The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XLVIII-1/W3-2023
2nd GEOBENCH Workshop on Evaluation and BENCHmarking of Sensors, Systems and GEOspatial Data
in Photogrammetry and Remote Sensing, 23–24 October 2023, Krakow, Poland

UNET NEURAL NETWORK IN AGRICULTURAL LAND COVER CLASSIFICATION USING SENTINEL-2

P.Kramarczyk [1], B.Hejmanowska [1] *

[1] AGH University of Science and Technology, Kraków galia@agh.edu.pl

KEY WORDS: UNet, Sentinel-2, landuse land cover, agriculture.

ABSTRACT:

The article discusses a method for classifying land cover types in rural areas using a trained neural network. The focus is on distinguishing agriculturally cultivated areas and differentiating bare soil from quarry areas. This distinction is not present in publicly available databases like CORINE, UrbanAtlas, EuroSAT, or BigEarthNet. The research involves training a neural network on multi-temporal patches to classify Sentinel-2 images rapidly. This approach allows automated monitoring of cultivated areas, determining periods of bare soil vulnerability to erosion, and identifying open-pit areas with similar spectral characteristics to bare soil. After training the U-Net network, it achieved an average classification accuracy of 90% (OA) in the test areas, highlighting the importance of using OA for multi-class classifications, instead of ACC. Analysis of our main classes revealed high accuracy, 99.01% for quarries, 92.3% for bare soil, and an average of 94.8% for annual crops, demonstrating the model's capability to differentiate between crops at various growth stages and assess land cover categories effectively.

1 INTRODUCTION

The article's topic concerns a universal method for classifying land cover types in rural areas. The idea is to train a neural network in such a way that it can perform classification (without training fields) of any Sentinel-2 image in any area to identify basic land cover types and land use forms. It is necessary to specify what this means. Some classes are the same in both urban and rural areas: buildings, industrial areas (including resource exploitation such as quarries and open-pit mines), roads, water bodies, permanent green areas, and forests. However, in rural areas, the focus is primarily on agriculturally cultivated areas. Depending on the season, plant phenology, or agrotechnical practices, the same surface may be bare soil or covered with vegetation at various stages of development. The novelty of our research lies in the detailed analysis of this group of land use types. We placed the main emphasis on distinguishing cultivated areas (classes 4, 5, and 9) and distinguishing bare soils (5) from quarry areas (10):

1. Coniferous forest
2. Mixed forest
3. Built-up areas, industrial areas
4. Crops: mature cereals (pre-harvest) or in spring before agricultural operations
5. Bare soil
6. Permanent grasslands
7. Roads
8. Water bodies
9. Crops: in the phase of vegetation growth
10. Quarry areas

Such a distinction of agriculturally cultivated areas is not present in publicly available databases such as CORINE, UrbanAtlas covering many countries, and local ones like S2GLC Global Land Cover - Sentinel 2 and POLSA Land Cover Classification. lntensive research is currently being conducted on various neural network models and the creation of benchmarks for land use and land cover classification (LULC). Examples of this research can be seen in EuroSAT and the BigEarthNet Sentinel-2 Multispectral Dataset (Helber et. al, 219, Zhang et. al., 2023). Both datasets consist of the same set of 10 classes, and it's worth noting that the 'annual crops' class is also excluded. Consequently, they cannot be utilized for the purposes defined above.

On the other hand, detailed monitoring of vegetation development at the crop level is carried out for the purpose of direct agricultural subsidies control, e.g. in Poland AMS (Area Monitoring System).

In our research, we have adopted some classes that are more generalized compared to CORINE, UrbanAtlas, EuroSAT or BigEarthNet and more detailed classes related to agricultural areas. However, these classes are not as detailed as needed for IACS (Integrated Administration and Control System) purposes. In this case, the goal is the detailed classification of cultivated plants, which also makes it impossible to use training data from this system (e.g., LPIS Land Parcel Identification System) for our purpose.

In our research, we trained a neural network on individual multi-temporal patches for the purpose of later using it to classify any single Sentinel-2 image. This approach allows for rapid and automated monitoring of agriculturally cultivated areas to determine the duration of periods when the soil is barren of vegetation and vulnerable to water erosion. Additionally, it is possible to monitor open-pit areas with spectral characteristics similar to bare soil during this process.

2 DATA AND METHODS

The test area was selected in southern Poland because of its fragmented land cover structure, where agricultural plots are mostly small, posing a challenge when utilizing Sentinel-2 images. Three test areas were chosen: Kolbuszowa, Strzegom, and Kraków, each with a different character.

1. Kolbuszowa covers a predominantly rural area with a high number of very small plots, often with a surface area below 0.5 hectares, and many of them have elongated shapes.
2. Strzegom is an area used for both agriculture and industry, with a significant presence of open-pit mining activities.
3. Due to the limited number of built-up areas in the first two patches, an additional area, Kraków, was selected. Kraków was used to represent various types of urban development.

These diverse test areas allow for a comprehensive evaluation of the neural network's performance across different land cover

* Corresponding author

and land use scenarios, ranging from rural agricultural landscapes to industrial and urban areas.

Figure 1. Test areas

The network training utilized Sentinel-2 data from various vegetative periods, spanning from March to September 2021-2022. This multi-temporal dataset likely allows the neural network to capture changes in land cover and vegetation over time, which can be valuable for tasks such as land cover classification, monitoring agricultural cycles, and assessing changes in the landscape throughout the growing season:

- strzegom_20210619
  strzegom_20220619
  strzegom_20220719
  strzegom_20221012
  strzegom_20230209
  strzegom_20230301
- kolbuszowa_20210327
  kolbuszowa_20210411
  kolbuszowa_20210509
  kolbuszowa_20210728
  kolbuszowa_20210906
- krakow_20220603
  krakow_20220603
  krakow_20220603

Testing was conducted on two patches:

- strzegom_20230709
- kolbuszowa_20210725

Each image was downloaded from the ESA hub in version 2A (after correction) and cropped to a patch size of 401x401 pixels. Ten channels were selected (B2, B3, B4, B5, B6, B7, B8, B8A, B11, B12), which were resampled to 10 meters and normalized to a 0-1 range.

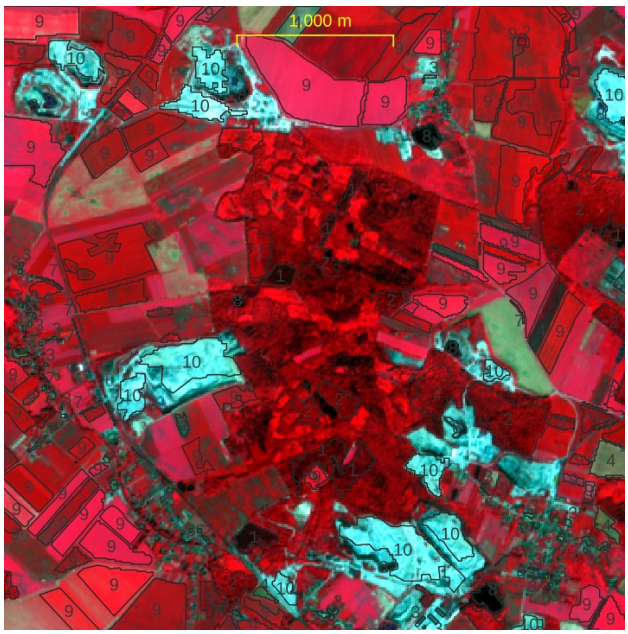

Figure 2. FCC - strzegom_20220619

Figure 3. FCC - strzegom_20220719

2.1 BENCHMARKS

The class patterns were vectorized using QGIS, particularly focusing on the false color compositions (FCC) of bands B8, B4, and B3. The following classes: forest (1,2), permanent grassland (6), and roads (9), were vectorized essentially once because they remain relatively stable over short time periods.
A more detailed interpretation was applied to the classes related to agricultural land use: crops in the mature cereals phase (pre-harvest) or in spring before agricultural operations (4) – appearing green on FCC, bare soils (6) – appearing cyan on FCC, and crops in the phase of vegetation growth (9) –

appearing red on FCC. The classes quarry areas (10) and water (8) were also verified on each composition.
Ultimately, you obtained 16 vector layers containing class patterns, with 14 for training and 2 for testing. These benchmark layers were then converted into a raster model, resulting in 16 raster masks where pixel values range from 1 to 9. The "no data" value was assigned to the background, while class values corresponded to land cover and land use types.

2.2 METHODS

Based on the publications (Selea 2023, Xu et al. 2023, Onojeghuo et al. 2023), the U-Net network model was selected, and its architecture is shown in Figure 5.1. The network was implemented using the PyTorch framework and the Python programming language. The basic convolution operation was designed as a block consisting of three layers:

- Convolution layer: Conv2d, kernel_size=3, stride=1, padding='same', bias=False
- Normalization layer: BatchNorm2d
- Activation layer: ReLU

At each level of the network (as shown in Figure 1), the image passes through the convolution block three times. Each passage through the block changes the number of image channels - it increases in the first part and decreases in the second part. After each convolution block, there is a double change in the size of the image (height and width): in the first part of the network, the image is downscaled (MaxPool2d), and in the second part, it is upscaled (ConvTranspose2d). All operations performed on the data by the U-Net model are provided as functions by the PyTorch framework.

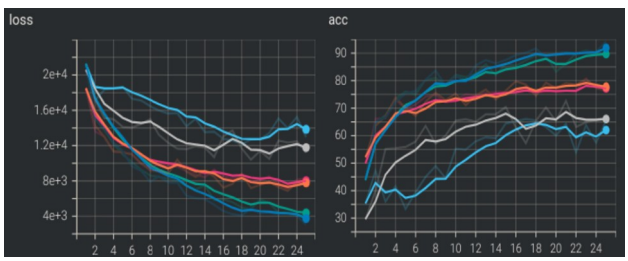

Figure 4. Sample plots of loss function values and accuracy. Hyperparameter testing for learning rate and weight decay.

The trained network was used for the classification of 2 test images. Accuracy analysis was conducted in QGIS using both pixel-based and object-based approaches.
Pixel-Based Approach:

1. Load the classified images into QGIS.
2. Utilize map algebra to subtract the benchmark mask from the classified images.
3. Calculate the accuracy metric, overall accuracy (OA), using the formula: OA = sum(pixels=0) / sum(pixels <> 0). This formula calculates the ratio of correctly classified pixels (pixels with a value of 0) to all non-background pixels.

Object-Based Approach:

1. Assign an attribute to each polygon in the benchmark mask, selecting a statistic such as the majority, computed from all pixels contained within that polygon.
2. Calculate the differences between the attribute value in the mask and the majority value from the classification result.
3. In this case, OA is calculated as the number of correctly classified polygons divided by the total number of polygons.

These methods allow for the assessment of classification accuracy in both a pixel-based and object-based manner, providing insights into how well the classification aligns with the benchmark data.

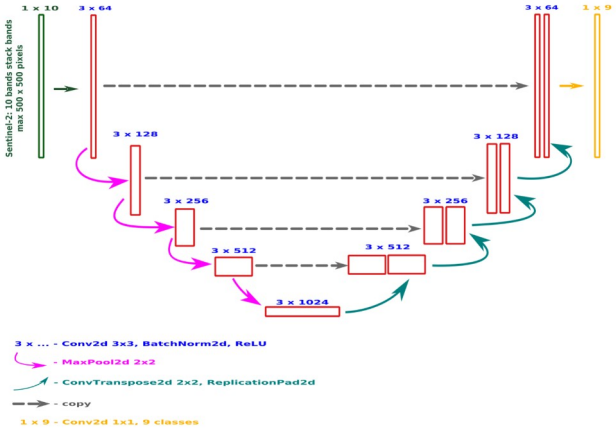

Figure 5. UNet scheme

A comprehensive accuracy analysis can be conducted based on the so-called confusion matrix, which contains all the information regarding errors, including underestimation and overestimation errors. In machine learning, terminology may not be entirely appropriate for assessing multi-class classification. Specifically, these terms correspond to binary classification, as commonly used in medical test evaluation: true positive (TP), true negative (TN), false positive (FP), and false negative (FN). However, due to the widespread use of these metrics in our research, we have calculated all of them using own scripts in Python.

Name	Formula
Producer accuracy (PA) Sensitivity True positive rate (TPR)	$\frac{TP}{TP+FN}$
Specificity True negative rate (TNR)	$\frac{TN}{TN+FP}$
User accuracy (UA) Precision Positive predictive value (PPV)	$\frac{TP}{TP+FP}$
Accuracy (ACC)	$\frac{TP+TN}{TP+TN+FP+FN}$
F1 score	$\frac{2\,TP}{2\,TP+FP+FN}$
Overall accuracy (OA) Percent of correct precision	$\frac{\sum_{i=1}^{n} TP_i}{\sum_{i=1}^{n} \left(TP_i+TN_i+FP_i+FN_i\right)}$

Table 1. Accuracy metrics

3 RESULTS

The false color composite (FCC) compositions of the test images are shown in Figures 6 and Figure 8, and the results of the classification of the test images are presented in Figures 7 and Figure 8. In Figure 7, class 6 (permanent grassland) is missing, and in Figure 9, class 10 (Quarry areas) is missing.
In both test areas, the classification accuracy in the pixel-based approach was higher than in the object-based approach. Slightly higher accuracy was achieved for the Strzegom area.
The confusion matrices and other accuracy metrics for both test areas and both approaches are provided in Tables 3 through Figure 10.

The summary average metric values in the classes can be found in Table 12. Notably, the average values of the commonly used machine learning metric ACC, often mistakenly reported as OA (Overall Accuracy), are worth mentioning, as they can artificially inflate the classification accuracy.

	piksel	object
kolbuszowa_20210725	93.1	86.5
strzegom_20230709	94.6	89.0

Table 2. OA for test images

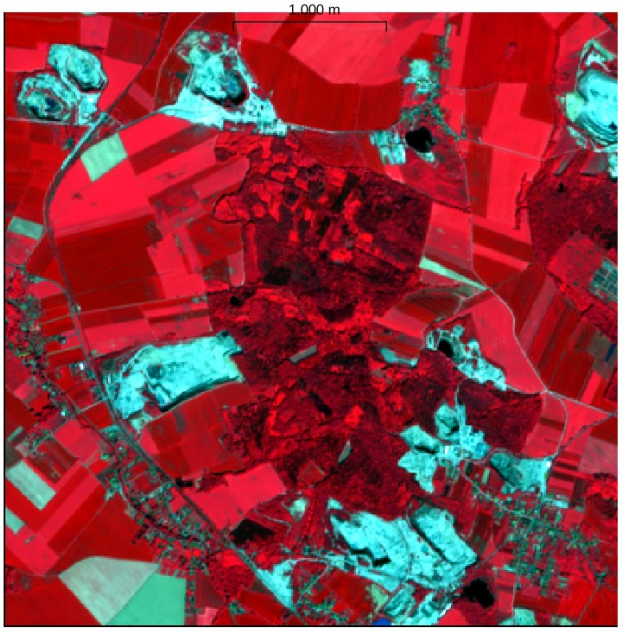

Figure 6. FCC - strzegom_20230709

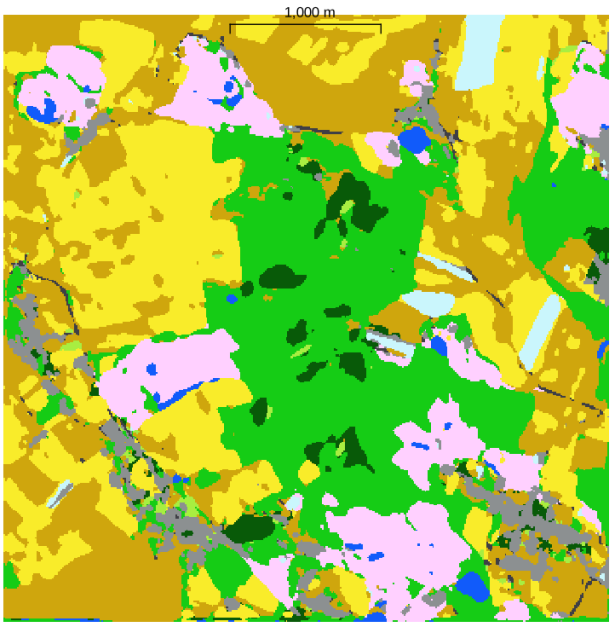

Figure 7. Classification result - strzegom_20230709

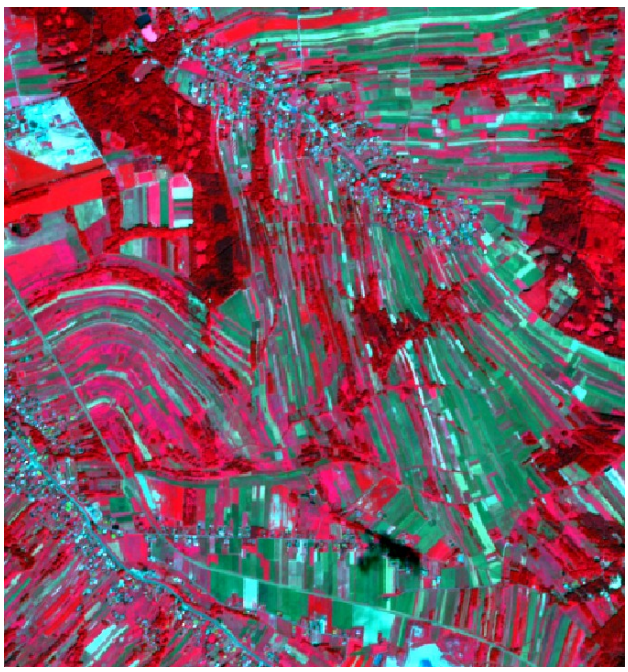

Figure 8. FCC - kolbuszowa_20210725

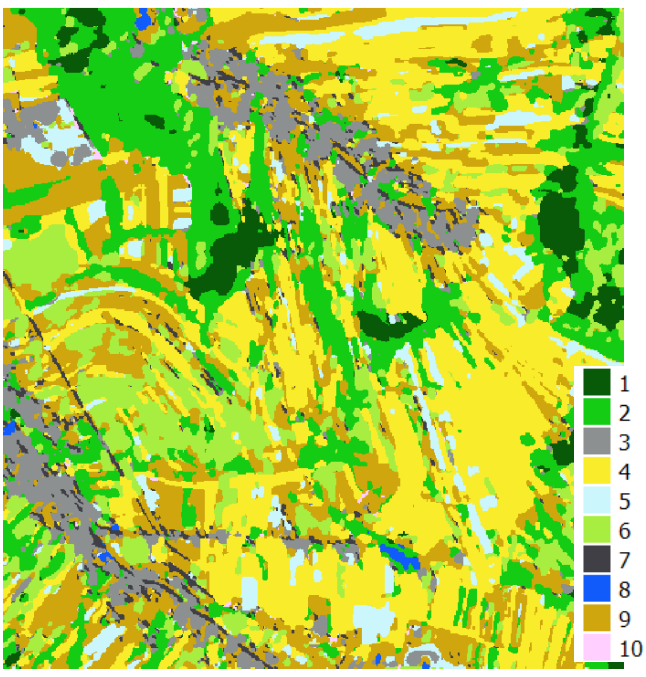

Figure 9. Classification result - kolbuszowa_20210725

	ground true									
ID	1	2	3	4	5	6	7	8	9	10
1	1202	0	6	47	0	0	1	0	0	0
2	549	7237	0	124	0	0	2	2	114	3
3	0	0	625	0	35	0	14	0	2	0
4	2	0	14	7341	0	0	17	0	15	7
5	0	0	0	6	1150	0	0	0	0	0
6	0	0	2	11	0	0	4	0	0	0
7	0	0	8	6	0	0	59	0	11	0
8	0	1	0	30	0	0	0	635	0	6
9	0	14	25	1717	0	0	67	0	9744	0
10	0	9	0	2	89	0	0	4	0	6998

Table 3. Strzegom, confusion matrix, pixel approach

	ground true									
ID	1	2	3	4	5	6	7	8	9	10
1	5	0	0	0	0	0	0	0	0	0
2	2	21	1	1	0	0	0	0	0	0
3	0	0	43	0	0	0	2	0	0	0
4	0	0	1	45	0	0	2	0	0	0
5	0	0	0	0	4	0	0	0	0	0
6	0	0	0	0	0	0	0	0	0	0
7	0	0	2	0	0	0	4	0	0	0
8	0	0	0	0	0	0	0	13	0	0
9	0	1	0	2	0	0	6	0	21	0
10	0	0	0	0	1	0	0	0	0	14

Table 4. Strzegom, confusion matrix, object approach

	TPR/PA	TNR	PPV/UA	ACC	F1
1	68.57	99.85	95.47	98.42	79.81
2	98.89	97.17	89.62	97.51	94.02
3	81.91	99.81	89.54	99.45	85.56
4	91.06	99.34	97.33	97.61	94.1
5	81.73	99.98	99.48	99.32	89.74
6		99.96	0	99.96	
7	35.98	99.93	70.24	99.66	47.58
8	99.06	99.9	94.49	99.89	96.73
9	98.14	97.56	94.22	97.73	96.14
10	99.76	99.61	98.27	99.64	99.01

Table 5. Strzegom, accuracy metrics, pixel approach

ID	TPR/PA	TNR	PPV/UA	ACC	F1
1	71.43	100	100	98.95	83.33
2	95.45	97.63	84	97.38	89.36
3	91.49	98.61	95.56	96.86	93.48
4	93.75	97.9	93.75	96.86	93.75
5	80	100	100	99.48	88.89
6					
7	28.57	98.87	66.67	93.72	40
8	100	100	100	100	100
9	100	94.71	70	95.29	82.35
10	100	99.44	93.33	99.48	96.55

Table 6. Strzegom, accuracy metrics, object approach

	ground true									
ID	1	2	3	4	5	6	7	8	9	10
1	1105	7	0	1	0	0	0	0	0	0
2	42	1591	1	9	6	0	0	0	0	0
3	2	0	324	22	4	0	23	0	0	0
4	0	0	9	2813	45	11	65	0	3	0
5	0	0	6	72	1221	2	2	0	0	0
6	10	102	18	4	5	530	0	0	1	0
7	0	0	12	3	6	0	105	0	0	0
8	0	0	1	0	0	0	1	21	0	0
9	15	5	11	21	0	105	0	0	1105	0
10	0	0	0	0	0	0	0	0	0	0

Table 7. Kolbuszowa, confusion matrix, pixel approach

	ground true									
ID	1	2	3	4	5	6	7	8	9	10
1	9	1	0	0	0	0	0	0	0	0
2	0	16	1	0	0	0	0	0	0	0
3	0	0	31	1	0	0	2	0	0	0
4	0	0	1	48	3	0	1	0	0	0
5	0	0	0	3	25	0	2	0	0	0
6	1	1	0	0	0	20	0	0	0	0
7	0	0	3	0	1	0	6	0	0	0
8	0	0	0	0	0	0	0	1	0	0
9	1	0	0	0	0	5	0	0	15	0
10	0	0	0	0	0	0	0	0	0	0

Table 8. Kolbuszowa, confusion matrix, object approach

	TPR/PA	TNR	PPV/UA	ACC	F1
1	94.12	99.9	99.28	99.19	96.63
2	93.31	99.25	96.48	98.18	94.87
3	84.82	99.44	86.4	98.85	85.6
4	95.52	97.96	95.49	97.2	95.5
5	94.87	99	93.71	98.44	94.29
6	81.79	98.41	79.1	97.27	80.42
7	53.57	99.77	83.33	98.82	65.22
8	100	99.98	91.3	99.98	95.45
9	99.64	98.12	87.56	98.3	93.21
10					

Table 9. Kolbuszowa, accuracy metrics, pixel approach

ID	TPR/PA	TNR	PPV/UA	ACC	F1
1	81.82	99.47	90	98.48	85.71
2	88.89	99.44	94.12	98.48	91.43
3	86.11	98.15	91.18	95.96	88.57
4	92.31	96.58	90.57	95.45	91.43
5	86.21	97.04	83.33	95.45	84.75
6	80	98.84	90.91	96.46	85.11
7	54.55	97.86	60	95.45	57.14
8	100	100	100	100	100
9	100	96.72	71.43	96.97	83.33
10					

Table 10. Kolbuszowa, accuracy metrics, object approach

	TPR/PA	TRN	PPV/UA	ACC	F1
1	83.9	99.31	82.87	**98.92**	86.97
2	84.52	98.57	89.26	**97.56**	85.30
3	88.63	99.09	90.29	**98.47**	89.02
4	85.54	98.23	85.73	**96.97**	85.27

Table 12. Accuracy metrics – summary, 1 – Strzegom, pixel approach, 2 – Strzegom, object approach, 3 – Kolbuszowa, pixel approach, 4 – Kolbuszowa, object approach

4 CONCLUSIONS

During the research, the analysis focused on LULC classes in agricultural areas, with a particular emphasis on distinguishing between areas used as annual crops and areas with industrial activity (quarry areas). After training the U-Net network, an average classification accuracy of OA = 90% was achieved on the test areas. It is important to emphasize that instead of reporting ACC (average accuracy), which can give the illusion of significantly higher accuracy (e.g., Zhang et. al. 2023) one should use OA. In our case, we could report an accuracy of 98%, which is 8% higher than OA. Similarly, it is not recommended to use the metric TRN (specificity) for the same reason (Hejmanowska et. al. 2021). Both accuracy and specificity metrics are suitable for binary classifications but not appropriate for multi-class classifications.

Regarding the specific classes 4, 5, 9, and 10, you can analyze the "remote sensing" metrics of PA (Producer's Accuracy) and UA (User's Accuracy) in Tables 5 and 9:

- The classification accuracy for class 10 (quarry area) is on average 99.01%.
- The classification accuracy for the very similar class, bare soil, is 92.3%.
- The classification accuracy for areas covered by annual crops, classes 4 and 9, is on average 94.8%.

This level of accuracy metrics the ability to distinguish between such crops in various growth stages (vegetation phase or resting state) or just before harvesting. These remote sensing metrics help assess the model's performance in correctly identifying and classifying specific land cover categories, including areas with different agricultural practices and land use.

Similar research was conducted by Yailymova et al. in 2022 in order to detect illegal dumpsites, reporting accuracy metrics kappa = 91%. A Sentinel-2 based multispectral convolutional neural network for detecting artisanal small-scale mining was researched (Gallwey et. al. 2020) with OAmax = 96%. On the other hand, multi-temporal analyses of marble quarry expansion allowed us to achieve an OAmax of 93.13% (Tercan and Dereli 2021). Interesting comparisons of various cases studies can be found in publications (Ang et. al. 2023, Chen et. al. 2018, De Fioravante et. al. 2021).

REFERENCES

Ang, M. L. E., Owen, J. R., Gibbins, C. N., Lèbre, É., Kemp, D., Saputra, M. R. U., Everingham, J.-A., & Lechner, A. M., 2023. Systematic Review of GIS and Remote Sensing Applications for Assessing the Socioeconomic Impacts of Mining. *The Journal of Environment & Development,* 32(3), 243-273. https://doi.org/10.1177/10704965231190126

Chen, W.; Li, X.; He, H.; Wang, L., 2018: A Review of Fine-Scale Land Use and Land Cover Classification in Open-Pit Mining Areas by Remote Sensing Techniques. *Remote Sens.* 10, 15. https://doi.org/10.3390/rs10010015

De Fioravante, P., Luti, T., Cavalli, A., Giuliani, C., Dichicco, P., Marchetti, M., Chirici, G., Congedo, L., Munafò, M. Multispectral Sentinel-2 and SAR Sentinel-1 Integration for Automatic Land Cover Classification. *Land 2021*, 10, 611. https://doi.org/10.3390/land10060611

Gallwey J., Robiati C., Coggan J., Vogt D., Eyre M., 2020: A Sentinel-2 based multispectral convolutional neural network for detecting artisanal small-scale mining in Ghana: Applying deep learning to shallow mining, *Remote Sensing of Environment,* Volume 248, 2020, 111970, ISSN 0034-4257,

Helber P., Bischke B., Dengel A., Borth D., 2019: Eurosat: A Novel Dataset and Deep Learning Benchmark for Land Use and Land Cover Classification, I*EEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, 12 (7) (2019), pp. 2217-2226, 10.1109/JSTARS.2019.2918242

Hejmanowska B., Kramarczyk P., Głowienka E., Mikrut S., 2021: Reliable Crops Classification Using Limited Number of Sentinel-2 and Sentinel-1 Images. *Remote Sensing.* 2021; 13(16):3176. https://doi.org/10.3390/rs13163176

Onojeghuo A. O., Miao Y., Blackburn G. A., 2023: Deep Resu-Net Convolutional Neural Networks Segmentation for Smallholder Paddy Rice Mapping Using Sentinel 1 SAR and Sentinel 2 Optical Imagery. *Remote Sens.*, 15(6)

Selea T., 2023: AgriSen-COG, a Multicountry, Multitemporal Large-Scale Sentinel-2 Benchmark Dataset for Crop Mapping Using Deep Learning. *Remote Sens.*, 15(12)

Tercan, E., Dereli, M.A., 2021: Monitoring of marble quarries expansion and land cover changes using satellite images and GIS on a rural settlement of Burdur province, Turkey. *El-Cezerî Journal of Science and Engineering,* 2021, 8 (2); 741-750.

Xu Y., Xue X., Sun Z., Gu W., Cui L., Jin Y., Lan Y., 2023: Deriving agricultural field boundaries for crop management from satellite images using semantic feature pyramid network. *Remote Sens.*, 15(11).

Yailymova H., Mikava P., Kussul N., Krasilnikova T., Shelestov A., Yailymov B., Titkov D., 2022: Neural Network Model for Monitoring of Landfills Using Remote Sensing Data. *IEEE 3rd International Conference on System Analysis & Intelligent Computing (SAIC),* Kyiv, Ukraine, 2022, pp. 1-4, doi: 10.1109/SAIC57818.2022.9923013.

Zhang P., Wu, Y., Li C., Li R., Yao H., Zhang, Y., Zhang, G., Li, D., 2023: National-Standards- and Deep-Learning-Oriented Raster and Vector Benchmark Dataset (RVBD) for Land-Use/Land-Cover Mapping in the Yangtze River Basin. *Remote Sens.* 2023, *15*, 3907. https://doi.org/10.3390/rs15153907