EVALUATION OF DIFFERENT PARAMETERS FOR PLANT CLASSIFICATION BY PRE-TRAINED DEEP LEARNING MODELS WITH BIGEARTHNET DATASET

F. Naali [1*], T. Alipour-Fard [1], H. Arefi [1,2]

[1] School of Surveying and Geospatial Engineering, University of Tehran, Tehran 1417414418, Iran – (f.naali, tayebalipour, hossein.arefi)@ut.ac.ir
[2] Institute for Spatial Information and Surveying Technology, Mainz University of Applied Sciences, Germany - hossein.arefi@hs-mainz.de

Commission IV, WG IV/3

KEY WORDS: Crop Classification, Convolution Neural Network, Deep Learning, Sentinel-2, ResNet

ABSTRACT:

Vegetation monitoring and mapping are essential for a diverse range of environmental problems such as forest management, food resources, and climate change assessment. Several methods have been developed to classify different vegetation types based on remote sensing (RS) data. Land use classification has been revolutionized with the advent of neural networks. Various vegetation types were classified using multispectral Sentinel-2 satellite images due to their high spatial resolution and spectral information. Deep Convolutional Neural Network is considered a promising method for classifying remote sensing images with high spatial resolution due to its powerful feature extraction capabilities. However, large labeled datasets are required for better classification performance, so we have used pre-trained ResNet networks with 152 layers, 50 layers, and 101 layers trained on Big Earth Net (BEN). In order to obtain the best network performance and evaluate the sensitivity of the parameters in this study, we have performed two experiments: 1) the effect of different patch sizes and 2) increasing the number of images. The results demonstrate that ResNet 152 shows the highest accuracy with patches of 120×120 pixels, with an accuracy of 76.62%, and ResNet 50 is the best with an accuracy of 76.2% since the process of this network does not take much time.

1. INTRODUCTION

Increasing population growth has significant implications for environmental problems and food shortages (Zhai et al. 2020). it is necessary to manage and monitor croplands and forests to preserve natural resources and food reserves for future generations. For this purpose, vegetation classification maps should be developed at local and global levels (Ahmed et al. 2017). Achieving environmental goals such as addressing climate change requires information about land use. (Cui et al. 2019), food security (Mutanga, Dube, and Galal 2017), ecosystem dynamics (Dubayah et al. 2020), etc. High-resolution satellite imagery plays a central role in studying land-use characteristics of the Earth's surface (Wulder et al. 2019). The launch of the Sentinel-2 satellite has promoted the development of a wide range of land studies and programs (Thanh Noi and Kappas 2017). Accurate classification of vegetation using Earth observation techniques is being used to assess stratification (Zhou et al. 2019), changes in cropping patterns (Weiss, Jacob, and Duveiller 2020), drought risk (Skakun et al. 2016), and levels of deforestation. Developing and improving satellite imagery requires advanced algorithms for processing and mapping.

A machine learning technique known as Deep Learning (DL) has become increasingly popular in recent years. Image classification with Deep Learning is used to automatically learn an internal feature representation and extract complex features (Li et al. 2019). Deep learning models automatically extract features to perform classification tasks without requiring a specialist. Automatic land cover mapping solves tasks such as

land cover change detection. The need for high-precision maps to detect changes on a weekly and daily time scale is essential for monitoring deforestation and illegal logging. Deep learning algorithms such as convolutional neural networks (CNN) are extremely successful. The best-known CNN architectures are (ResNet (He et al. 2016), VGG (Simonyan and Zisserman 2015), and Inception (Längkvist, Karlsson, and Loutfi 2014)), which are used in computer vision, medicine, and remote sensing image processing. One of the advantages of DL networks is the flexibility in reducing and increasing the number of layers for different tasks (Lecun, Bengio, and Hinton 2015). Evidence suggests that network depth plays a critical role in network performance; the more layers added, the more complex features the network can extract (Shrestha and Mahmood 2019). Deeper networks produce better results but require many labeled samples. Collecting labeled datasets is complex and requires a lot of manpower and resources to label remote sensing data. If the number of labeled samples is not large enough, the problem of overfitting will occur. There are two ways to overcome this problem: 1) data augmentation (Perez and Wang 2017) and 2) transfer learning (Hu et al. 2022).

As a deep learning technique, transfer learning overcomes the problem of limited labeling, where a model trained from a large dataset is transferred to a new, related task. The ImageNet dataset has caused a revolution in transfer learning due to the large number of images it contains. ImageNet is a computer vision dataset created with 3 RGB images (Russakovsky et al. 2015) and is therefore insufficient for remote sensing tasks since images with multiple bandwidths and resolutions are commonly used in remote sensing to identify different features.

---

* Corresponding author

Besides, the semantic contents in the CV and RS datasets are different, so the class labels must be different. Although RS images have a variety of semantic contents, they represent various features on one image, so more than one label should be assigned to each image to solve feature complexity for image classification and retrieval tasks. Most remote sensing datasets consist of a single label. In the DFC15 dataset (Hua, Mou, and Zhu 2019, aerial images with a 5 cm spatial resolution and multiple labels were used, but not in large enough numbers for the development of DL models. Table 1 shows a list of remote sensing benchmarks. To address this issue, Sumbul et al. (Sumbul et al. 2019) provided a multi-label multi-resolution BigEarthNet dataset with 590,326 images 2019, which yielded significant results for image classification (Sumbul et al. 2020) and image retrieval (Sumbul et al. 2021) with DL networks. In this study, the BigEarthNet pre-training models' performance is presented in a classification map.

Dataset Name	Image Type	Annotation Type	Number of Images	Number of Classes
UC Merced	Aerial RGB	Single Label	2100	21
WHU-RS19	Aerial RGB	Single Label	1005	19
RSSCN7	Aerial RGB	Single Label	2800	7
AID	Aerial RGB	Single Label	10000	30
NWPU-RESISC45	Aerial RGB	Single Label	31500	45
RSI-CB	Aerial RGB	Single Label	36707	12
PatternNet	Aerial RGB	Single Label	30400	38
EuroSat	Satellite Multispectral	Single Label	27000	10
DFC15	Aerial RGB	Multi Label	3342	8
BigEarthNet	Sentinel-2	Multi Label	590326	43

Table 1. The list of RS benchmarks

2. STUDY AREA AND DATA

2.1 Study area

In this article, we used two study areas. The first region is Switzerland in Western Europe with (45 49 2 N, 5 57 22 E) geographical coordinates. This area has a completely similar land use pattern to BEN classes. On the other hand, this area has included images taken in BEN, so it has been chosen for investigation in this study. The first study area is shown in Figure 1.
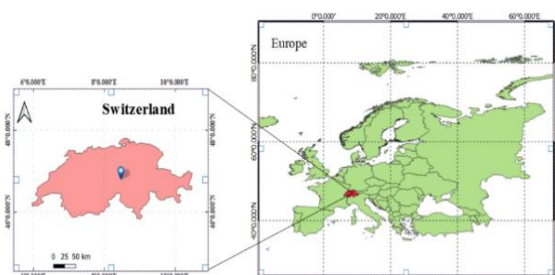

Figure 1. A study area's location

The second study area is the Strasbourg, Saarbrucken, and Karlsruhe regions, located in southwestern Germany and part of France. More data was needed to implement the second strategy. Although not in the BigEarthNet archive images, the images taken from this area have similar vegetation classes and low cloud cover.

2.1.1 BigEarthNet Dataset
Sumbul et al. introduced the BigEarthNet(BEN) dataset in 2019 (Sumbul et al. 2019). BEN is the first multi-label, and the multi-resolution dataset has been created from remote sensing images. The 125 tiles used in this dataset were taken from June 2017 to May 2018 in ten European countries (Switzerland, Belgium, Austria, Ireland, Kosovo, Lithuania, Serbia, Portugal, Luxembourg, and Finland). In detail, the sentinel-2 images are in the UTM coordinate system. Atmospheric corrections were made using Sen2cor on these images. There are 590,326 patches total, and they are divided into three groups: 120 by 120 pixels for 10m bands; 60 by 60 pixels for 20m bands; and 20 by 20 pixels for 60m bands. The dataset is available to download from http://bigearthnet.net/

2.1.2 Sentinel-2
In this research, Sentinel 2 high-resolution images were used. Figure 2 shows the RGB of the Sentinel 2 image of Switzerland, a Level 1C product; on this date, cloud cover was 7%. The tile we used has an area of 110×110 km and It is georeferenced using the coordinate system UTM 31N based on WGS 84. The 10th band has been removed because it did not have information from the surface.
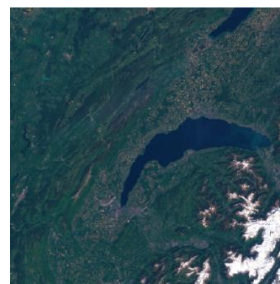

Figure 2. RGB image from Sentinel-2

The second data used, related to southwestern Germany and part of France, consists of 4 tiles, 2 of which are level 2A products, and the other two are level 1C products (Figure 3). The imaging system of all tiles is WGS, and the date of taking the images was the same as the Switzerland data. We are using the second data to study the performance of the network when more images are uploaded because the images in the neighborhood of first image had a cloud coverage of about 70%, which was not suitable for automatic classification; we used secondary data.

Figure 3. RGB image of Sentinel-2 from the second study area

3. METHOD

3.1 pre-processing of ground truth and sentinel-2 images

In order to quantitatively evaluate the classification results, we need a reference map. Since the study area is Switzerland, Germany, and France, we downloaded the land use map of Europe from **https://land.copernicus.eu.** The CLC map includes 43 classes with a spatial resolution of 100 meters and the coordinate system ETRS_1989_LAEA. We first reclassified the ground truth (GT) map according to the 19 classes in the BigEarthNet dataset with ArcGIS software. The map coordinate system and Sentinal-2 image must be the same to match geographically. Using the "Data management tool," we converted the map coordinate system to WGS_1984_UTM_Zone_31N for Switzerland and WGS_1984_UTM_Zone_32N for the second data. We considered the spatial resolution of the map to match the image. After fitting the image and the map, we cut the image area with the "spatial analyst tools." Figures 4 and 5 show GT for the first and second data, respectively.
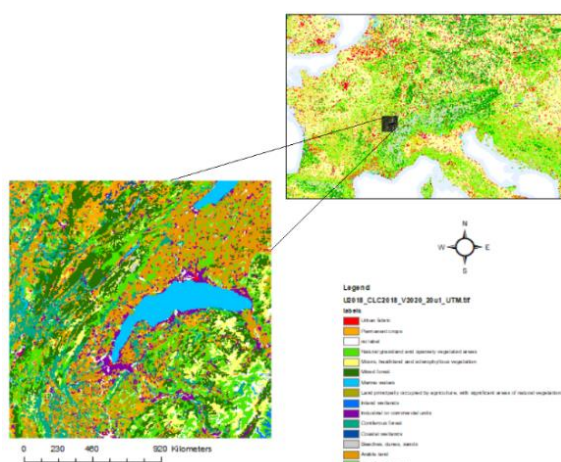

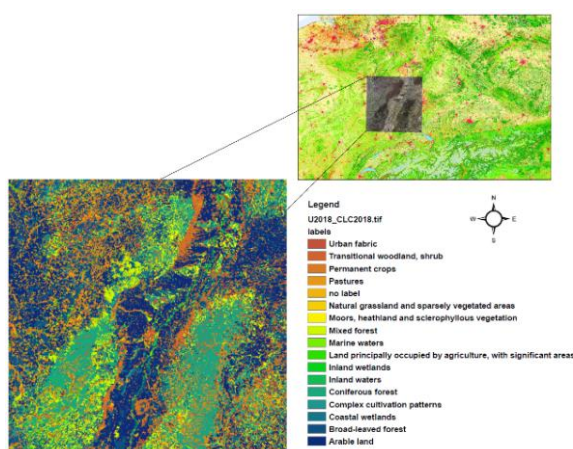**Figure 4**. Ground truth for first data

Figure 5. Ground truth for second data

Atmospheric corrections on the 1C level product of Sentinel-2 images using the Sen2cor plugin (the sen2cor algorithm designed in SNAP software provides the ability to eliminate the effect of the atmosphere with high accuracy on this data) SNAP

software did it. To feed images to the network, we have to convert the images into patches and then into tensors. In this study, to achieve the best network performance and evaluate the sensitivity of the parameters, in sections 3.1.1 and 3.1.2, we explain two strategies we used. The difference between variant experiments is in the number and dimensions of the patches.

3.1.1 The effect of different sizes of patches

The first strategy is to survey the impact of patch dimensions on the efficiency of CNN for the image with a spatial resolution of 20 meters, which we divided into 60×60 dimensions in the previous step; we create patches with 90×90 and 120×120 pixels. We also turn these patches into tensors. Figure 6 shows patches with different dimensions.
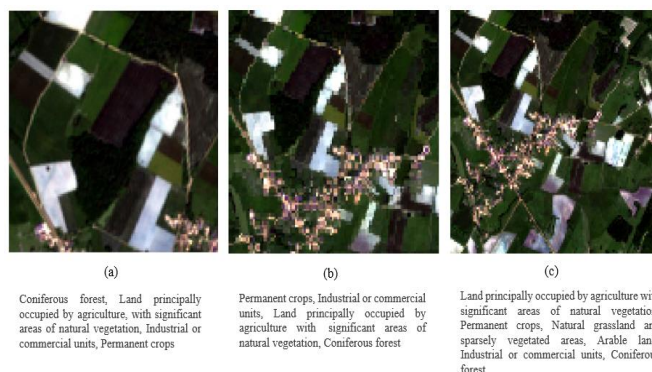

Figure 6. a, b, c patches with dimensions 60,90,120 pixels, respectively

3.1.2 increasing the number of images

In the second experiment, we used the second data, four images, to investigate the increase in the number of images on the network performance. After atmospheric correction, similar to the previous data processing, we converted all the bands to a spatial resolution of 20 meters using the nearest neighbourhood method. We create patches with dimensions of 60×60 pixels, then turn them into a tensor. Since the number of images has quadrupled; as a result, the number of patches also increases and is equal to 33124.

3.2 Classification

To model the classes, by transferring knowledge from BigEarthNet to Sentinel-2 images, using pre-trained ResNet50, ResNet101, and ResNet152 models. In these networks, in order to decrease the sigmoid cross entropy loss, the Adam approach (Kingma and Ba 2015) with a learning rate of 10^{-3} has been used, and all models were trained in 100 epochs. The batch size for the ResNet 50,101 was 500, and the ResNet 152 was 256. In the first step, we tested all three networks with images with a resolution of 10 m obtained from the P + XS fusion method. To increase the accuracy of the networks, we fine-tuned them to do this from 70% of the data (5799 patches) for training and the remaining 30% (2485 patches) for testing. We used. We trained all three networks with 20 epochs and batch size 128 for ResNet 50 and ResNet 101 and 64 for ResNet 152.

To compare the increase in the size of the patches, the networks with patches with dimensions of 90×90 pixels and 120×120 pixels, which we explained in section 3.1.1, and the parameters that we set in the previous step, fine-tuned and then tested. The final step of the research is the effect of increasing training samples on network performance, so we used the second data, which includes four images belonging to regions of Germany

and France. First, we fine-tuned the networks with a quarter of the data, one of the images (5799 patches), and tested it with the rest of the patches (2485 patches). In the next step, we used all four images (23196 patches) for fine-tuning the networks with the new data and 9940 patches for testing. In this step, all the parameters for fine-tuning are the same as in the previous steps. The results of this section are displayed in the results chapter.

We used Tensorflow 1.3 (Agarwal et al. 2015)) and the NVIDIA GeForce GTX 1080 Ti GPU to run these networks.

4. RESULTS AND DISCUSSION

The outcome of the first experiment, which increased the patch size on the precision of the classifier networks, are presented in the following chart (Figure 7).
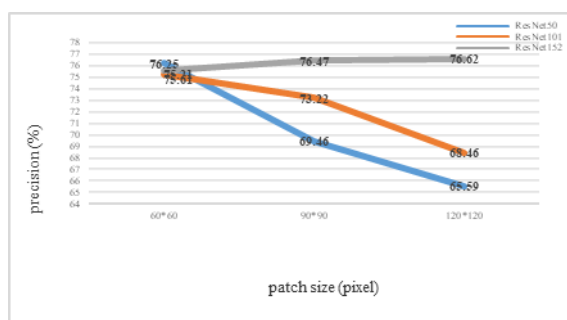

Figure 7. Network precision based on different patch dimensions

According to the results of evaluating networks with different patch dimensions, it is observed that with increasing patch dimensions, the accuracy of ResNet50 and ResNet101 networks decreases, and with increasing patch dimensions, the variety of features (classes) per patch increases. It is observed that the deeper network (ResNet152) has better accuracy for large patches because deeper networks have a better ability to learn complex features. Still, the computations in these networks are high and require more time.

In Table 2, the precision of the networks is shown as a function of increasing the number of images. Increasing the number of images will increase the average precision of the networks by 7% and improve the precision of the classes more.

Classes	ResNet 50		ResNet 101		ResNet 152	
	One image	Four images	One image	Four images	One image	Four images
Arable land	77.32	79.15	78.93	81.39	82.02	74.55
Permanent crop	53.75	72.31	56.57	**83.37**	66.82	**87.19**
Pastures	71.56	**73.79**	65.35	66.07	64.67	61.27
Complex cultivation patterns	24.32	29	50	20.17	0	14.31
Land principally occupied by agriculture	20.83	11.45	0	16.6	0	0
Broad-leaved forest	68.41	69.53	68.99	67.54	71.23	71.97
Coniferous forest	76	72.13	79.24	74.82	71.3	87.46
Mixed forest	42.16	65.57	73.63	63.65	72.63	64.19
Transitional woodland, shrub	21.77	28.77	0	18.78	0	6.6
Recall	60.59	65.52	54.59	**68.76**	70.16	62.75
Precision	66.47	73.23	68.78	**76.19**	73.79	75.24

Table 2. The precision of each class obtained from pre-trained ResNet 50,101,152 based on BEN dataset with fine-tuned networks on one and four sentinel-2 images

By examining the precision of each class and the visual interpretation of the classes from the land cover, it can be concluded that the distribution of classes also affects their accuracy in addition to the number of training samples.

According to Fig 8, it can be seen that although the Pastures class has more pixels (about six times) than the Permanent crop class and is scattered on the map, the Permanent class crop is integrated the accuracy of these two classes is relatively equal. Besides, it can be seen that the Land principally occupied by the agriculture class has decreased in accuracy with increasing the number of images; the same conclusion can be drawn for this class.
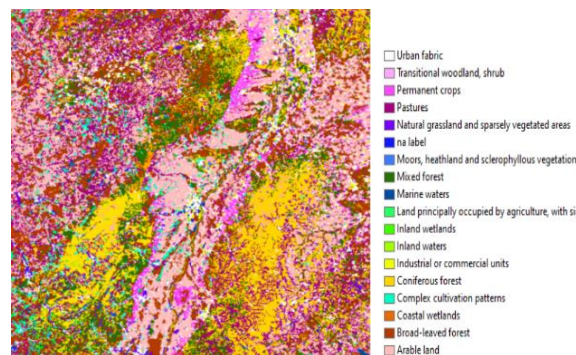

Figure 8. Ground truth

5. CONCLUSIONS

Vegetation types classification maps are critical tools for managing and planning environmental problems. The goal of this study was to create accurate classification maps using pre-trained networks trained on BigEarthNet datasets. In order to get the best efficiency, we do various tests on ResNet 50, ResNet 101, and ResNet 152 networks. According to the first experiment, which aimed to examine the effect of varying patch scales, increasing patch size has a direct effect on increasing feature complexity. The ResNet 152 network has a resolution of 76.62 for a 120×120 pixel patch because the deeper network has a better ability to learn complex features but to ensure this conclusion, we need to look at more different dimensions of the patches. In the second experiment, which aimed to increase the ability to identify vegetation classes, we increased the number of training samples. The overall precision of the classes increased by an average of 7%; in addition to increasing training samples, cohesion also affected network performance. Based on the overall accuracy achieved for the networks, it can be concluded that the BigEarthNet dataset stands out from other machine vision databases as it allocates multiple labels to each patch and uses multiband satellite imagery. (In a dataset like ImageNet, only three RGB bands are used). For each class, this advantage can be observed; however, complexity and similarity of features may make some classes difficult to distinguish. The results of this study show that there is still space for progress in the classification performance for deep CNN. In future studies, sentinel-1 and sentinel-2 images should be merged.

REFERENCES

Agarwal, Ashish, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S Corrado, Andy Davis, et al. 2015. "TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems."

Ahmed, Oumer S., Adam Shemrock, Dominique Chabot, Chris Dillon, Griffin Williams, Rachel Wasson, and Steven E. Franklin. 2017. "Hierarchical Land Cover and Vegetation Classification Using Multispectral Data Acquired from an Unmanned Aerial Vehicle." *International Journal of Remote Sensing* 38 (8–10): 2037–52. https://doi.org/10.1080/01431161.2017.1294781.

Cui, Xintong, Xiaoyu Guo, Yidi Wang, Xuelei Wang, Weihong Zhu, Jianghong Shi, Chunye Lin, and Xiang Gao. 2019. "Application of Remote Sensing to Water Environmental Processes under a Changing Climate." *Journal of Hydrology* 574 (April): 892–902. https://doi.org/10.1016/j.jhydrol.2019.04.078.

Dubayah, Ralph, James Bryan Blair, Scott Goetz, Lola Fatoyinbo, Matthew Hansen, Sean Healey, Michelle Hofton, et al. 2020. "The Global Ecosystem Dynamics Investigation: High-Resolution Laser Ranging of the Earth's Forests and Topography." *Science of Remote Sensing* 1 (September 2019): 100002. https://doi.org/10.1016/j.srs.2020.100002.

He, Kaiming, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. "Identity Mappings in Deep Residual Networks." *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 9908 LNCS: 630–45. https://doi.org/10.1007/978-3-319-46493-0_38.

Hu, Yueran, Hongwei Zeng, Fuyou Tian, Miao Zhang, Bingfang Wu, Sven Gilliams, Sen Li, Yuanchao Li, Yuming Lu, and Honghai Yang. 2022. "An Interannual Transfer Learning Approach for Crop Classification in the Hetao Irrigation District, China." *Remote Sensing* 14 (5): 1–25. https://doi.org/10.3390/rs14051208.

Hua, Yuansheng, Lichao Mou, and Xiao Xiang Zhu. 2019. "ImageNet Large Scale Visual Recognition Challenge." *ISPRS Journal of Photogrammetry and Remote Sensing* 149 (December 2018): 188–99. https://doi.org/10.1016/j.isprsjprs.2019.01.015.

Kingma, Diederik P., and Jimmy Lei Ba. 2015. "Adam: A Method for Stochastic Optimization." *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*, 1–15.

Längkvist, Martin, Lars Karlsson, and Amy Loutfi. 2014. "Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning." *Pattern Recognition Letters* 42 (1): 11–24. http://arxiv.org/abs/1512.00567.

Lecun, Yann, Yoshua Bengio, and Geoffrey Hinton. 2015. "Deep Learning." *Nature* 521 (7553): 436–44. https://doi.org/10.1038/nature14539.

Li, Shutao, Weiwei Song, Leyuan Fang, Yushi Chen, Pedram Ghamisi, and Jon Atli Benediktsson. 2019. "Deep Learning for Hyperspectral Image Classification: An Overview." *IEEE Transactions on Geoscience and Remote Sensing* 57 (9): 6690–6709. https://doi.org/10.1109/TGRS.2019.2907932.

Mutanga, Onisimo, Timothy Dube, and Omer Galal. 2017. "Remote Sensing of Crop Health for Food Security in Africa: Potentials and Constraints." *Remote Sensing Applications: Society and Environment* 8: 231–39. https://doi.org/10.1016/j.rsase.2017.10.004.

Perez, Luis, and Jason Wang. 2017. "The Effectiveness of Data Augmentation in Image Classification Using Deep Learning." http://arxiv.org/abs/1712.04621.

Russakovsky, Olga, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, et al. 2015. "ImageNet Large Scale Visual Recognition Challenge." *International Journal of Computer Vision* 115 (3): 211–52. https://doi.org/10.1007/s11263-015-0816-y.

Shrestha, Ajay, and Ausif Mahmood. 2019. "Review of Deep Learning Algorithms and Architectures." *IEEE Access* 7: 53040–65. https://doi.org/10.1109/ACCESS.2019.2912200.

Simonyan, Karen, and Andrew Zisserman. 2015. "Very Deep Convolutional Networks for Large-Scale Image Recognition." *ArXiv Preprint ArXiv:1409.1556*.

Skakun, Sergii, Nataliia Kussul, Andrii Shelestov, and Olga Kussul. 2016. "The Use of Satellite Data for Agriculture Drought Risk Quantification in Ukraine." *Geomatics, Natural Hazards and Risk* 7 (3): 901–17. https://doi.org/10.1080/19475705.2015.1016555.

Sumbul, Gencer, Marcela Charfuelan, Begum Demir, and Volker Markl. 2019. "Bigearthnet: A Large-Scale Benchmark Archive for Remote Sensing Image Understanding," 5901–4. https://doi.org/10.1109/igarss.2019.8900532.

Sumbul, Gencer, Jian Kang, Tristan Kreuziger, Filipe Marcelino, Hugo Costa, Pedro Benevides, Mario Caetano, and Begüm Demir. 2020. "BigEarthNet Dataset with A New Class-Nomenclature for Remote Sensing Image Understanding," 1–14. http://arxiv.org/abs/2001.06372.

Sumbul, Gencer, Arne de Wall, Tristan Kreuziger, Filipe Marcelino, Hugo Costa, Pedro Benevides, Mário Caetano, Begüm Demir, and Volker Markl. 2021. "BigEarthNet-MM: A Large Scale Multi-Modal Multi-Label Benchmark Archive for Remote Sensing Image Classification and Retrieval," 1–5. https://doi.org/10.1109/MGRS.2021.3089174.

Thanh Noi, Phan, and Martin Kappas. 2017. "Comparison of Random Forest, k-Nearest Neighbor, and Support Vector Machine Classifiers for Land Cover Classification Using Sentinel-2 Imagery." *Sensors (Basel, Switzerland)* 18 (1). https://doi.org/10.3390/s18010018.

Weiss, M., F. Jacob, and G. Duveiller. 2020. "Remote Sensing for Agricultural Applications: A Meta-Review." *Remote Sensing of Environment* 236 (December 2018): 111402. https://doi.org/10.1016/j.rse.2019.111402.

Wulder, Michael A., Thomas R. Loveland, David P. Roy, Christopher J. Crawford, Jeffrey G. Masek, Curtis E. Woodcock, Richard G. Allen, et al. 2019. "Current Status of Landsat Program, Science, and Applications." *Remote Sensing of Environment* 225 (November 2018): 127–47. https://doi.org/10.1016/j.rse.2019.02.015.

Zhai, Yongguang, Nan Wang, Lifu Zhang, Lei Hao, and Caihong Hao. 2020. "Automatic Crop Classification in Northeastern China by Improved Nonlinear Dimensionality

Reduction for Satellite Image Time Series." *Remote Sensing* 12 (17). https://doi.org/10.3390/RS12172726.

Zhou, Keqi, Dongping Ming, Xianwei Lv, Ju Fang, and Min Wang. 2019. "CNN-Based Land Cover Classification Combining Stratified Segmentation and Fusion of Point Cloud and Very High-Spatial Resolution Remote Sensing Image Data." *Remote Sensing* 11 (17): 1–28. https://doi.org/10.3390/rs11172065.