AGILE: GIScience Series, 3, 42, 2022. https://doi.org/10.5194/agile-giss-3-42-2022
Proceedings of the 25th AGILE Conference on Geographic Information Science, 2022.
Editors: E. Parseliunas, A. Mansourian, P. Partsinevelos, and J. Suziedelyte-Visockiene.
This contribution underwent peer review based on a full paper submission.

Comparison of CNN-based segmentation models for forest type classification

Kevin Kocon [1,2], Michel Krämer [1,2], and Hendrik M. Würz [1,2]

[1]Fraunhofer Institute for Computer Graphics Research IGD, Fraunhoferstraße 5, Darmstadt, Germany
[2]Technical University of Darmstadt, Karolinenplatz 5, Darmstadt, Germany

Correspondence: Kevin Kocon (kevin.kocon@igd.fraunhofer.de)

Abstract. We present the results from evaluating various Convolutional Neural Network (CNN) models to compare their usefulness for forest type classification. Machine Learning based on CNNs is known to be suitable to identify relevant patterns in remote sensing imagery. With the availability of free data sets (e.g. the Copernicus Sentinel-2 data), Machine Learning can be utilized for forest monitoring, which provides useful and timely information helping to measure and counteract the effects of climate change. To this end, we performed a case study with publicly available data from the federal state of North Rhine-Westphalia in Germany. We created an automated pipeline to preprocess and filter this data and trained the CNN models UNet, PSPNet, SegNet, and FCN-8. Since the data contained large rural areas, we augmented the imagery to improve classification results. We reapplied the trained models to the data, compared the results for each model, and evaluated the effect of augmentation. Our results show that UNet performs best with a categorical accuracy of 73% when trained with augmented imagery.

Keywords. Machine Learning, Augmentation, Remote Sensing, Convolutional Neural Network

1 Introduction

Climate change, one of the most important topics of our time, leads to increasing temperatures, extreme weather events, floods, wildfires, and other global effects (Watts et al., 2021). It not only has a negative impact on the environment, but also directly on humans. According to Watts et al., "in the past two decades, heat-related mortality in the over-65 population has increased by 53.7%, reaching 296,000 deaths in 2018". In addition, there is an increased risk of infectious diseases, respiratory diseases, and undernutrition due to food insecurity (Watts et al., 2021).

An important factor in the issue of climate change are forests. On the one hand, they are severely damaged by climate change itself (Senf and Seidl, 2021). On the other hand, forests are a crucial parameter to counteract its effects. In Germany alone, forests absorb about 52 million tons of carbon dioxide annually (Federal Ministry of Food and Agriculture, 2022a). Therefore, it is particularly important to act against forest dieback. For this, precise environmental monitoring is of great importance. The forests in Germany (and other global regions) are currently inventoried manually every 10 years (Federal Ministry of Food and Agriculture, 2022b). This is associated with very high costs and large efforts. At the same time, this interval is too long, as it is particularly important to be able to quickly react to the dynamic changes (Banskota et al., 2014).

For these reasons, there is great potential in forest monitoring based on remote sensing data, which is available with high temporal resolution (mostly for free, e.g. Copernicus Sentinel-2 imagery). It has already been shown that various Machine Learning approaches, especially those involving Convolutional Neural Networks (CNN), have the ability to recognize relevant patterns from these recordings (Wessel et al., 2018; Yang et al., 2018).

However, there is a gap in literature as there is, to the best of our knowledge, no publication yet that compares different CNN models and investigates their usefulness for forest type classification. We believe that such a comparison would provide useful insights for researchers and other stakeholders involved in forest monitoring and would help them select a suitable model for their use case.

In this paper, we therefore present the results of a case study we conducted with the aim to evaluate various CNN models for forest type classification. For this case study, we selected a suitable study area, trained the widely used UNet, FCN-8, SegNet, and PSPNet models with publicly available remote sensing data from this area, performed an evaluation, and then compared the results. In this process, we also identified the lack of ground-truth data as a common problem during training. To improve the accuracy of the trained models, we augmented the training data by randomly transforming the source images.

2 Related Work

Analyzing remote sensing data with Machine Learning has become increasingly popular in recent years (Ma et al., 2019). One of the most active application areas in this respect is land use and land cover (LULC) classification (Feng et al., 2017). Random Forest (RF), Support Vector Machine (SVM), and CNN approaches have been used for classification, which all deliver good results (Thanh Noi and Kappas, 2017). Raczko and Zagajewski (2017) compared SVMs, RFs and CNNs for LULC classification and showed that CNNs outperform the other two approaches.

However, there still is a gap in knowledge in the field of forest type classification (Mäyrä et al., 2021). Existing works have focused on the use of hyperspectral data (Pan et al., 2018) or even the combination of hyperspectral and LiDAR images (Mäyrä et al., 2021), but these data are not freely available and their temporal resolution is quite limited. Wessel et al. (2018) have already obtained good results using Sentinel-2 imagery. They used RF and SVM approaches and claim that images from the month of May are best for forest type classification.

Selecting the right Sentinel-2 bands is crucial for classification quality. Wessel et al. (2018) obtained good results with bands 6 (mid red edge), 7 (long red edge), and 8 (NIR). A drawback with this selection is that bands 6 and 7 only have a resolution of 20 meters per pixel. Persson et al. (2018) used all bands and also obtained good results. More channels provide a deeper input layer. However, a deeper network is harder to train (Du et al., 2019). Another approach is to use bands 2 (blue), 3 (green) and 8 (NIR). Ng et al. (2017) achieved good results with them and showed that the higher resolution of these bands has a positive effect on accuracy.

Various papers have compared the usefulness of CNN models for LULC classification. For example, Zhang et al. (2020) covered UNet (Ronneberger et al., 2015), PSPNet (Zhao et al., 2017), and SegNet (Badrinarayanan et al., 2017) models. The PSPNet and UNet performed best. On the other hand, Storie and Henry (2018) obtained the best results with a FCN-8 (Long et al., 2015). Additionally, Stivaktakis et al. (2019) showed that CNN models with prior augmentation outperform all CNN models without augmentation. This is a relevant aspect since labeled training data in the field of LULC and forest type classification are rare (Chen et al., 2014; Pan et al., 2018).

The papers mentioned above cover LULC, but, to the best of our knowledge, there is no existing publication comparing different CNN models for forest type classification. This is a research gap that we aim to close with this paper

3 Case study

In the following sections, we describe a case study we performed with the aim to classify forest types using a CNN. For this, we had to select a study area with publicly available imagery and then prepared and filtered the data with an automated pipeline. Since the selected area was mostly rural, we had to augment the data to generate more input images. We then trained UNet, PSPNet, SegNet, and FCN-8 with 80% of the data, applied the CNNs to the remaining 20% and compared the results.

3.1 Study Area

The German state of North Rhine-Westphalia (NRW) provides geodata about forest type distribution for free (MULNV Nordrhein-Westfalen, 2022). It is served through an OGC Web Map Service (WMS). The data is based on cadastral information and distinguishes between deciduous, coniferous, mixed forests, as well as areas without forests. NRW has an area of 34,000 square kilometers and is located in the northwest of Germany. Predominant parts of it are populated, which means that the relative proportion of forest is low. We queried the WMS in March 2022.

In addition, we used Sentinel-2 imagery from the Copernicus project. The data is freely available and can be queried via an API (Copernicus, 2022). We used six cloud-free images taken on May 7, 2020.

3.2 Data Preparation

To prepare the data for training, we created a processing pipeline with several steps. The pipeline executes these steps automatically one after the other and generates training data suitable for the different CNN models. Most of the steps are the same for all models, only the *Split* step produces images in different resolutions.

1) Search The pipeline searches the Copernicus project database for satellite images that were taken in the desired time period, show the surface of NRW, and have as few clouds as possible. Since the images contain metadata, the search is easily possible via the Copernicus API.

2) Download The selected satellite images are downloaded to our servers. Copernicus delivers them as archives containing different bands and several metadata files.

3) Filter Bands The pipeline extracts the files relevant for us from the archives. The training requires bands 2 (blue), 3 (green), and 8 (NIR).

4) WMS An image pair is created that consists of ground truth data from the North Rhine-Westphalia WMS and the already downloaded satellite image for the same area. For this, we download the PNG maps from the WMS and interpret the pixels as forest type classifications.

5) Split This image pair covers an area of 100×100 kilometers and has a resolution of 10×10 meters. However, for training, we need several small image pairs. The pipeline splits the image pair into sub-images with a resolution suitable for the network to be trained. SegNet and FCN-8 use a resolution of 224×224 pixels as input, UNet and PSPNet a resolution of 512×512 pixels.

6) Filter Forest Images Finally, the image pairs are filtered according to their forest proportion. For the subsequent training, only those image pairs are retained that contain at least a forest fraction of 50% per image. This results in the following distribution of the classes to be trained.

- **Non-forest:** 28%
- **Coniferous forest:** 40%
- **Deciduous forest:** 19%
- **Mixed forest:** 12 %
- **Invalid data:** 1%

Although this distribution is not ideal, more filtering would further reduce the data set. With this, the pipeline creates about 1000 image pairs with a resolution of 224×224 pixels, as well as about 240 image pairs with a resolution of 512×512 pixels.

3.3 Data Augmentation

As described above, large parts of NRW are rural. The number of images we could use for training our CNNs was rather low compared to the whole data set. We therefore investigated how to deal with this. One way is to augment the data. The idea is to get new information by transforming the existing images. This increases the size of the data set, which counteracts overfitting and improves accuracy (Shorten and Khoshgoftaar, 2019). During augmentation, we randomly applied between one and three of the following geometric transformations to each image:

- Flip left-right
- Flip up-down
- Random rotation: [-180°,180°]
- Random scaling x, y: [0.7,1.3]
- Random translation x, y: [-20px,20px]

This ensured the augmenting was maximally randomized, which resulted in strongly differing new images. We augmented every image seven times in each epoch. Figure 1 shows some visual results of the augmentation.

3.4 CNN Models

To compare the classification quality of different CNN models, we implemented UNet, PSPNet, SegNet, and FCN-8. As mentioned above, these models already achieved good results in LULC classification (Zhang et al., 2020; Storie and Henry, 2018). We implemented them based on the original papers in Keras 2.4.3 (Chollet et al., 2022) with the TensorFlow (Abadi et al., 2022) backend in version 2.4.1. Our training parameters are shown in Table 1. They are based on the work of Maeda-Gutiérrez et al. (2020) who achieved good results with them.

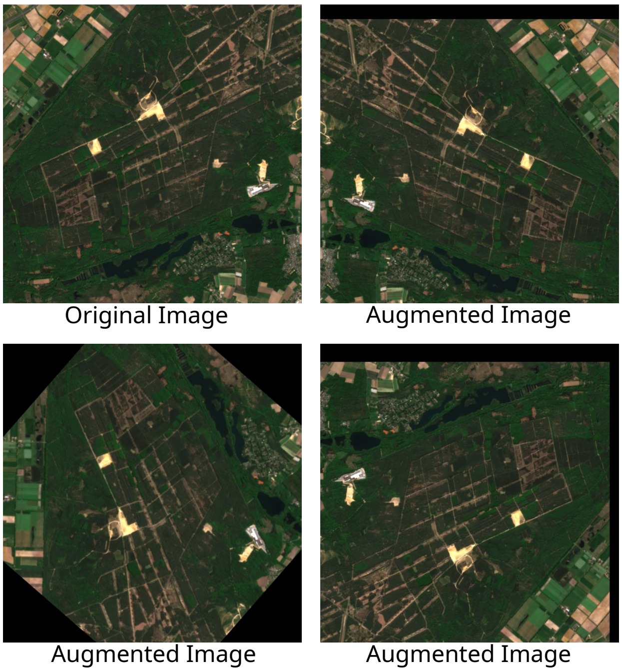

Figure 1. Augmented images based on Copernicus Sentinel data (2022)

Table 1. Used Hyper-Parameters for Training.

Parameter	Value
Optimizer	Stochastic Gradient Descent
Loss Function	Categorical Crossentropy
Learning Rate	10^{-3}
Decay	5^{-4}
Momentum	0.9
Batch Size	16
Epochs	30

For training, we used 80% of randomly selected image pairs from the overall data. Furthermore, we implemented a batch-based approach. This means that the entire data did not have to be prepared as a whole and then pushed into the GPU for training, but rather only the data of one batch, one after the other. This approach had the benefit that the training could already begin in parallel with the preparation of the remaining data. In addition, the processed data did not have to be stored on the hard disk first, but could be pushed directly into the GPU.

4 Results

After we had used 80% of the data for the training above, we applied the models to the remaining 20% to evaluate

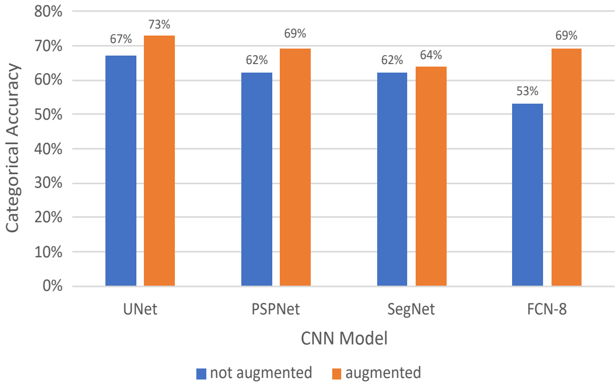

Figure 2. Comparison of Categorical Accuracies between the different CNN models.

Table 2. Normalized Confusion Matrix of the augmented UNet. **(1)** Non-forest, **(2)** Coniferous forest, **(3)** Mixed forest, **(4)** Desiduous forest

		Predicted class			
		1	**2**	**3**	**4**
Ground truth	**1**	0.85	0.10	0.01	0.04
	2	0.08	0.87	0.02	0.03
	3	0.15	0.52	0.15	0.18
	4	0.18	0.22	0.01	0.59

their quality. Figure 2 summarizes the results. It shows the *Categorical Accuracy* of each model when trained with the original input data as well as the augmented images.

Augmentation has a positive effect on the accuracy of all models. For the FCN-8 model, we achieved an improvement of 16%. A potential reason for this is the number of parameters of the FCN-8, which is significantly higher compared to the rest of the models but also requires more data for training. This clearly shows that augmenting the data is a good way to extend the training data in this use case and achieves better accuracy. Furthermore, we noticed during our evaluation that the maximum accuracy is reached with the generation of 7 new images per image. Both fewer and more augmentations per image have a negative impact on accuracy.

Overall, the UNet achieves the best accuracy with 73%. This result is also in line with the publication of Zhang et al. (2020) on the LULC classification, where UNet also performed best. In the study of Storie and Henry (2018), the FCN-8 model performed best, but it is not known exactly with which other models the comparison was made.

For a more detailed evaluation of the augmented UNet model, the normalized confusion matrix is shown in Table 2. It can be seen that the detection of non-forest areas and coniferous forest works very well. The recognition of deciduous forest does not perform as well. One reason for this is that deciduous forest is underrepresented in the training data.

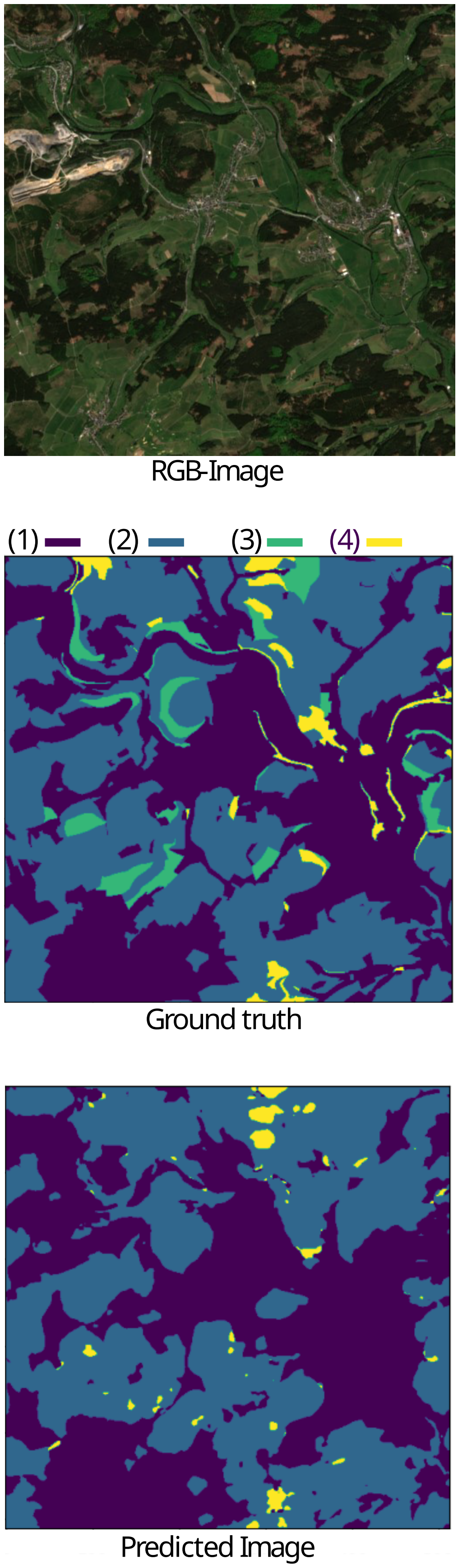

Figure 3. A RGB-, ground truth- and predicted image from our model.**(1)** Non-forest, **(2)** Coniferous forest, **(3)** Mixed forest, **(4)** Desiduous forest. The RGB image contains modified Copernicus Sentinel data (2022)

Mixed forest is almost not recognized at all. This is also due to underrepresentation but mainly to the characteristics of a mixed forest, which consists of deciduous trees and conifers. Therefore, the model recognizes the components of the mixed forest instead of classifying it as such. Figure 3 shows a satellite image, a ground-truth image as well as the image as classified by our model. In this predicted image, the described behaviors already seen in the confusion matrix are visible.

5 Conclusion

In this paper, we compared four different CNN models and evaluated their usefulness for forest type classification. To this end, we performed a case study with publicly available data including imagery from the Copernicus Sentinel-2 mission. We created an automated pipeline to preprocess and filter the data and augmented it to improve classification accuracy. We then trained all four models and reapplied them to the input data to evaluate their quality.

Our results show that UNet performs best with an accuracy of 73%. This result also shows that data augmentation has a positive impact on the classification, as UNet only achieved an accuracy of 67% with the original, unaugmented data.

It must be pointed out that, for the sake of comparability, we applied training and classification only to those images that our pipeline retained after filtering in step 6 (those that consisted at least of 50% of forest). For this filtered set of images, UNet already achieved a good accuracy, but, as a reference, we also applied UNet to the entire unfiltered data set (all of NRW). Here, it achieved an even higher accuracy of 93%.

These results show that CNNs are well-suited for forest type classification. Using them for forest monitoring could be of great potential. We believe that this is a necessary step towards understanding the impact of climate change on our Earth and to save the forests, which play an important role in counteracting negative global effects.

References

Abadi, M. et al.: TensorFlow, available at: https://www.tensorflow.org, last access: 22 March 2022, 2022.

Badrinarayanan, V., Kendall, A., and Cipolla, R.: Segnet: A deep convolutional encoder-decoder architecture for image segmentation, IEEE transactions on pattern analysis and machine intelligence, 39, 2481–2495, 2017.

Banskota, A., Kayastha, N., Falkowski, M. J., Wulder, M. A., Froese, R. E., and White, J. C.: Forest monitoring using Landsat time series data: A review, Canadian Journal of Remote Sensing, 40, 362–384, 2014.

Chen, Y., Lin, Z., Zhao, X., Wang, G., and Gu, Y.: Deep learning-based classification of hyperspectral data, IEEE Journal of Selected topics in applied earth observations and remote sensing, 7, 2094–2107, 2014.

Chollet, F. et al.: Keras, available at: https://www.keras.io, last access: 22 March 2022, 2022.

Copernicus: Open Access Hub, available at: https://scihub.copernicus.eu/, last access: 22 March 2022, 2022.

Du, S., Lee, J., Li, H., Wang, L., and Zhai, X.: Gradient descent finds global minima of deep neural networks, in: International conference on machine learning, pp. 1675–1685, PMLR, 2019.

Federal Ministry of Food and Agriculture: Surveying the forest, available at: https://www.bundeswaldinventur.de/en/third-national-forest-inventory/surveying-the-forest/, last access: 22 March 2022, 2022a.

Federal Ministry of Food and Agriculture: The forests as climate protectors – still a carbon sink, available at: https://www.bundeswaldinventur.de/en/third-national-forest-inventory/the-forests-as-climate-protectors-still-a-carbon-sink/, last access: 22 March 2022, 2022b.

Feng, Y., Lu, D., Moran, E. F., Dutra, L. V., Calvi, M. F., and De Oliveira, M. A. F.: Examining spatial distribution and dynamic change of urban land covers in the Brazilian Amazon using multitemporal multisensor high spatial resolution satellite imagery, Remote Sensing, 9, 381, 2017.

Long, J., Shelhamer, E., and Darrell, T.: Fully convolutional networks for semantic segmentation, in: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 3431–3440, 2015.

Ma, L., Liu, Y., Zhang, X., Ye, Y., Yin, G., and Johnson, B. A.: Deep learning in remote sensing applications: A meta-analysis and review, ISPRS journal of photogrammetry and remote sensing, 152, 166–177, 2019.

Maeda-Gutiérrez, V., Galvan-Tejada, C. E., Zanella-Calzada, L. A., Celaya-Padilla, J. M., Galván-Tejada, J. I., Gamboa-Rosales, H., Luna-Garcia, H., Magallanes-Quintanar, R., Guerrero Mendez, C. A., and Olvera-Olvera, C. A.: Comparison of convolutional neural network architectures for classification of tomato plant diseases, Applied Sciences, 10, 1245, 2020.

Mäyrä, J., Keski-Saari, S., Kivinen, S., Tanhuanpää, T., Hurskainen, P., Kullberg, P., Poikolainen, L., Viinikka, A., Tuominen, S., Kumpula, T., et al.: Tree species classification from airborne hyperspectral and LiDAR data using 3D convolutional neural networks, Remote Sensing of Environment, 256, 112 322, 2021.

MULNV Nordrhein-Westfalen: Waldinfo.NRW 2.26.0, available at: https://www.waldinfo.nrw.de/waldinfo2/?lang=en, last access: 22 March 2022, 2022.

Ng, W.-T., Rima, P., Einzmann, K., Immitzer, M., Atzberger, C., and Eckert, S.: Assessing the Potential of Sentinel-2 and Pléiades Data for the Detection of Prosopis and Vachellia spp. in Kenya, Remote sensing, 9, 74, 2017.

Pan, B., Shi, Z., and Xu, X.: MugNet: Deep learning for hyperspectral image classification using limited samples, ISPRS Journal of Photogrammetry and Remote Sensing, 145, 108–119, 2018.

Persson, M., Lindberg, E., and Reese, H.: Tree species classification with multi-temporal Sentinel-2 data, Remote Sensing, 10, 1794, 2018.

Raczko, E. and Zagajewski, B.: Comparison of support vector machine, random forest and neural network classifiers for tree species classification on airborne hyperspectral APEX images, European Journal of Remote Sensing, 50, 144–154, 2017.

Ronneberger, O., Fischer, P., and Brox, T.: U-net: Convolutional networks for biomedical image segmentation, in: International Conference on Medical image computing and computer-assisted intervention, pp. 234–241, Springer, 2015.

Senf, C. and Seidl, R.: Persistent impacts of the 2018 drought on forest disturbance regimes in Europe, Biogeosciences, 18, 5223–5230, 2021.

Shorten, C. and Khoshgoftaar, T. M.: A survey on image data augmentation for deep learning, Journal of big data, 6, 1–48, 2019.

Stivaktakis, R., Tsagkatakis, G., and Tsakalides, P.: Deep learning for multilabel land cover scene categorization using data augmentation, IEEE Geoscience and Remote Sensing Letters, 16, 1031–1035, 2019.

Storie, C. D. and Henry, C. J.: Deep learning neural networks for land use land cover mapping, in: IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium, pp. 3445–3448, IEEE, 2018.

Thanh Noi, P. and Kappas, M.: Comparison of random forest, k-nearest neighbor, and support vector machine classifiers for land cover classification using Sentinel-2 imagery, Sensors, 18, 18, 2017.

Watts, N., Amann, M., Arnell, N., Ayeb-Karlsson, S., Beagley, J., Belesova, K., Boykoff, M., Byass, P., Cai, W., Campbell-Lendrum, D., et al.: The 2020 report of the Lancet Countdown on health and climate change: responding to converging crises, The Lancet, 397, 129–170, 2021.

Wessel, M., Brandmeier, M., and Tiede, D.: Evaluation of different machine learning algorithms for scalable classification of tree types and tree species based on Sentinel-2 data, Remote Sensing, 10, 1419, 2018.

Yang, C., Rottensteiner, F., and Heipke, C.: Classification of land cover and land use based on convolutional neural networks, ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences 4 (2018), Nr. 3, 4, 251–258, 2018.

Zhang, X., Han, L., Han, L., and Zhu, L.: How well do deep learning-based methods for land cover classification and object detection perform on high resolution remote sensing imagery?, Remote Sensing, 12, 417, 2020.

Zhao, H., Shi, J., Qi, X., Wang, X., and Jia, J.: Pyramid scene parsing network, in: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 2881–2890, 2017.