SVM), and Classification and Regression Trees (CART). The choice of these algorithms was driven by their complementary strengths in addressing the complexities of land use and land cover (LULC) classification. Each algorithm offers unique advantages tailored to specific challenges of the dataset and classification objectives.

- Random Forest (RF): RF is a powerful ensemble learning algorithm capable of handling large, high-dimensional datasets. Its ability to aggregate multiple decision trees helps mitigate overfitting and improves classification accuracy, particularly in complex and heterogeneous environments.
- Support Vector Machine (SVM): SVM is well suited for datasets with non-linear decision boundaries, a common characteristic in LULC classification due to the overlapping spectral signatures of various land cover types. The radial basis function (RBF) kernel enables SVM to model intricate relationships effectively, making it especially useful for distinguishing subtle differences between classes such as urban areas and arable

land. Furthermore, SVM's efficiency with relatively small training datasets adds to its suitability.

- Classification and Regression Trees (CART): While simpler than RF and SVM, CART offers an interpretable baseline model. Its clear decision tree structure provides insights into the classification process, making it valuable for understanding the influence of specific features on class differentiation.

By leveraging these algorithms within GEE, the study aims to establish an optimal methodology for LULC classification in Central European landscapes, while creating a replicable workflow applicable to other regions with diverse and complex environments.

Due to the lack of advanced accuracy metric computation functionality in GEE and other available tools, a detailed accuracy analysis was conducted using custom Python 3.10 scripts. These scripts enabled the calculation of comprehensive metrics commonly used in both remote sensing and machine learning to evaluate classification performance.

2.1. Test Area

The test site for this study is located near Krakow, Poland, encompassing approximately 50 × 100 km. This area was selected for its highly heterogeneous landscape, which includes agricultural fields, grasslands, woodlands, and urban zones. These features reflect the diverse land use and land cover (LULC) patterns characteristic of Central Europe.

The LUCAS (Land Use/Cover Area frame Survey) dataset provides detailed field-collected information on land use and land cover across Europe, making it valuable for remote sensing and land classification studies. Key attributes in LUCAS include land cover type, land use type, and additional descriptors like soil, vegetation, and crop type specifics. The survey categorizes land cover into broad classes such as cropland, woodland, grassland, shrubland, and artificial surfaces, with detailed subclasses for more precise analysis. Updated periodically, LUCAS serves as ground-truth data for validating satellite-based classifications and tracking land cover changes across Europe.

LUCAS 2018 and LUCAS 2022 data are available and can be downloaded from the internet (https://figshare.com/articles/dataset/LUCAS_2018_Copernicus/12382667, accessed on 25 December 2024) as vector files. The 2018 versions were chosen for the study due to the greater diversity of land cover/use forms (Table 1).

Table 1. LUCAS data.

Title 1	PL	ROI
LUCAS 2018	23,087	438 more balanced classes
LUCAS 2022	10,576	304 mainly crops in detail

The LUCAS 2018 database was used as reference data, which was initially prepared by selecting the following classes: Artificial land, Arable land, Woodland and Shrubland, Grassland, and Water (Table 2).

Table 2. LUCAS classes selected for analysis.

Level-1	Classes	LUCAS Code	Code in Our Research	Color on the Maps
100	Artificial land	A	3	red
200	Arable land	B	4	yellow
300	Woodland and Shrubland	C	1	dark green
500	Grassland	E	6	light green
700	Water	G	**8**	blue

Figure 1 displays the location of the test site, situated in southern Poland near the major urban area of Krakow (Region of Interest, ROI). The entire area of Poland is densely covered, with 23,097 in situ measurement points marked in pink on the map. Within the ROI, there are 438 measurement points, highlighted in various colors. Figure 2 provides a zoomed-in view of the ROI.

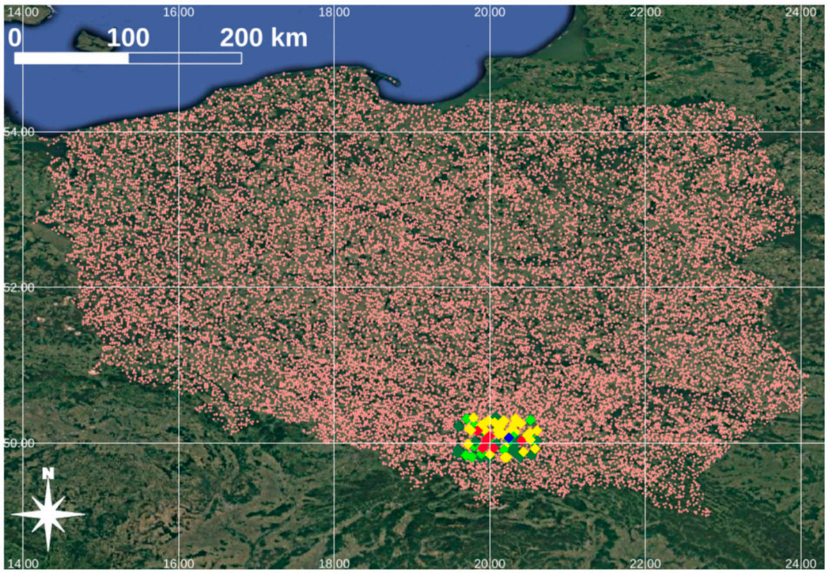

Figure 1. LUCAS 2018 (23,087 points) image of the whole country (Poland) with the ROI (Region Of Interest) near Kraków (438 points).

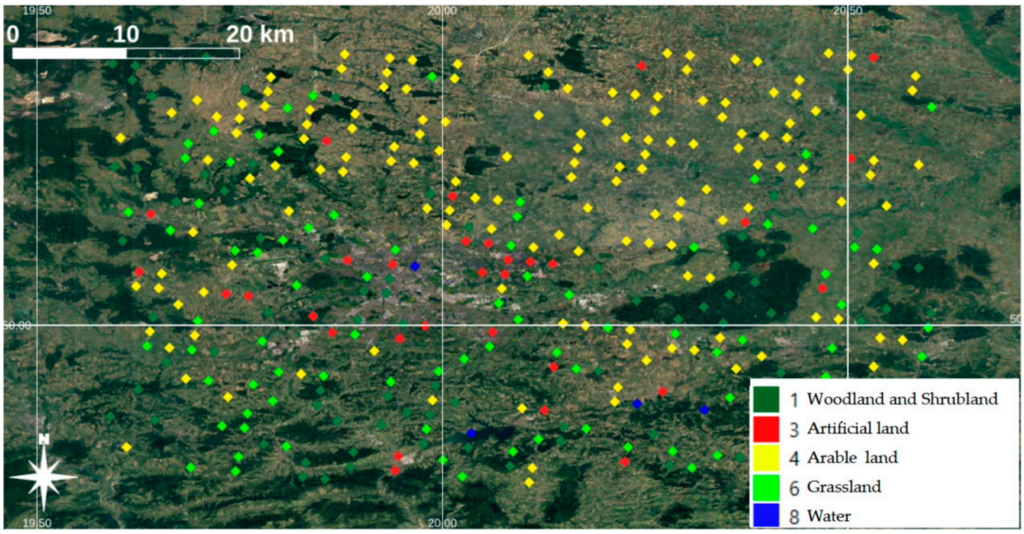

Figure 2. LUCAS 2018 (438 points) overlaid on Kraków ROI (100 × 50 km).

2.2. Workflow

The schematic outline of the conducted research is shown in Figure 3. In the first stage, data were extracted from the LUCAS 2018 database, and samples with specific codes were selected and validated using up-to-date orthophotos from 2023. The data were then divided into training (80%) and testing (20%) sets. In the next step, Sentinel-2 satellite data from 2023 were prepared, and classification was performed using three methods: Random Forest (RF), Support Vector Machine (SVM), and CART. Finally, an accuracy analysis of the results was conducted. The classification was performed in GEE, which meant that any registration period for the collection could be specified. In our study, we conducted classifications for the data from both 2023 and 2018. The code is publicly available, including the LUCAS 2018 data, so anyone can perform classification for any time period.

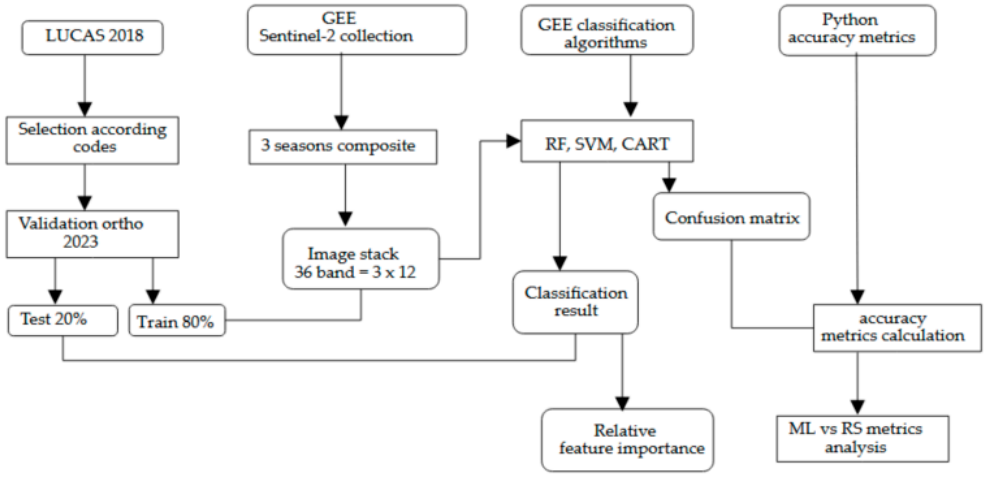

Figure 3. Data processing scheme.

2.3. Preparation of the LUCAS Database

In the LUCAS 2018 database, which covers all of Poland, a total of 23,087 points were surveyed to capture data on land cover, land use, and related environmental features. For this specific research study, a subset of 438 points was initially selected within the designated study area. To ensure the accuracy and relevance of these points for the current analysis, each point underwent a visual validation process using recent high-resolution orthophoto imagery available from Google Maps.

The validation process for the LUCAS 2018 database involved a careful visual verification of Land Use and Land Cover (LULC) types to ensure accuracy in QGIS. Each data point was cross-checked with high-resolution satellite imagery to confirm its LULC classification. When discrepancies were identified between the database entries and the observed land cover in the imagery, the correct LULC code was assigned to replace the inaccurate entry. This manual adjustment aimed to improve the reliability of the data for further analysis and classification tasks. Examples of corrected classifications are provided in the accompanying Figure 4, illustrating common types of discrepancies and the validation approach used to rectify them. This visual verification step was essential to ensure that the LUCAS database reflected current land cover accurately, thereby enhancing its utility as a reference dataset in remote sensing and land cover studies.

During this verification, 63 points were identified as potentially inaccurate or inconsistent with the present land cover. These points were subsequently excluded from the dataset, resulting in a refined set of 375 points. This removal represents approximately 15% of the original sample points in the study area, underscoring the importance of careful validation when using large, multi-year databases like LUCAS. Finally, the LUCAS 2018 point dataset, after validation, was divided into two subsets: a training set (80%, 298 points) and a testing set (20%, 77 points). Both subsets were imported into Google Earth Engine as shapefiles (https://code.earthengine.google.com/?asset=projects/urban-atlas-benchmark/assets/lucas_ver_train_80_validation_after, accessed on 25 December 2024, https://code.earthengine.google.com/?asset=projects/urban-atlas-benchmark/assets/lucas_ver_train_20_validation_after, accessed on 25 December 2024).

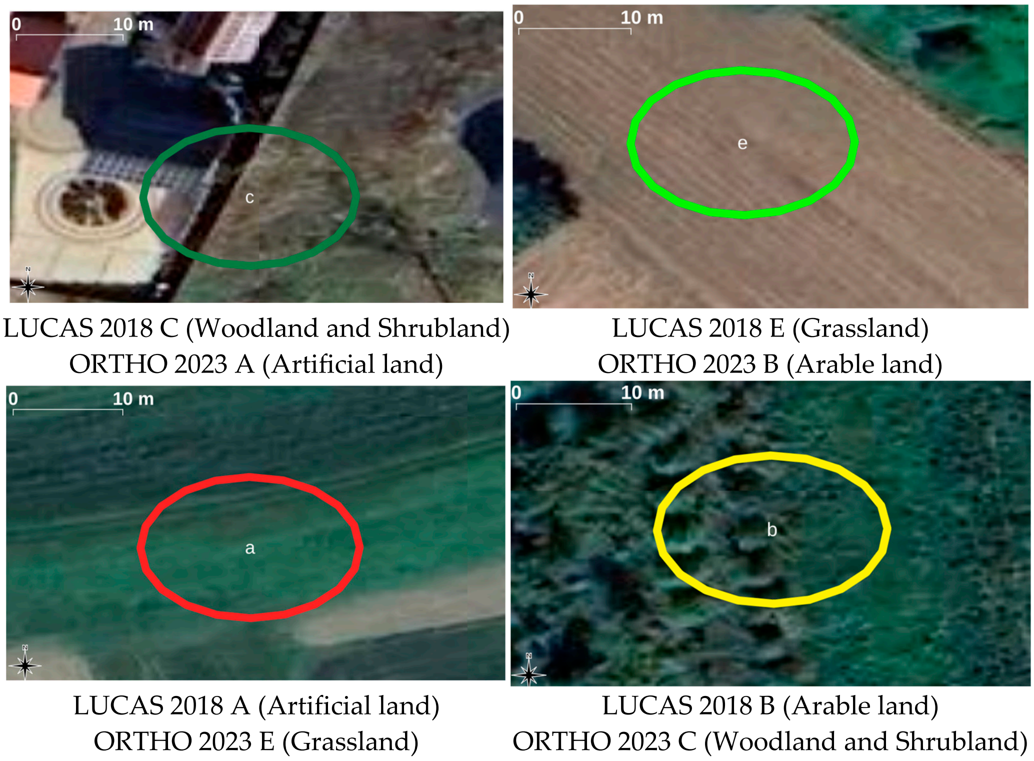

Figure 4. LUCAS 2018 validation; codes and colors are specified in Table 2.

2.4. Remote Sensing Data Processing

The research employed Sentinel-2 satellite images, accessed from the Google Earth Engine (GEE) collection ee.ImageCollection ("COPERNICUS/S2_SR_HARMONIZED"), provided by the European Union, ESA, and the Copernicus program. To ensure high-quality imagery suitable for land cover analysis, a cloud and shadow masking approach was applied. Specifically, only pixels with a cloudiness or shadow probability of 10% or lower were retained, effectively filtering out much of the interference from clouds or cloud shadows that can hinder accurate land cover classification. Additionally, only scenes with a total cloud cover not exceeding 30% were selected from the designated observation period. This selection criterion helped maximize the clarity and usability of the data for accurate mapping and classification. Following this pre-processing, the selected images were grouped into three distinct seasonal composites, representing spring, summer, and autumn phenological phases of vegetation growth [33]. These seasonal composites were then combined into a multi-layered image stack, enabling more comprehensive data analysis across different periods within the growing season (Figure 3). A Sentinel-2 image for a single season consists of 12 bands: B1, B2, B3, etc. (excluding the B10—Cirrus band). Three seasons were analyzed: spring, summer, and autumn, resulting in a total of 36 bands in the stack (3×12).

The classification was performed in Google Earth Engine using the imported training dataset and the Sentinel-2 stack for 2023. The code is publicly accessible, allowing anyone to perform classification for any time period (https://code.earthengine.google.com/526e3bb5ff3da58af2a06ff5e9725b32, accessed on 25 December 2024).

The built-in classifiers available in GEE, Random Forest (RF), Support Vector Machine (SVM), and CART, were applied with default settings. The Random Forest (RF) algorithm [34], based on random decision trees, provides high accuracy for large datasets. The following parameters were adapted for the classification: number of trees = 10; maximum number of variables used for splitting = $(12)^{0.5}$, approximately 3 or 4. The Support Vector Machine (SVM) classifier [35], based on a decision boundary function, performs well with complex data. The classification used the following parameters: kernel type = radial basis function (RBF); error tolerance = 0.001. The CART (Classification and Regression Trees) algorithm [36] employs a single decision tree. In our study, the maximum tree depth was not limited, and the minimum number of data points in a leaf node was set to 1. The matching of model parameters and their validation were performed on the training set. The final accuracy evaluation was performed on a test set that the models had never seen. We utilized the confusion matrix, generated in GEE from the test dataset and exported as a text file, to compute all accuracy metrics using a Python script.

2.5. Accuracy Metrics

Classification accuracy metrics have been a focus of research for many years [37–40], always based on confusion matrix (Tables 3 and 4). Nevertheless, remote sensing experts consistently utilize specific metrics, such as those listed in Table 5 (Overall Accuracy (OA), Producer Accuracy (PA), User Accuracy (UA)) [38]. With the increasing adoption of machine learning algorithms, many publications now report metrics that may be more familiar to machine learning practitioners, who may lack expertise in remote sensing.

Table 3. Binary confusion/error matrix.

	Predicted Class	
True Class	True positive (TP)	False negative (FN)
	False positive (FP)	True negative (TN)

OA = ACC = (TP + TN)/(FP + FN).

Table 4. Many-class confusion/error matrix.

	Predicted Class		
True Class	A	B	C
	a11	**a12**	**a13**
	a21	a22	a23
	a31	a32	a33

OA = (a11 + a22 + a33)/(a11 + a12 + a13 + a21 + a22 + a23 + a31 + a32 + a33); ACC = (a11 + a22 + a23 + a32 + a33)/(a11 + a12 + a13 + a21 + a22 + a23 + a31 + a32 + a33).

Table 5. Accuracy metrics used in remote sensing.

Metrics	Formula
OA (overall_accuracy)	sum (TP)/(TP + TN + FP + FN)
PA (producer_accuracy)	TP/(TP + FN)
UA (user_accuracy)	TP/(TP + FP)
OME (omission errors)	FN/(TP + FN)
CME (errors_of_commision)	FP/(TP + FP)

In the medical field, image processing commonly leads to classification metrics that assess whether a patient has a particular disease. This represents a classic binary classification scenario: a positive test result indicates a sick patient (True Positive, TP), a positive test result indicates a healthy patient (False Positive, FP), a negative test result indicates a healthy patient (True Negative, TN), and a negative test result indicates a sick patient (False Negative, FN). Common evaluation metrics for such tests include accuracy, sensitivity, and specificity. In medical testing, accuracy and sensitivity metrics are critical, especially concerning TN cases. However, this emphasis differs in remote sensing, where TN instances are not as significant.

This distinction can be illustrated with an example. Table 3 presents a confusion matrix typical of the binary classification found in medical tests, where OA equals ACC. However, in Table 4, which contains only three classes, the scenario changes. Here, ACC is notably greater than OA due to a substantial presence of TN (olive color), TP (red), FP (orange), and FN (red).

Table 5 lists the metrics used in remote sensing, Table 6 presents metrics relevant to machine learning. It is important to note that while the names of these metrics may vary, their underlying concepts remain the same. For instance, Sensitivity, also referred to as the True Positive Rate (TPR), is equivalent to PA (Producer Accuracy), and Precision or Positive Predictive Value (PPV) is synonymous with UA (User Accuracy), as summarized in Table 7.

Table 6. Accuracy metrics used in machine learning.

Metrics	Formula
ACC (accuracy)	(TP + TN)/(TP + TN + FP + FN)
PPV (precision or positive predictive value)	TP/(TP + FP)
TPR (sensitivity, recall, hit rate, or true positive rate)	TP/(TP + FN) = 1 − FNR
TNR (specificity, selectivity, or true negative rate)	TN/(TN + FP) = 1 − FPR
NPV (negative predictive value)	TN/(TN + FN) = 1 − FOR
FNR (miss rate or false negative rate)	FN/(FN + TP) = 1 − TPR
FPR (fall-out or false positive rate)	FP/(FP + TN) = 1 − TNR
FDR (false discovery rate)	FP/(FP + TP) = 1 − PPV
FOR (false omission rate)	FN/(FN + TN) = 1 − NPV
S/CSI (Threat score (TS) or critical success index (CSI))	TP/(TP + FN + FP)
MCC (Matthews correlation coefficient)	$(TP * TN - FP * FN)/[(TP + FP) * (TP + FN) * (TN + FP) * (TN + FN)]^{0.5}$

Table 7. Common accuracy metrics used in remote sensing and machine learning.

Metrics
OA (overall_accuracy) NOT ACC
PA (producer_accuracy) = sensitivity, recall/true positive rate (TPR)
UA (user_accuracy) = precision/positive predictive value (PPV)
OME = FNR (miss rate or false negative rate)
CME = FDR (false discovery rate)

2.6. Relative Feature Importance

Relative feature importance measures the contribution of each feature to the performance of a machine learning model. It highlights which variables have the greatest impact on the model's predictions, allowing for better understanding and interpretation of the results. In land use and land cover (LULC) classification, feature importance can reveal which spectral bands, indices, or temporal characteristics are most influential in distinguishing between different classes. This insight helps refine models, optimize data inputs, and prioritize variables for future analyses. In our study, the analysis of relative feature importance focused on the spectral bands of Sentinel-2 across three seasons: spring, summer, and autumn. A total of 36 features were analyzed, corresponding to the 12 spectral bands for each season: B1_spring through B12_spring, B1_summer through B12_summer, and B1_autumn through B12_autumn. This approach allowed us to assess the seasonal contributions of individual bands to the classification process and understand their relative significance in distinguishing land use and land cover types.

3. Results

In this chapter, the results of the conducted research are presented. The first part discusses the results of the validation of the LUCAS 2018 data based on the current orthophoto map. The second part presents and compares the results of classification using the algorithms RF, SVM, and CART. The final subsection contains an analysis of the relative feature importance of the classified Sentinel-2 stack, which includes three seasonal Sentinel-2 sets (36 features/bands).

3.1. LUCAS 2018 Validation

To evaluate the discrepancies in land use and land cover types between 2018 and 2023, the previously mentioned metrics can be employed. According to the discrepancy matrix (Table 8), the overall accuracy (OA) is reported as 0.8562, while the average accuracy (ACC) is approximately 10% higher, at 0.9425 (Table 9). Currently, analyzing the LUCAS 2018 database makes it challenging to ascertain whether these discrepancies are genuinely attributable to changes in land use and land cover or if the database already contained inaccuracies in 2018. Although it would theoretically be possible to analyze archival orthophotos from that period, such analyses were not conducted as part of this research due to the lack of available orthophotos. Ultimately, any data that did not correspond to the current land use, according to the LUCAS 2018 database, were excluded. As a result, the classifications were carried out using all data from the LUCAS 2018 database, as well as the revised database, which was filtered to remove questionable points. The highest accuracy, equal to or above 90%, was observed for the classes: Water, Grassland, and Arable land. The accuracy for the other two classes, Artificial land as well as Woodland and Shrubland, was around 78%.

Table 8. Matrix of discrepancies between 2018 and 2023.

Predicted/True	1	3	5	6	8
1	91	5	10	12	0
3	0	30	4	4	0
4	0	1	176	16	0
6	2	4	5	73	0
8	0	0	0	0	5

Table 9. Accuracy metrics: OA= 0.8562, ACC average = 0.9425.

Code	ACC	PA	UA	OME	CME
1	0.9338	0.7712	0.9785	0.2288	0.0215
3	0.9589	0.7895	0.75	0.2105	0.25
4	0.9178	0.9119	0.9026	0.0881	0.0974
6	0.9018	0.869	0.6952	0.131	0.3048
8	1	1	1	0	0

3.2. Evaluation of Classification Methods and Accuracy Metrics for Land Cover Mapping Using the LUCAS Database

Figure 5 illustrates the classification results for the Kraków ROI obtained from the RF, SVM, and CART algorithms. Based on a visual assessment of the maps, the RF algorithm provided the most accurate classification, achieving an OA of 0.8309 (compare Table 10 for 2019 with Table 11 for 2023). The SVM algorithm exhibited notable weaknesses in the classification of built-up areas, while the CART algorithm faced challenges with water body classification. In the figures, unclassified pixels are displayed in white, and LUCAS 2018 points are marked as black dots. Visual comparisons of the classification results with the accuracy metrics in Table 11 show that both RF and SVM yielded the highest accuracy (OA = 0.8309). In contrast, the CART algorithm achieved an OA of 0.7347.

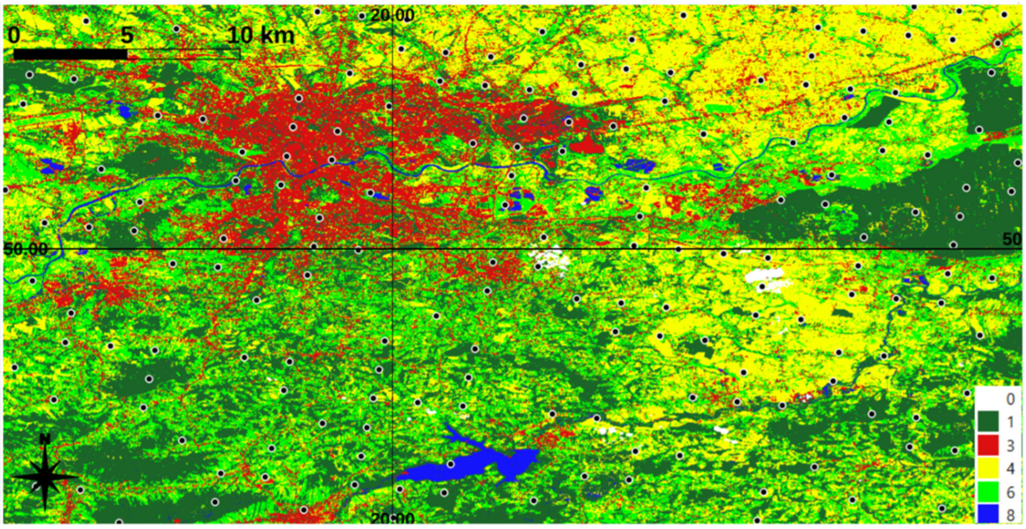

Figure 5. Results of RF classification for 2023: unclassified areas are shown in white (code 0), while LUCAS 2018 data points are represented as black dots.

Table 10. Classification accuracy metric (OA) for the 2019 Sentinale-2 collection.

2019	CART	RF	SVM
LUCAS 2018	0.6615	0.7891	0.7708
LUCAS 2018 validated	0.7085	**0.8630**	0.8338

Table 11. Classification accuracy metric (OA) for the 2023 Sentinale-2 collection.

2023	CART	RF	SVM
LUCAS 2018	0.6979	0.7838	0.7760
LUCAS 2018 validated	0.7347	**0.8309**	**0.8309**

Figures 6 and 7 highlight specific misclassification patterns, particularly for the Water class (Class 8). In the SVM results, significant overestimation of built-up areas as water bodies is evident. This issue is even more pronounced in CART classifications, where most water-covered areas are incorrectly identified as Woodland and Shrubland (Class 1). Zoomed-in sections (Figures 8 and 9) provide detailed views of these misclassifications for RF and SVM, respectively.

Tables 11 and 12 summarize the classification accuracy metrics for 2019 and 2023, comparing results derived from the original LUCAS 2018 database with those from the modified, validated database. The highest accuracy (OA = 0.8630) was achieved in 2019 using the RF algorithm on the validated database. In 2023, the accuracy slightly decreased to 0.8309 for both RF and SVM algorithms. The original database consistently yielded lower accuracy, with discrepancies of 5% to 7%, highlighting potential errors in the LUCAS 2018 dataset.

Table 12. Confusion matrix RF 2023 on LUCAS 2018 validated data.

Predicted/True	1	3	5	6	8
1	73	0	4	5	2
3	1	26	1	3	0
4	4	0	150	8	0
6	4	7	17	34	0
8	2	0	0	0	2

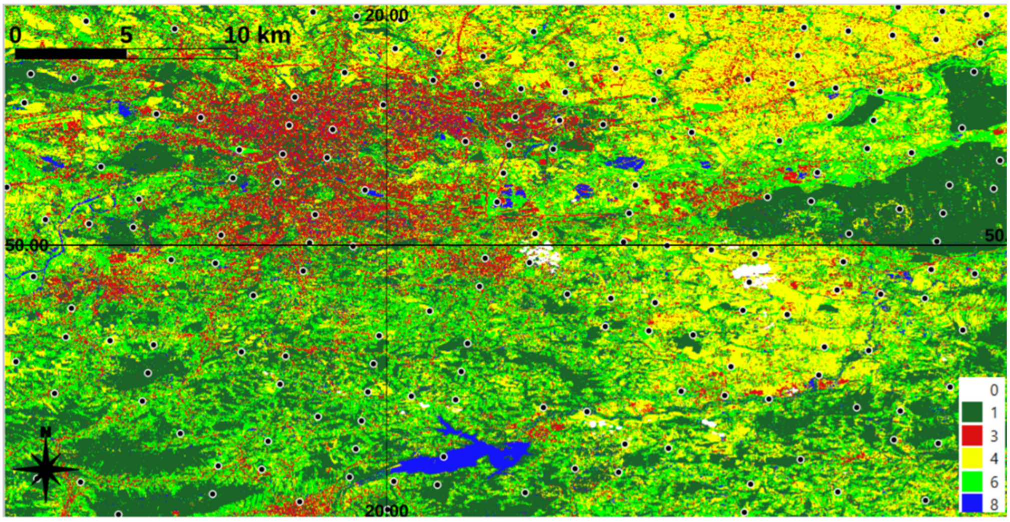

Figure 6. Results of SVM classification for 2023: unclassified areas are shown in white (code 0); overestimated areas are marked as code 1; underestimated areas are marked as code 3. LUCAS 2018 data points are represented as black dots.

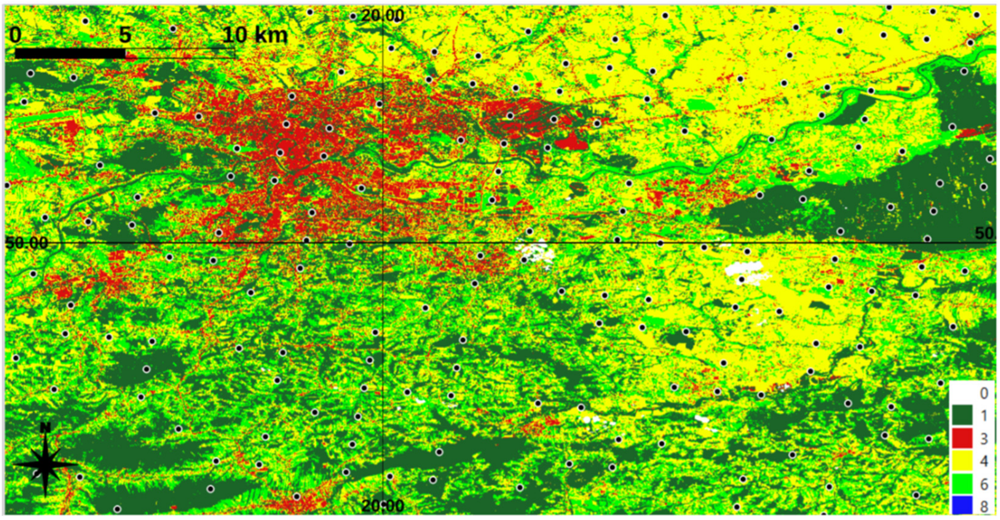

Figure 7. Results of CART Classification for 2023: Unclassified areas are shown in white (code 0); code 8 indicates improper classification of water; LUCAS 2018 data points are represented as black dots.

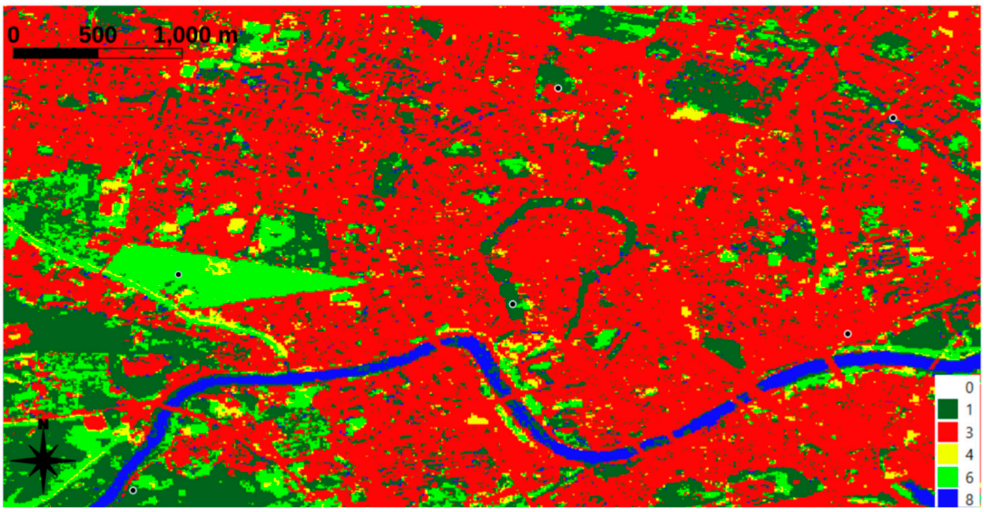

Figure 8. Zoomed-in section of Figure 5: RF classification—Center of Krakow City.

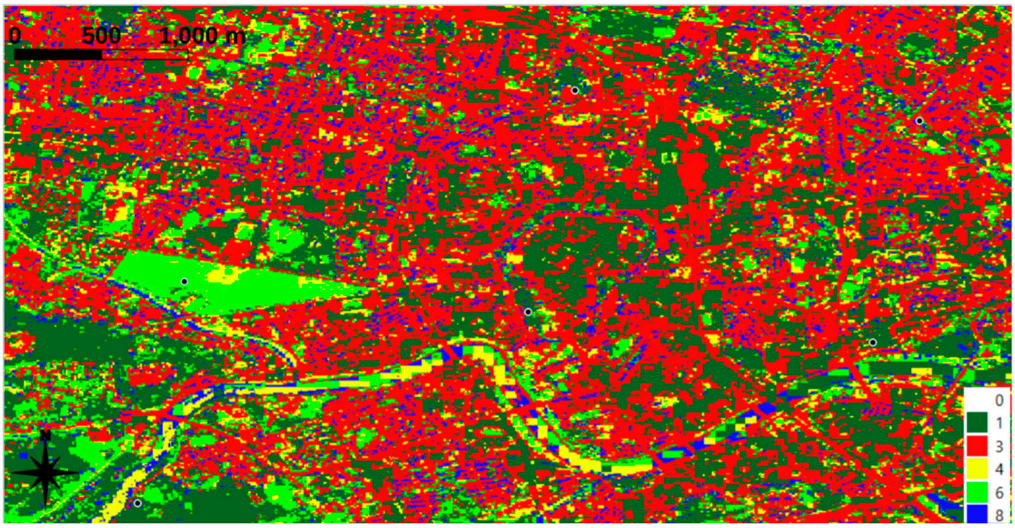

Figure 9. Zoomed-in section of Figure 6: SVM classification—center of Krakow City; overestimation is class code 8; incorrect Water classification, see Vistula River.

Interestingly, the average accuracy (ACC) metric, at 0.9324, appeared inflated compared to OA. This discrepancy emphasizes the need to carefully interpret accuracy metrics in the context of classification performance.

From the results summarized in Tables 12–16, the following insights can be drawn:

- The average accuracy (ACC) is significantly higher than the Overall Accuracy (OA) (0.93 vs. 0.83, respectively).
- For each class, ML accuracy metrics (ACC, NPV, TNR) are consistently high, averaging around 0.90 or above.
- ML error metrics are low, typically below 0.10, and much smaller than remote sensing metrics such as OME and CME.
- The corresponding metrics from remote sensing and ML approaches are closely aligned.

Table 13. Binary confusion matrix RF 2023 on LUCAS 2018 validated data (notice TN).

Code	1	3	5	6	8
TP	73	26	150	34	2
TN	248	305	159	265	337
FP	11	7	22	16	2
FN	11	5	12	28	2

Table 14. RF 2023 accuracy metrics: OA= 0.8309, ACC average = 0.9324 (compare PA and UA with tpr and ppv in Table 16; OME and CME with FNR and FDR in Table 15).

Code	ACC	PA	UA	OME	CME
1	**0.9359**	0.8690	0.869	0.131	0.131
3	**0.9650**	0.8387	0.7879	0.1613	0.2121
4	**0.9009**	0.9259	0.8721	0.0741	0.1279
6	**0.8717**	0.5484	0.6800	0.4516	0.3200
8	**0.9883**	0.5000	0.5000	0.5000	0.5000

Table 15. RF 2023 ML accuracy metrics (notice "error" metrics foRate and fpr and compare with OME and CME Table 14).

Code	fdr	fnr	foRate	fpr	mcc
1	0.1310	0.1310	0.0425	0.0425	0.8266
3	0.2121	0.1613	0.0161	0.0224	0.7937
4	0.1279	0.0741	0.0702	0.1215	0.8031
6	0.3200	0.4516	0.0956	0.0569	0.5359
8	0.5000	0.5000	0.0059	0.0059	0.4941

Table 16. RF 2023 ML accuracy metrics.

Code	npv	ppv	tnr	tpr	ts
1	**0.9575**	0.8690	**0.9575**	0.8690	0.7684
3	**0.9839**	0.7879	**0.9776**	0.8387	0.6842
4	**0.9298**	0.8721	**0.8785**	0.9259	0.8152
6	**0.9044**	0.6800	**0.9431**	0.5484	0.4359
8	**0.9941**	0.5000	**0.9941**	0.5000	0.3333

The analysis demonstrates that the RF algorithm performs the best in classifying the Kraków ROI, with the highest overall accuracy achieved using a validated LUCAS database. The findings underscore the importance of database quality, as inaccuracies in the original LUCAS dataset led to lower classification performance. Despite high average accuracy metrics, the observed discrepancies with OA highlight the need for a nuanced evaluation of classification methods, particularly when assessing specific classes like Water. These results contribute to refining classification methodologies and improving land cover mapping reliability.

3.3. Feature Importance

In the analysis of feature importance presented in the Figure 10, we focus on a collection of Sentinel-2 images captured during three distinct vegetation periods: spring, summer, and autumn. This dataset comprises multiple sets of images, each containing 12 spectral channels, with the exception of channel 10, which corresponds to the Cirrus band, known for its limited utility in vegetation analysis.

Out of the total of 36 individual images in the entire collection, a significant finding is that 14 of these images account for 50% of the overall information captured. This statistic highlights the effectiveness of using a selective subset of images for analysis, indicating that not all images contribute equally to understanding the vegetation dynamics over the seasons. Notably, the distribution of these key images reveals a dominance of autumn and spring data, with seven images from autumn and five from spring being particularly informative. In contrast, only two images from the summer season were identified as important, suggesting that the spectral characteristics of vegetation may be more pronounced or more easily discernible in the spring and autumn periods.

Moreover, the analysis indicates a prevalence of infrared channels over visible channels in this dataset. Specifically, there are a total of nine infrared channels compared to just five visible channels. This trend suggests that infrared spectral information is more relevant for capturing the variations in vegetation during these periods. The infrared channels are known for their sensitivity to plant health, moisture content, and biomass, making them critical for understanding vegetation conditions. Therefore, the emphasis on infrared data may provide deeper insights into seasonal changes in vegetation, enhancing the overall analysis and interpretation of the data.

This exploration of feature importance not only underscores the temporal dynamics of vegetation as observed through Sentinel-2 imagery but also emphasizes the significance of selecting appropriate spectral channels for effective monitoring and assessment of ecological and environmental conditions across different seasons.

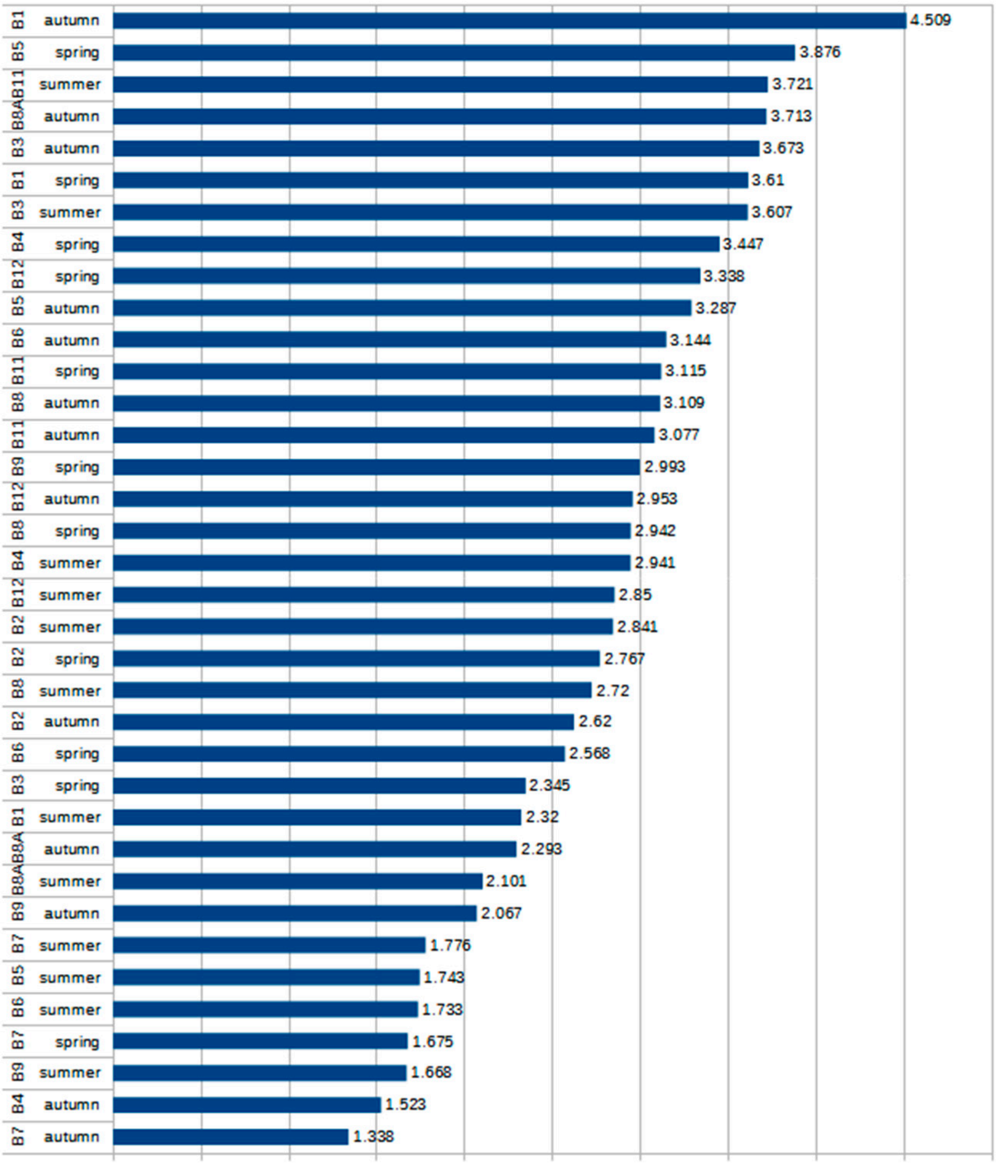

Figure 10. Relative feature importance (Sentinel-2 bands: B1. . .B12 registered in spring, summer, and autumn).

4. Discussion

In the discussion section, we compare our results on the accuracy analysis of the LUCAS database and classification outcomes with findings from other studies in the literature. In the second part of the discussion, we highlight instances of ambiguous accuracy metrics to illustrate the challenges in comparative interpretations.

Preliminary work on crowdsourcing in relation to the LUCAS database can be reviewed in [6]. As part of the research, a mobile application called FotoQuest Go was developed, although it is no longer available. Fieldwork using this application allowed for verification of the LUCAS database. It was found that the overall accuracy for the primary level of detail (Artificial, Cropland, Woodland, Shrubland, Grassland, Bareland, Water, Wetland) is 0.89 (Table 5 [6]). In our study, the discrepancy between LUCAS 2018 and the

2023 orthophotomap was 0.86. Some differences likely resulted from land use changes, such as Grassland to Agricultural land and Woodland/Shrubland to Artificial land. Other discrepancies were due to the unsuitable positioning of the measurement point for remote sensing analyses, and some were due to random error. But, in general, we obtained results consistent with [6]. In our case, it would be more accurate to discuss the currency of the LUCAS 2018 database or the changes that occurred between 2018 and 2023, rather than its accuracy. This implies that for the codes we examined, the current reliability of the LUCAS 2018 database should be considered to be approximately 85%.

Our research results demonstrated that it is feasible to use the old LUCAS 2018 database to classify basic land cover/land use types, with an accuracy level around 80% (78% using LUCAS 2018 and 83% using the 2023-validated LUCAS 2018 database). The accuracy achieved aligns with findings in the existing literature. For instance, in [26], a study combining Sentinel-1 and Sentinel-2 data created a detailed land use and land cover map for the EU-28, using LUCAS 2018 for validation. Monthly and annual features from S2 along with 10-day composites from S1 were classified with a Random Forest algorithm, achieving 78.3% accuracy across 21 classes, which increased to 82.7% when grouped into eight broader categories—improving both accuracy and spatial coverage over using S1 or S2 data alone.

To support detailed crop monitoring in the EU, a new 10 m resolution map was developed for 2022, detailing 19 crop types as an update to the 2018 version [27]. Using LUCAS 2022 field data, Sentinel-1 and Sentinel-2 imagery, land surface temperature, and elevation data, the approach generated two classification layers with Random Forest: a primary map and a gap-filling map for cloud-covered areas. This map, covering 27 EU countries, achieved 79.3% accuracy for major land cover classes and 70.6% for all crop types.

In another study [28], Eurostat's LUCAS 2018 data with Sentinel-2 imagery was used to create a high-detail LULC map for Europe, distinguishing 19 crop types and two broader land categories. The approach generated S2 composites with spectral reflectances, indices, and bio-geophysical indicators, selecting the most relevant features for classification via support vector machine and random forest algorithms. The results showed an overall accuracy of 77.6%.

Similarly, our results are on par with those reported in the literature [26–28]. It should also be noted that we used the LUCAS 2018 archive database in the classification of the Sentinel-2 collection from 2023. Furthermore, in most cases, researchers also achieve the best results using the RF method.

All of the above-mentioned research results, including accuracy metrics, come from remote sensing contexts. The metrics reported adhere to standards that have been in use within remote sensing for years, making the results clear, understandable, and comparable [37–40]. Unfortunately, not all publications present results in a similar manner, which can make interpretation difficult [41–47]. While the issue is complex, it can be illustrated through the use of accuracy metrics, specifically the interchangeable use of "average accuracy" (ACC) and "overall accuracy" (OA), often unintentionally. When a publication reports only the ACC metric as "accuracy" without also providing OA, it can give the impression that the metric reflects the commonly used OA. This can lead to confusion, as average ACC values in multi-class classifications often exceed 90%. Interpreting such results is challenging, especially when dealing with new ML models, where higher accuracy levels are typically expected.

Unfortunately, in the intriguing publication [41] on various ML models, the authors use the metrics ACC and OA interchangeably, comparing them directly. Naturally, the reported accuracies exceed 90%. On the other hand, reference [46] focuses on precision and recall metrics, which correspond to UA and PA rather than OA. This makes it challenging to draw reliable conclusions about the analyzed models.

A typical approach to accuracy estimation in ML image classification on satellite images with deep learning, including code, can be found in [42]. This work explicitly presents the ACC metric along with high values for this metric when comparing different models. Eight models were compared (three variants of RESNET 50, two variants of CNN, and three variants of MobileNet). The differences between seven of the models ranged from 0.946 to 0.993, while one model (MobileNet) recorded an accuracy of 0.852.

In [42], the formula for ACC is presented as OA, and the accuracy of the seven analyzed models exceeds 95%, with differences among them being only a few percentage points. Similarly, other publications [43,44] have employed the ACC metric, reporting accuracies above 90–95%. However, in those cases, the metrics were not referred to as OA.

Our paper demonstrates the discrepancies between ACC and OA values. In both cases examined here (LUCAS 2018 accuracy/divergence analysis and classification accuracy analysis), reporting accuracy with ACC creates an artificial impression of higher accuracy, overestimating it by approximately 10%. Our other studies demonstrate that the smaller the OA, the greater this difference, as ACC often exceeds 90% in multi-class classifications [48].

A notable aspect of our study is the use of the widely accessible Google Earth Engine (GEE) cloud platform, which provides both harmonized Sentinel-2 image collections and seasonal composite analysis algorithms, as well as machine learning algorithms for image classification. Additionally, the LUCAS database is freely available across Europe, enabling its application in any region within Europe. The data and GEE code used in this research have also been shared.

Accuracy analyses, including the computation of accuracy metrics, were conducted using custom Python scripts, which have been made available on GitHub.

5. Conclusions

This article presents research on the feasibility of using the LUCAS 2018 database, even in its original form, to classify basic land cover and land use types in Europe. It also analyzes discrepancies between commonly used accuracy metrics in machine learning (ML) and remote sensing, highlighting the implications of these metrics in land cover classification tasks. The topic should be considered significant, particularly because the entirety of Europe is covered by LUCAS data. Over the years, this project has amassed a vast amount of data. The study utilizes existing datasets (LUCAS 2018) for land classification and validates their accuracy, which can be seen as a substantial contribution to the integration of remote sensing and ML in land use research.

In conclusion, potential future research directions can be outlined as follows:

- Utilizing updated LUCAS datasets and analyzing the performance of additional machine learning algorithms, including Convolutional Neural Networks (CNNs) and deep learning methods.
- Validating classification results using independent datasets from different test areas, accompanied by an in-depth analysis of inconsistencies within the reference datasets.

Author Contributions: Conceptualization, B.H.; methodology, B.H.; software, B.H. and P.K.; validation, B.H. and P.K.; writing, B.H.; project administration, B.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by of the subvention of AGH University of Science and Technology No. 16.16.150.545 and Excellence Initiative—Research University (IDUB D9).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: LUCAS 2018 training data set: https://code.earthengine.google.com/?asset=projects/urban-atlas-benchmark/assets/lucas_ver_train_20_validation_after, accessed on 25 December 2024; LUCAS 2018 testing data set: https://code.earthengine.google.com/?asset=projects/urban-atlas-benchmark/assets/lucas_ver_train_80_validation_after, accessed on 25 December 2024; GEE code: https://code.earthengine.google.com/526e3bb5ff3da58af2a06ff5e9725b32, accessed on 25 December 2024.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Getting to Know LUCAS: The Land Use-Land Cover Area Frame Survey. Available online: https://ec.europa.eu/eurostat/statistics-explained/index.php?title=LUCAS_-_Land_use_and_land_cover_survey (accessed on 25 December 2024).
2. D'andrimont, R.; Yordanov, M.; Martinez-Sanchez, L.; Eiselt, B.; Palmieri, A.; Dominici, P.; Gallego, J.; Reuter, H.I.; Joebges, C.; Lemoine, G.; et al. Harmonised LUCAS in-situ land cover and use database for field surveys from 2006 to 2018 in the European Union. *Sci. Data* **2020**, *7*, 352. [CrossRef]
3. D'Andrimont, R.; Verhegghen, A.; Meroni, M.; Lemoine, G.; Strobl, P.; Eiselt, B.; Yordanov, M.; Martinez-Sanchez, L.; van der Velde, M. LUCAS Copernicus 2018: Earth-observation-relevant in situ data on land cover and use throughout the European Union. *Earth Syst. Sci. Data* **2021**, *13*, 1119–1133. [CrossRef]
4. D'Andrimont, R.; Yordanov, M.; Martinez-Sanchez, L.; Haub, P.; Buck, O.; Haub, C.; Eiselt, B.; van der Velde, M. LUCAS cover photos 2006–2018 over the EU: 874 646 spatially distributed geo-tagged close-up photos with land cover and plant species label. *Earth Syst. Sci. Data* **2022**, *14*, 4463–4472. [CrossRef]
5. Copernicus and its In-Situ Component. Available online: https://insitu.copernicus.eu/ (accessed on 25 December 2024).
6. Laso Bayas, J.C.; See, L.; Bartl, H.; Sturn, T.; Karner, M.; Fraisl, D.; Moorthy, I.; Busch, M.; van der Velde, M.; Fritz, S. Crowdsourcing LUCAS: Citizens Generating Reference Land Cover and Land Use Data with a Mobile App. *Land* **2020**, *9*, 446. [CrossRef]
7. Sharma, S.; Sedona, R.; Riedel, M.; Cavallaro, G.; Paris, C. Sen4Map: Advancing Mapping with Sentinel-2 by Providing Detailed Semantic Descriptions and Customizable Land-Use and Land-Cover Data. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2024**, *17*, 13893–13907. [CrossRef]
8. Bessardon, G.; Rieutord, T.; Gleeson, E.; Pálmason, B.; Oswald, S. High-Resolution Land Use Land Cover Dataset for Meteorological Modelling—Part 1: ECOCLIMAP-SG+ an Agreement-Based Dataset. *Land* **2024**, *13*, 1811. [CrossRef]
9. Al Shoumik, B.A.; Khan, Z.; Błońska, E.; Lasota, J. Dynamics of Soil Organic Carbon and Total Nitrogen in Particulate and Mineral-Associated Organic Matter Fractions Under Different Continuous Land Use Patterns Across Europe. *Agric. Ecosyst. Environ.* **2025**, *381*, 109411. [CrossRef]
10. Geng, Y.; Zhou, T.; Zhang, Z.; Cui, B.; Sun, J.; Zeng, L.; Yang, R.; Wu, N.; Liu, T.; Pan, J.; et al. Continental-Scale Mapping of Soil pH with SAR-Optical Fusion Based on Long-Term Earth Observation Data in Google Earth Engine. *Ecol. Indic.* **2024**, *165*, 112246. [CrossRef]
11. Chen, S.; Chen, Z.; Zhang, X.; Luo, Z.; Schillaci, C.; Arrouays, D.; Richer-De-Forges, A.C.; Shi, Z. European Topsoil Bulk Density and Organic Carbon Stock Database (0–20 cm) Using Machine-Learning-Based Pedotransfer Functions. *Earth Syst. Sci. Data* **2024**, *16*, 2367–2383. [CrossRef]
12. Heikkinen, J.; Kostensalo, J.; Keskinen, R.; Soinne, H.; Nuutinen, V. Temporal Trends in Finnish Agricultural Soils: A Comparative Analysis of National and LUCAS Soil Monitoring Datasets. *Eur. J. Soil Sci.* **2024**, *75*, e13525. [CrossRef]
13. Ji, X.; Purushothaman, B.; Prasad, R.V.; Aravind, P. Developing a Digital Mapping of Soil Organic Carbon on a National Scale Using Sentinel-2 and Hybrid Models at Varying Spatial Resolutions. *Ecol. Indic.* **2024**, *167*, 112654. [CrossRef]
14. Bruni, E.; Lugato, E.; Chenu, C.; Guenet, B. European Croplands under Climate Change: Carbon Input Changes Required to Increase Projected Soil Organic Carbon Stocks. *Sci. Total. Environ.* **2024**, *954*, 176525. [CrossRef] [PubMed]
15. Feeney, C.J.; Bentley, L.; De Rosa, D.; Panagos, P.; Emmett, B.A.; Thomas, A.; Robinson, D.A. Benchmarking Soil Organic Carbon (SOC) Concentration Provides More Robust Soil Health Assessment than the SOC/Clay Ratio at European Scale. *Sci. Total. Environ.* **2024**, *951*, 175642. [CrossRef] [PubMed]
16. Wang, X.; Geng, Y.; Zhou, T.; Zhao, Y.; Li, H.; Liu, Y.; Li, H.; Ren, R.; Zhang, Y.; Xu, X.; et al. Mapping the Soil C:N Ratio at the European Scale by Combining Multi-Year Sentinel Radar and Optical Data via Cloud Computing. *Soil Tillage Res.* **2025**, *245*, 106311. [CrossRef]

17. Panagos, P.; De Rosa, D.; Liakos, L.; Labouyrie, M.; Borrelli, P.; Ballabio, C. Soil bulk density assessment in Europe. *Agric. Ecosyst. Environ.* **2024**, *364*, 108907. [CrossRef]
18. Demir, S.; Demir, S. Determination of suitable agricultural areas and current land use in Isparta Province, Türkiye, through a linear combination technique and geographic information systems. *Environ. Dev. Sustain.* **2024**, *26*, 13455–13493. [CrossRef]
19. De Rosa, D.; Ballabio, C.; Lugato, E.; Fasiolo, M.; Jones, A.; Panagos, P. Soil organic carbon stocks in European croplands and grasslands: How much have we lost in the past decade? *Glob. Chang. Biol.* **2024**, *30*, e16992. [CrossRef] [PubMed]
20. Bormpoudakis, D.; Giannarakis, G.; Sanchez-Cueto, P.; Sanchez, S.G.; Llado, S.; Hartmann, M.; Kontoes, C. Exploring the Links between Bacterial Diversity with Vegetation and Soil Parameters Using Soil Metabarcoding Data and Sentinel-2 Indices. In Proceedings of the International Geoscience and Remote Sensing Symposium (IGARSS), Athens, Greece, 7–12 July 2024; pp. 4261–4265.
21. Verhegghen, A.; D'Andrimont, R.; Waldner, F.; Van der Velde, M. Accuracy Assessment of the First Eu-Wide Crop Type Map with Lucas Data. In Proceedings of the IGARSS 2021—2021 IEEE International Geoscience and Remote Sensing Symposium, Brussels, Belgium, 11–16 July 2021; pp. 1990–1993.
22. Yailymov, B.; Yailymova, H.; Kussul, N.; Shelestov, A. Semi-Supervised European Forest Types Mapping using High-Fidelity Satellite Data. In *ProfIT AI 2024, Proceedings of the 4th International Workshop of IT-Professionals on Artificial Intelligence (ProfIT AI 2024), Cambridge, MA, USA, 25–27 September 2024*; CEUR Workshop Proceedings: Tenerife, Spain, 2024; Volume 3777, pp. 1–15.
23. Weigand, M.; Staab, J.; Wurm, M.; Taubenböck, H. Spatial and semantic effects of LUCAS samples on fully automated land use/land cover classification in high-resolution Sentinel-2 data. *Int. J. Appl. Earth Obs. Geoinf.* **2020**, *88*, 102065. [CrossRef]
24. Pflugmacher, D.; Rabe, A.; Peters, M.; Hostert, P. Mapping pan-European land cover using Landsat spectral-temporal metrics and the European LUCAS survey. *Remote Sens. Environ.* **2019**, *221*, 583–595. [CrossRef]
25. Ballot, R.; Guilpart, N.; Jeuffroy, M.-H. The first map of crop sequence types in Europe over 2012–2018. *Earth Syst. Sci. Data* **2023**, *15*, 5651–5666. [CrossRef]
26. Ghassemi, B.; Immitzer, M.; Atzberger, C.; Vuolo, F. Evaluation of Accuracy Enhancement in European-Wide Crop Type Mapping by Combining Optical and Microwave Time Series. *Land* **2022**, *11*, 1397. [CrossRef]
27. Ghassemi, B.; Dujakovic, A.; Żółtak, M.; Immitzer, M.; Atzberger, C.; Vuolo, F. Designing a European-Wide Crop Type Mapping Approach Based on Machine Learning Algorithms Using LUCAS Field Survey and Sentinel-2 Data. *Remote Sens.* **2022**, *14*, 541. [CrossRef]
28. Ghassemi, B.; Izquierdo-Verdiguier, E.; Verhegghen, A.; Yordanov, M.; Lemoine, G.; Martínez, Á.M.; De Marchi, D.; van der Velde, M.; Vuolo, F.; D'andrimont, R. European Union crop map 2022: Earth observation's 10-meter dive into Europe's crop tapestry. *Sci. Data* **2024**, *11*, 1048. [CrossRef] [PubMed]
29. Franco, A.; Vieira, D.; Clerbaux, L.; Orgiazzi, A.; Labouyrie, M.; Köninger, J.; Silva, V.; van Dam, R.; Carnesecchi, E.; Dorne, J.L.C.M.; et al. Evaluation of the Ecological Risk of Pesticide Residues from the European LUCAS Soil Monitoring 2018 Survey. *Integr. Environ. Assess. Manag.* **2024**, *20*, 1639–1653. [CrossRef] [PubMed]
30. Benevides, P.J.; Silva, N.; Costa, H.; Moreira, F.D.; Moraes, D.; Castelli, M.; Caetano, M. Land cover mapping at national scale with Sentinel-2 and LUCAS: A case study in Portugal. In Proceedings of the Remote Sensing for Agriculture, Ecosystems, and Hydrology XXIII, Online, 13–17 September 2021; p. 1185606.
31. Martinez-Sanchez, L.; See, L.; Yordanov, M.; Verhegghen, A.; Elvekjaer, N.; Muraro, D.; D'andrimont, R.; van der Velde, M. Automatic classification of land cover from LUCAS in-situ landscape photos using semantic segmentation and a Random Forest model. *Environ. Model. Softw.* **2023**, *172*, 105931. [CrossRef]
32. Paris, C.; Gasparella, L.; Bruzzone, L. A Scalable High-Performance Unsupervised System for Producing Large-Scale HR Land Cover Maps: The Italian Country Case Study. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2022**, *15*, 9146–9159. [CrossRef]
33. Gandhi, U. End-to-End Google Earth Engine Course. Spatial Thoughts. 2021. Available online: https://courses.spatialthoughts.com/end-to-end-gee.html (accessed on 25 December 2024).
34. Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
35. Vapnik, V.N. *The Nature of Statistical Learning Theory*; Springer: New York, NY, USA, 1995; p. 314.
36. Breiman, L.; Friedman, J.; Olshen, R.; Stone, C. *Classification and Regression Trees*; Wadsworth and Brooks; Chapman and Hall/CRC: New York, NY, USA, 1984.
37. Russell, G.C. A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sens. Environ.* **1991**, *37*, 35–46.
38. Foody, G.M. Impacts of Sample Design for Validation Data on the Accuracy of Feedforward Neural Network Classification. *Appl. Sci.* **2017**, *7*, 888. [CrossRef]
39. Canran, L.; Paul, F.; Lalit, K. Comparative assessment of the measures of thematic classification ac-curacy. *Remote Sens. Environ.* **2007**, *107*, 606–616.
40. Foody, G.M. Impacts of ignorance on the accuracy of image classification and thematic mapping. *Remote Sens. Environ.* **2021**, *259*, 112367. [CrossRef]

41. Adegun, A.A.; Viriri, S.; Tapamo, J.-R. Review of deep learning methods for remote sensing satellite images classification: Experimental survey and comparative analysis. *J. Big Data* **2023**, *10*, 93. [CrossRef]
42. Image Classification on Satellite Images with Deep Learning. 2024. Available online: https://medium.com/@ozgunhaznedar/image-classification-on-satellite-images-with-deep-learning-baa9813dde4e (accessed on 11 November 2024).
43. Lake, T.A.; Runquist, R.D.B.; Moeller, D.A. Deep Learning Detects Invasive Plant Species across Complex Landscapes Using Worldview-2 and Planetscope Satellite Imagery. *Remote Sens. Ecol. Conserv.* **2022**, *8*, 875–889. [CrossRef]
44. Tehsin, S.; Kausar, S.; Jameel, A.; Humayun, M.; Almofarreh, D.K. Satellite Image Categorization Using Scalable Deep Learning. *Appl. Sci.* **2023**, *13*, 5108. [CrossRef]
45. Dïker, F.; Erkan, I. Classification of satellite images with deep convolutional neural networks and its effect on architecture. *Eskişehir Tech. Univ. J. Sci. Technol. A-Appl. Sci. Eng.* **2022**, *23*, 31–41. [CrossRef]
46. Vaishnnave, M.P.; Devi, K.S.; Srinivasan, P. A Study on Deep Learning Models for Satellite Imagery. *Int. J. Appl. Eng. Res.* **2019**, *14*, 881–887.
47. Pritt, M.; Chern, G. Satellite image classification with deep learning. In Proceedings of the 2017 IEEE Applied Imagery Pattern Recognition Workshop (AIPR), Washington, DC, USA, 10–12 October 2017; pp. 1–7.
48. Hejmanowska, B.; Kramarczyk, P.; Głowienka, E.; Mikrut, S. Reliable Crops Classification Using Limited Number of Sentinel-2 and Sentinel-1 Images. *Remote Sens.* **2021**, *13*, 3176. [CrossRef]